Molecular genetic studies on Brassica napus L.

Napis, Suhaimi

MOLECULAR GENETIC STUDIES ON BRASSICA NAPUS L.

by

Suhaimi Napis

**A Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy**

Department of Biological Sciences

The University of Durham
1991

DECLARATION

No part of this thesis has previously been submitted for a degree in this or any other university. I hereby declare that the work presented herein is entirely my own work.

Suhaimi Napis

ABSTRACT

The feasibility of using two different methods of assaying for DNA polymorphisms has been assessed. They were Restriction fragment length polymorphisms (RFLPs) as revealed by a range of characterised *Brassica* cDNA sequences and Random amplified polymorphic DNA (RAPD). These techniques have been shown to reveal DNA polymorphisms between varieties of *Brassica napus* L.. Furthermore, the sequence and organisation of a *Hind* III family of highly repetitive DNA sequences were also studied on *Brassica napus* L.

RFLPs associated with rape extensin, *ext A*, and *Brassica oleraceae* self-incompatibility genes were observed when DNA samples from 19 commercial varieties of *B. napus* were analysed using the cDNA probes pRR566 (coding for root-specific extensin) and pBOS2 (coding for S_5 self-incompatibility allele in *B. oleraceae*). Both cDNA clones were able to reveal RFLP patterns with varying degrees of polymorphism depending on the restriction enzymes used in the digestion of genomic DNAs. Although both probes could generate complex RFLP band patterns, those revealed by pBOS2 were generally easier to analyse and more suitable for DNA fingerprinting while those revealed by pRR566 were less distinct as a result of extensive background hybridisations. The probe pRR566, with certain restriction enzymes generated simpler RFLP band patterns that were more suitable for segregation analyses. Segregation analysis of F_1 individuals revealed additive RFLP band patterns of both parental varieties, while that of F_2 individuals revealed RFLP band patterns of each parental varieties as well as the additive pattern. When analysed for possible association with varietal glucosinolate content, none of the RFLP band patterns showed such linkage.

A cDNA library was constructed from pod material of a high glucosinolate variety in an attempt to obtain clones which could reveal RFLP patterns associated with glucosinolate content. Differential screening using total cDNAs from pod materials of high and low glucosinolate varieties failed to isolate any cDNA clones useful as RFLP markers.

Another DNA polymorphism assay studied, RAPD, was able to detect inter- and intraspecific variation in *Brassica sp.*. Analysis of six phylogenetically-related but distinct *Brassica sp.* revealed extensive variation in the RAPD band patterns

of amplification products; with some amphidiploid species sharing conserved band patterns with their ancestral species. RAPD analysis on 17 varieties of rape revealed polymorphic as well as highly conserved RAPD band patterns depending on the primer used. One of the primers was able to amplify a polymorphic band which could be associated with low glucosinolate varieties i.e. present almost exclusively in low glucosinolate varieties. Species-specific as well as variety-specific band patterns were also observed during the RAPD analysis.

Finally, sequence and organisation of a *Hind* III family of repetitive sequences was studied. The monomeric and polymeric forms (trimer and tetramer) of the repetitive sequences were successfully cloned into pUC18. Sequence analysis of the two clones containing the polymeric forms revealed that the monomers were arranged in tandem array and that all internal *Hind* III recognition sites were lost due to point mutation(s) which occurred within the six basepair recognition site. A consensus monomeric sequence was deduced from sequence comparison of the 8 copies of the monomeric sequences present in the 3 clones and the deviation from the consensus sequence of each of the eight monomers was less than 3%. No two monomeric sequences were identical. It was estimated that the number of copies of the monomeric sequences in a haploid genome was approximately 0.3 million copies. Estimates of the proportional representation of each of the polymeric sequences based on the number of copies of the monomers in each polymer were also calculated.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitute to my supervisor Dr. R.R.D. Croy for his excellent supervision throughout my study.

I would like to thank the Public Services Department of Malaysia, University of Agriculture, Malaysia, and Malaysian Students Department, London for their financial support. I am especially indebted to Professor Mohd. Mahyudin Mohd. Dahan, Associate Professor Zaliha Christine Alang, and Associate Professor Abdullah Sipat for giving me the opportunity and encouragement to further my study.

I also appreciate the help of Miss Julia Bryden for the DNA sequencing, Mr. John Gilroy for oligonucleotide synthesis, Mr. Dave Hutchinson and Mr. Paul Sidney for photographic work, Dr. Andrew Morby and Mr. Amit Gupta for computer analysis and presentation of DNA sequences, and the rest of the staff members (academic, administrative, as well as technical) at the Department of Biological Sciences, University of Durham for making it possible for me to complete my study.

I wish to thank Dr. Glyn Edwards of Shell Research Limited for supplying the oligonucleotide primers. I also wish to thank the following for generously supplying the seed materials of respective oilseed rape varieties:

1. Danish Plant Breeders, Cheltenham, UK. Varieties – Optima and Willi.

2. Nickerson RPB Ltd., Rothwell, Lincoln, UK. Varieties – Astec, Doublol, Libravo, Lictor, Liradona, Lirawell, Pasha, Score, and Tapidor.

3. Semundo, Cambridge, UK. Variety – Ariana.

4. Svalov AB, S268 00 Svalov, Sweden. Varieties – Global, Karat, and Topas.

5. Twyford Seeds Ltd., Kings Sutton, Banbury, Oxon, UK. Varieties – Bienvenu, Brutor, Cobra, Drakkar, and Rafal.

6. Weibullsholm Plant Breeding Institute, S261 24, Landskrona, Sweden. Varieties – Comet, Elin, and Hanna.

7. Dr. I. Bartkowiak Broda, Instytut Hodowli IAklimatyzacji Roslin, 61-771 Poznan 2, ul Sieroca 1a. Poland. Varieties – Bolko, Bronowski, and Jantar.

8. Dr. D. J. Murphy, John Innes Institute, Colney Lane, Norwich, UK. Varieties – Jet Neuf, and an unnamed double high variety.

Finally, I dedicate this thesis especially to my wife, Faridah Mohd. Noor for her continuous support, sacrifice and patience during the course of the study and also to my two children, Mohd. Haidar and Safia Najwa for giving me their smiles.

LIST OF ABBREVIATIONS

AP-PCR – Arbitrarily primed polymerase chain reaction

ATP – Adenosine triphosphate

bp – basepairs

BSA – Bovine serum albumin

cDNA – Complementary DNA

CIP – Calf intestinal phosphatase

cpm – Counts per minute

DABA – Diaminobenzoic acid

DEPC – Diethyl pyrocarbonate

DMF – Dimethyl formamide

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic acids

dATP – Deoxyadenosine-5'-triphosphate

dCTP – Deoxycytidine-5'-triphosphate

dGTP – Deoxyguanosine-5'-triphosphate

dTTP – Deoxythymidine-5'-triphosphate

DTT – Dithiothreitol

EtBr – Ethidium bromide

EDTA – Ethylene diaminotetraacetic acid

HEPES – N-2-hydroxyethypiperrazine-N'-2-ethanesulfonic acid

IPTG – Isopropyl-2-D-thiogalactoside

kbp – Kilobasepair

kd – Kilodaltons

5-MC – 5-methyl cytosine residue

mRNA – Messenger RNA

MOPS – 3-Morpholinopropanesulfonic acid

MS – Murashige and Skoog

MW – Molecular weight

PCR – Polymerase chain reaction

PEG – Polyethylene glycol

Poly-A^+ – Polyadenylated rich RNA

POPOP – 1,4-*bis* (5-phenyl-2-oxazolyl) benzene

PVP – Polyvinyl pyrolidone

RAPD – Random amplified polymorphic DNA

RFLP – Restriction fragment length polymorphism

RNA – Ribonucleic acid

RNAse A – Ribonuclease A

SDS – Sodium dodecyl sulphate

SLS – Sodium lauryl sarcosinate

SSC – Standard saline citrate

ssDNA – Single-stranded DNA

stDNA – Satellite DNA

TCA – Trichloroacetic acid

uv – Ultraviolet

X-gal – 5-bromo-4-chloro-3-indolyl-2-D-galactoside

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 GENERAL INTRODUCTION

In recent years, oil seed rape (*Brassica napus*) has become increasingly more important as a crop in many parts of the world where intensive farming is practised. In the EEC, Common Agricultural Policy (CAP) has greatly stimulated the production of oilseed rape. As a result, there has been a big decline in the importation of oilseed rape and its products making the EEC virtually self-sufficient in rape oil and meals in 1981. Unfortunately, the contribution of rape oil amounted to only 33 percent in 1981 (39 percent in 1982) of the total vegetable oils required by the EEC (Austin *et al.*, 1986). Today, importation of soybean oil and many tropical oils is still necessary to meet the demand for vegetable oil.

As already shown by Canada, one of the major exporters of unprocessed oilseed rape and its processed products such as oil and meals for livestocks in Europe, it is possible to exploit oilseed rape as a crop to its fullest potential. In the U.K., the total area where oilseed rape is grown rose from 10,000 ha in 1974 to 269,000 ha in 1984 and this is considered as the most rapid rise for any crop in temperate agriculture (Bunting, 1986). The rapid rise, besides the influence of CAP, may have been because rape is also an ideal 'break crop' where farmers can use the same machinery that they already have for cereal crops. A break crop is a one-season alternative crop inserted in the rotation of crops grown to prevent build-up of pathogens and pests and also to re-establish the nutrient balance in the soil. Thus, oilseed rape has become the most important crop that can be harvested by combined harvesters after barley and wheat (Bunting, 1986). By increasing its production, oilseed rape can become the substitute for presently imported proteins and vegetable oils. In the long run, with sound planning and rigorous agronomical research and development, the cost of production of oilseed rape can be lowered and its yield increased, thus becoming a major product for export. Today, agronomists

and plant breeders, especially in the EEC, have made oilseed rape research one of their top priorities in an effort to make it more and more commercially viable.

The improvement of the yield, resistance to diseases and pests, fatty acid composition, and lowered erucic acid and glucosinolate contents can be achieved by breeders carrying out selection experiments and exploiting the variability in these already-established characters. The improvement of oil quality has received a lot of attention from breeders. Selection for lower linolenic acid content is desirable as linolenic acid is very unstable and is easily oxidised to give unpleasant-smelling substances in both margarine and cooking oil made with rape oil. Moreover, breeding varieties with a reduced proportion of the C-18 fatty acids (Oleic acid (18:1) and linoleic acid (18:2)) and an increased proportion of palmitic acid (16:1) may overcome the problem of margarine recrystallisation during storage. Erucic acid is undesirable because margarine made from high erucic acid rape oil will not spread when taken from a household refrigerator. Furthermore, erucic acid is deposited in the heart muscles of experimental animals, and consequently is a potential health hazard to man. Another very important consideration for breeders is the selection of varieties with lower glucosinolate content. Glucosinolates are sulphur-containing substances which when broken down, give bitter-tasting, toxic, and goitrogenic compounds in rape meal (Bunting, 1986). Breeders select desired characteristics by classical plant genetics using established morphological markers or by direct chemical analyses. The use of biochemical markers, isozyme polymorphisms or isozyme markers can help in such selections where certain characters can be linked to a particular polymorphism observed. The information on these biological and biochemical markers is normally contained in a classical genetic map constructed from many years of laborious plant breeding. One of the disadvantages of such a genetic map is that the markers selected for the analysis are products of transcription and translation of genes and as such their expressions can be influenced by environmental factors such as temperature, light, soil conditions and so on. Moreover, breeders have to wait for the expression of desired phenotypic characters before they can be visually analysed. For example, morphological markers of mature seeds such as round/wrinkle, green/white, etc. can only be analysed after the seeds have matured. So, an alternative marker which can be rapidly analysed for, and is independent of environmental influence is highly desirable.

Recently, restriction fragment length polymorphisms (RFLPs) have been increasingly utilised in plant breeding to generate another form of genetic markers. RFLPs can be defined as differences observed in the size of fragments resulting from the digestion of DNA with restriction endonucleases. Sequence changes in genomic DNA can lead to the creation or loss of particular cleavage sites, thus altering the size of fragments generated from a particular region of the DNA when digested with a restriction endonuclease. There are many direct applications for RFLP technology to help breeders in their plant breeding programmes. For example, RFLPs can provide a clear and precise analysis of the genotype of a crop species and thus establish genetic linkage to the characters to be selected. Other uses of RFLPs include developing molecular linkage maps, organising germplasms, protecting lines, varieties, and hybrids for commercial purposes, introgressing important traits from exotic germplasm sources, and aiding breeders in recovering recurrent-parent genotypes in backcross breeding programmes (Walton, 1988). One important advantage of RFLP maps is that since the variation is detected directly at the DNA level, alleles behave in a codominant manner and normally have no visible effect on the phenotype (McCough *et al.*, 1988).

RFLP technology has already been applied in many crops including tomato (Bernatzky and Tanksley, 1986a; Nienhuis *et al.*, 1987; Young and Tanksley, 1989), maize (Helentjaris, 1987; Burr *et al.*, 1988), lettuce (Landry *et al.*, 1987a, 1987b, 1987c), potato (Bonierbale *et al.*, 1988; Rivard *et al.*, 1989; Gebhardt *et al.*, 1989), lentils (Havey and Muehlbauer, 1989) pea (Lee *et al.*, 1988), wheat (May and Appels, 1987; Sharp *et al.*, 1989; Dvorak *et al.*, 1988; Kam-Morgan *et al.*, 1989), barley (Saaghai-Marrof *et al.*, 1984), soybean (Apuya *et al.*, 1988; Keim *et al.*, 1989; Close *et al.*, 1989), and *Brassica sp.*, (Song *et al.*, 1990, 1988a). RFLPs have also been applied in the study of *Arabidopsis* genome, a non-crop plant extensively studied as a plant molecular biology tool due to its small genome and low frequency of repetitive sequences (Chang *et al.*, 1988; Nam *et al.*, 1989).

Another method for analysing DNA polymorphisms was recently described independently by Williams *et al.*, (1990) and Welsh and McClelland (1990). This method is called Random Amplified Polymorphic DNA (RAPD) (Williams *et al.*, 1990) or Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) (Welsh and McClelland, 1990) which henceforth will be referred to as RAPD only. This

method was based on the amplification using PCR, of random DNA segments with single oligonucleotide primers of arbitrary nucleotide sequence. The polymorphisms observed when genomic DNAs were used as templates were simply detected as sequences or fragments of DNA some of which amplify from one genetic line but not from another as visualised by electrophoresis on 1.4% (w/v) agarose gels stained with ethidium bromide. The pattern of polymorphic bands or single unique bands generated can serve as genetic markers as they follow Mendelian inheritance, and thus can be used to generate genetic maps in a wide variety of species by analysing segregating F_2 populations. The utility of RAPD technology in revealing consistent and reproducible DNA polymorphisms has been shown in many species including man, soybean, maize, *Neurospora crassa*, and bacteria (Williams *et al.*, 1990). Williams *et al.* (1990) have also successfully demonstrated the Mendelian genetic segregation of the RAPD markers and established their map positions in the molecular linkage map and confirmed the positions using closely linked RFLP markers in relation to established genetic markers in soybean, maize and *Neurospora crassa.* In another study, Welsh and McClelland (1990) were able to demonstrate the utility of RAPD technology in generating simple, reproducible fingerprints of complex genomes of 24 strains of *Staphylococcus sp.*, 11 strains of *Streptococcus sp.* and 3 varieties of rice (*Oryza sativa*). In their latest work, Welsh and McClelland (1991) have further demonstrated that genomic fingerprints can also be generated in human and maize cultivars which can help distinguish between species and genera.

Genetic mapping using RAPD markers has several advantages over other methods: (a) a universal set of primers can be used for genomic analysis in a wide variety of species, (b) no preliminary work to isolate cloned probes nor in the preparation of filters for hybridisations is required, (c) greatly simplified information transfer in collaborative works – only sequences of primers need be communicated instead of exchange of the actual clones.

The use of DNA polymorphism assays such as RFLP and RAPD has led to a greater understanding of the complexity and variability of genome architecture of plant and animal species. In recent years, the understanding of genome architecture in plants has been facilitated by the use of genome fingerprinting, molecular linkage maps based on RFLPs and to a lesser extent on RAPD technologies.

DNA polymorphisms revealed by these methods can be brought about by point mutations or larger scale changes ("macromutations") as a result of chromosome inversions, translocations, deletions or transpositions of DNA fragments. These major changes can, in part, be brought about by repetitive sequences that are often present in large quantities (greater than 75% in species with genomic DNA content above 2 pg (Flavell *et al.*, 1982). Flavell *et al.* (1982) contended that repetitive sequences not only create much of the architecture of the plant genome but may also predispose it to instability and changes during evolution. The relationship between DNA polymorphism as detected by RFLP and the structure and arrangements of repetitive sequences in plants, so far, remains unknown.

1.2 AIMS OF THE RESEARCH PROJECT

It can therefore be justified that the application of RFLP and RAPD technologies to oilseed rape breeding programmes should be explored. There are many commercial varieties of oilseed rape (See list in Results and Discussion, Section 3.1.1) available to farmers throughout Europe, where each variety is classified as single low ('0', seeds are low in erucic acid, high in glucosinolate content and double low ('00', seeds are low in erucic acid, low in glucosinolate content). The ability to screen varieties for these characters and for other desirable features conveniently and rapidly (i.e. using RFLP and RAPD technologies) is likely to be extremely valuable to Oilseed Rape Breeders where characters can be linked to RFLP and RAPD patterns. Moreover, genomic fingerprints realised by RFLP and RAPD technology will be extremely useful for cultivar identification and commercial patent rights. The structure of oilseed rape repetitive sequences and their contribution to DNA polymorphisms would also be a useful study. The research programme can be divided into 4 sections as described below:

1.2.1 A Survey of RFLPs in oilseed rape varieties

The application of RFLP technology has not been utilised in *Brassica napus* especially in relation to commercial varieties. Feasibility study was therefore initiated to determine the utility of RFLP technology in rape. The most important step towards realising RFLPs in genomic DNA samples is the selection of combinationsof probes and restriction enzymes to be used in the analyses. In this study, a

number of characterised rape cDNA clones were used as probes for genomic DNA samples from several varieties of oilseed rape digested with different restriction enzymes. The objective of this study was to identify combinations of cloned rape DNA sequences and restriction enzyme digests of genomic DNA which can reveal RFLPs and for subsequent use in segregation studies. At the time of starting this research programme, no RFLP study was reported in oilseed rape.

1.2.2 RFLP markers and glucosinolate content

One of the genetic characters of great economic importance in oilseed rape breeding programmes is the glucosinolate content of the seeds. Breeding programmes for double low rape varieties have intensified in recent years and the low glucosinolate progenies have been selected by determining the glucosinolate content of the seeds using labour-intensive extractions and HPLC analysis (Heaney *et al.*, 1986). The use of RFLP markers linked to glucosinolate content in seeds would therefore facilitate the selection process. The objective of this study was to construct a cDNA library and differentially screen the library to find suitable cDNA sequences which may be used as RFLP markers linked to glucosinolate content. At the start of this research programme, little or no information was available on site and mechanism of glucosinolate metabolism.

1.2.3 *Brassica napus* Repetitive Sequences

During the initial studies involving digestion and analysis of oilseed rape genomic DNAs using *Hind* III, a "step-ladder" pattern of bands, characteristic of repetitive DNAs, was observed in the EtBr-stained agarose gel after electrophoresis. As discussed earlier in the introduction, repetitive sequences can contribute to genomic DNA variability (Flavell *et al.*, 1982) which may then be detected as RFLPs. Apart from academic interest in the nature of these highly abundant sequences, the question of whether highly repetitive sequences could be used as probes for RFLPs should be answered. So far, the use of highly repetitive sequences as probe for RFLP studies has not been explored. Further information on the structure of repetitive sequences in *Brassica napus* especially in the sequences and arrangements should also be obtained. The objective of this part of the study was to explore the feasibility of using highly repetitive DNA as a

probe for RFLPs and to obtain information about the structure of the repetitive sequences in *Brassica napus.*

1.2.4 RAPD Analysis in Related *Brassica sp.*

The development of RAPD analysis has shown many advantages over RFLP analysis in the detection of DNA polymorphisms; its ability to detect DNA polymorphisms in many species indicates that it should be possible to study *Brassica napus* with this technique. The success of RAPD analysis and its ability to detect DNA polymorphisms is dependent on the choice of oligonucleotide used as primers in the amplification reactions. Several oligonucleotides are employed in this technique, based on the sequences described by Williams *et al.* (1990). The objective in this part of the programme was to test the feasibility of using RAPD technology to detect interspecific DNA polymorphisms by analysing genomic DNA samples from six phylogenetically-related *Brassica sp.* and secondly, to use RAPD to detect varietal differences in oilseed rape and any possible associations with their glucosinolate contents.

1.3 REVIEW OF LITERATURE

1.3.1 Oilseed rape (*Brassica napus)* – Background

There are four Brassica oilseed species grown extensively throughout the world; namely: *Brassica carinata, Brasica juncea, Brassica napus,* and *Brassica campestris.* Cytological relationships between these species and their close allies were first established by U (1935), who stated that *Brassica nigra* (black mustard; 2n = 16, bb), *Brassica oleracae* (cabbages and kales; 2n = 18, cc) and *Brassica campestris* (turnip rape; 2n = 20, aa) are the primary species while *Brassica carinata* (Ethiopian mustard, 2n = 34, bbcc), *Brassica juncea* (Oriental mustard, 2n = 36, aabb) and *Brassica napus* (oilseed rape, 2n = 38, aacc) are the amphidiploids resulting from crosses between the primary species (2n represents the number of chromosomes in a diploid while the designation 'aa', 'bb', 'cc', 'aabb', and so on represent the genotype of a particular species). Synthesis of these amphidiploids provided proof of their close cytological relationships and the recent advances in genetic manipulation techniques can facilitate gene transfer (e.g. somatic hybrids) among *Brassica sp.* as means of oilseed rape improvement (Bunting, 1986).

B. napus is the primary Brassica oilseed species (*B. campestris* (turnip rape) being the other species) grown in Europe and probably accounts for about 95% of current production (Bunting, 1986). There are spring-sown and autumn-sown varieties of these two species, thus giving rise to four crop types; namely: winter rape (autumn-sown *B. napus*), spring rape (spring-sown *B. napus*), winter turnip rape (autumn-sown *B. campestris*), and spring turnip rape (spring-sown *B. campestris*). Bunting (1986) also stated that, in general, autumn-sown varieties (winter rape) have a potential yield about 20 percent higher than spring-sown varieties of the same species. Furthermore, *B. napus* outyields *B. campestris*, which is widely grown in Sweden, by about the same amount, but *B. campestris* is earlier maturing and the winter varieties are less succeptible to winter kill (Bunting, 1986).

There are several important breeding objectives to improve the commercial value of oilseed rape as a crop. Thompson and Hughes (1986) have discussed the subject in great detail and suggest the following:

1. Good resistance to stem canker (*Leptosphaeria maculans*), the most widespread and damaging disease, and to a lesser extent, resistance to leaf spot (*Pyrenopeziza brassicae*).

2. Low erucic acid content (less than 0.5% in the oil) and low glucosinolate content with good winter hardiness. The desirability of lower erucic acid and glucosinolate contents has been discussed in the introduction.

3. Higher yield than current varieties unless the new varieties can include other desirable characteristics such as yellow seed for better oil colour or short straw for easy husbandry and harvesting. Selection must also be made for large, plump seeds as this type produces a more balanced ratio of oil to rape meal.

4. Better oil and rape meal qualities. Selection for lower linolenic acid content gives good frying oils. On the other hand, lower linoleic and linolenic acids but higher palmitic acids will prevent margarines or shortenings from crystallising during storage. Low glucosinolate rape meal means a greater portion can be put in cattle diets and possibly in poultry feeds if the glucosinolate content is low enough.

Austin *et al.* (1986) has also suggested other characters for improvement including the following:

1. Selection for drought resistance. Winter rape often experiences drought after sowing or drought during seed development in summer.

2. Improved uniformity of light distribution within the canopy. Plants with reduced petal size or apetalous lines and more erect pods should be bred. Since petals absorb up to 60% of solar radiation, reduced petal size or apetaly will result in less interception of solar radiation during flowering. This character, combined with more erect pods would permit light to penetrate deeper into the lower pod canopy, increasing photosynthesis in these pods and consequently reducing seed and pod abortion (Thompson and Hughes, (1986).

3. Plants with thicker stems and well-developed root systems should be bred for resistance to lodging during heavy rains and strong winds.

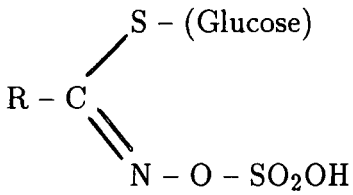
4. Resistance to shattering of pods and loss of seeds. Brown mustard (*B. juncea* which is resistant to shattering could be a source of such resistance.

5. Resistance to selective herbicides. Oilseed rape is a dicot, like many other weed species, and is therefore susceptible to selective herbicides (atrazine-based) used in cereal crop growing. Resistance would allow the use of such herbicides for oilseed rape crops.

1.3.2 Glucosinolates in Oilseed Rape

Low glucosinolate content is now a major breeding objective for most winter rape breeders. Glucosinolates, found in all tissues of rape, are sulphur-containing substances that are broken down by the enzyme myrosinase. Glucosinolates in rape seeds, upon the breakdown by the enzyme, gave bitter-tasting, toxic and goitrogenic compounds in rape meal (Bunting,1986).

The generic structure of glucosinolate is shown below:



There are at least two classes of these compounds, one in which R is an aliphatic chain and the other in which it is a substituted indole (Austin *et al.*, 1986). These compounds when present in rape meal used for animal feed will reduce its palatibility and hence restricts the animals' food intake and rate of growth (Bowland *et al.*, 1965). Moreover, there is a health hazard factor involved as oxazolidinethione by-products of glucosinolate breakdown have been shown to inhibit the function of the thyroid gland (Greer, 1950).

Compared with soybean meal and many other seed meals, rape seed meal has a better balanced amino acid composition (Austin *et al.*, 1986). Consequently, it can be added in a greater proportion to animal feeds. The inclusion of rape meal in cattle diets could be doubled using low-glusinolate meal, but is still not suitable for feeding to laying poultry as a fishy taint can develop in the eggs (Bunting, 1986).

Thus, breeding for low glucosinolate oilseed rape is highly desirable. The Polish spring variety Bronowski contains only one-sixth of the normal content of glucosinolate and has been used as the only major genetic source from which to breed low glucosinolate varieties (Bunting, 1986). Glucosinolate content is determined by at least three partially recessive genes, which are maternally inherited and the products which are present in the seeds are apparently mainly synthesised in the pods (Kondra and Stefansson, 1970). So far, the pathway for the biosynthesis of glucosinolates is not fully understood and extensive research is now underway to try to understand the pathway , localization and enzymes controlling the production of the final glucosinolate products. Figure 1.1 shows the proposed partial pathway for the biosynthesis of glucosinolates. Currently, varieties with

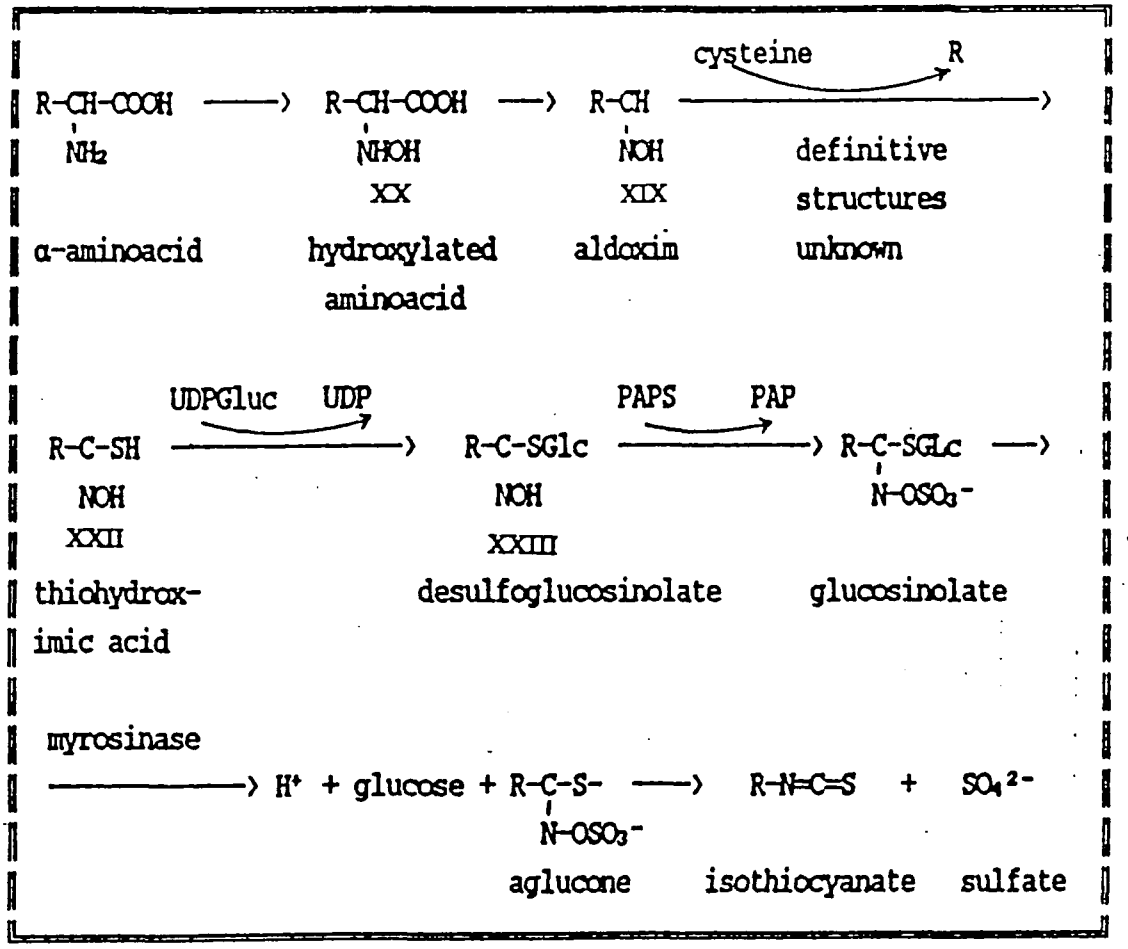
low erucic acid (less than 0.5 percent in oil) and low glucosinolate content (20 to 30 μ moles glucosinolates per g defatted meal) ('double zero') are commercially available. However, it is argued that their yield and disease resistance is inferior to the high glucosinolate varieties (Bunting, 1986). From 1989, due to new legislation from the EEC, farmers throughout Europe are now banned from growing high glucosinolate varieties. So, in conclusion, the breeding of varieties having low seed glucosinolate content but still maintaining high level of leaf glucosinolates is therefore a highly desirable objective.

1.3.2.1 Glucosinolate Biosynthesis and Control in Brassica

The biosynthesis of glucosinolate is so far not fully understood. A generally accepted pathway of glucosinolate biosynthesis where all the glucosinolates are presumably derived from amino acids in a series of metabolic steps, is outlined in Figure 1.1. Aldoximes (XIX) are the most important intermediates which give rise to not only glucosinolates but also the cyanogenic glucosides. The intermediates before the synthesis of the aldoximes are not well characterised and understood. Aldoximes are then transformed into glucosinolates via thiohydroximic acids (XXII) and desulfoglucosinolates (XXIII). The intermediates between the aldoximes and the thiohydroximic acid have not been identified but it is generally accepted that the sulfur in cysteine is readily incorporated into the thiohydroximic acid (1-thioglucose has been excluded as a precursor) (Larsen, 1981). Desulfoglucosinolates are the last intermediates in the biosynthetic pathway and act as a precursor for many types of individual glucosinolates in plants. An enzyme, UDP-glucose:thiohydroximate glucotransferase, catalyses the formation of desulfobenzyl glucosinolate by glucosyl transfer from UDPG to phenylacetothiohydroximic acid (Larsen, 1981).

Not surprisingly, very little is known about the control of glucosinolate biosynthesis in plants. Larsen (1981) proposed that some sort of control is exerted in the first step (i.e. the production of aldoximes), the activation (probably N-hydroxylation) of the amino acids. This was later confirmed by Schnug (1990). Extensive breeding programmes have been carried out in oilseed rape to produce low glucosinolate varieties. As mentioned earlier, variety Bronowski is the only source of low glucosinolate genes. By using a number of labelled

Figure 1.1


cysteine
R
R-CH-COOH
NH2
α-aminoacid
R-CH-COOH
NHOH
XX
hydroxylated aminoacid
R-CH
NOH
XIX
aldoxim
definitive structures unknown
UDPGluc
UDP
PAPS
PAP
R-C-SH
NOH
XXII
thiohydroximic acid
R-C-SGlc
NOH
XXIII
desulfoglucosinolate
R-C-SGLc
N-OSO3-
glucosinolate
myrosinase
H+ + glucose + R-C-S-
N-OSO3-
aglucone
R-N=C=S
isothiocyanate
SO4 2-
sulfate

precursors including methyl thiopentanaldoximes and desulfo-3- butenylglucosinolate, it was concluded that 2-hydroxy-3- butenylglucosinolate was derived from 3-butenylglucosinolate in Regina II (high glucosinolate variety) and that metabolic blocks occurred both in a step prior to the oxime and in the hydroxylation in 'Bronowski' (Josefsson, 1971, 1973; Underhill, 1990)

It was suggested that the glucosinolates which are eventually found in the seeds are apparently synthesised in the pods and that there are at least three partially recessive, maternally inherited genes controlling the synthesis (Lein, 1972a, 1972b; Austin *et al,* 1986).

1.3.3 Recent Molecular Biological Studies on *Brassica napus* L.

Due to its potential economic importance, research on *Brassica napus* L. has received much recent attention especially in the field of molecular biology.

Perhaps, the earliest and most extensive research programmes done on this species is the study of its seed storage proteins. Bhatty *et al.*, (1968) first described the water-soluble rapeseed proteins which were later referred to as 1.7S storage proteins. These basic storage proteins with isoelectric points close to 11.0, were then referred to as napins by Crouch *et al.* (1983). Synthesised in developing embryos during seed maturity, napins have an average molecular weight of 13 kd and are composed of a large and a small subunit (9 kd and 4 kd, respectively) linked by disulphide bonds (Scofield and Crouch, 1987). At seed maturity, napins constitute about 20% of total seed protein and are broken down rapidly during germination (Crouch and Sussex, 1981). Later studies on napins dealt mainly with their structure and precursor polypeptides using cloned cDNAs (pN1 and pN2, and gNa) coding for both subunits (Ericson *et al.*, 1986; and Scofield and Crouch, 1987, respectively).

Another type of rape storage protein, called cruciferin, a legumin-like 12S storage protein is a large, neutral, oligomeric protein synthesised during seed development (Rodin *et al.*, 1990). Cruciferin is the major seed storage protein and is composed of six subunit pairs where each subunit pair is made up of 30 kd α polypeptide and 20 kd β polypeptide, and was cleaved post-translationally from a 50 kd precursor (Simon *et al.*, 1985; Rodin *et al.*, 1990). Earlier work on cruciferin

included its isolation and characterisation as a 'neutral' protein (Schwenke *et al.*, 1983). Simon *et al.*, 1985 successfully cloned one of the members of the cruciferin family from embryo cDNA (designated as pC1) which codes for a precursor of cruciferin subunits $\alpha 2/3$, $\beta 2$, and $\beta 3$. Simon *et al.* (1985) also reported about 40% amino acid sequence homology between cruciferin and legumin from pea. Another cruciferin clone from an embryo cDNA library has recently been isolated (designated pCRU1) by Rodin and co-workers (1990) which encodes a precursor that contains the $\alpha 1 - \beta 1$ subunit of cruciferin. The nucleotide sequences, as well as the amino acid sequences of the two cDNA clones, have about 64% homology (Rodin *et al.*, 1990). At maturity, cruciferin constitutes about 60% of the total seed protein (Crouch and Sussex, 1981).

mRNA levels for both cruciferin and napin during seed development have also been studied. DeLisle and Crouch (1989) reported that mRNA levels of cruciferin and napin are about 11% and 8% respectively of the total embryo mRNA during cell expansion phase.

Another type of rape seed protein that is currently being studied is a class of amphipathic proteins associated with the lipid storage bodies. Oil is stored in the rape seed inside compartments enclosed by a membrane. The proteinaceous components of the lipid storage bodies are called oleosins and are not associated with any other cellular structures (Murphy *et al.*, 1991). The major rapeseed oleosin was found to be made up of at least two 19 kDa polypeptides designated as nap-I and nap-II, which have closely related but different amino acid sequences (Murphy *et al.*, 1990). Murphy and co-workers (1991) also concluded that oleosins exhibit structural and functional similarities with some animal serum apolipoproteins.

At Durham University, as part of a programme to isolate plant genes expressed in an organ-specific manner, a family of cross-hybridising cDNA sequences highly expressed in rape roots have been isolated, characterised, and shown to encode proteins showing homology to carrot and tomato extensin (Evans *et al.*, 1990). By sreening a rape genomic library with the cDNAs, an extensin gene, designated as *extA*, has been isolated. The gene is a member of a multigene family, consisting of about 3 members per haploid genome with strong homology, and 20 or so members with weaker homology to the cDNA probe used (Evans *et al.*, 1990). Expression

studies indicated that *extA* was only expressed in root and was transcribed into an mRNA species of about 1.3 kb in size. Another extensin gene with similar characteristics has recently been isolated and designated as *extB* (Gatehouse *et al.*, 1990). This extensin gene, *extB*, represents a sub-family of extensin genes different from the previously described subfamily of *extA* in that it encodes a polypeptide relatively rich in tyrosine (Gatehouse *et al.*, 1990).

The resynthesis of the allotetraploid *Brassica napus* from its progenitors, *B. campestris* and *B. oleraceae* is regarded as a powerful tool for the introduction of new genetic variations into this crop for economic improvement. The traditional sexual methods between the two parents (U, 1935, Akbar, 1989) has now been replaced by more modern methods. One such method is somatic hybridisation which protoplasts from the two parents, are fused together with the help of Polyethylene glycol-4000 or electroporation (Akbar, 1989 and Rosen *et al.*, 1988). Plants are regenerated and screened for true hybrids based on available morphological, chromosomal, and molecular data. Another method of resynthesis is through embryo rescue where embryos, developed as a result of interspecific crosses, were "rescued" during development, cultured and regenerated into normal plants. Colchicine treatment has sometimes been employed to generate dihaploid plants (Chen *et al.*, 1988 and Akbar, 1989).

Plant transformation is also another powerful method for the introduction of genetic materials of agronomic importance. At Durham University, work has been carried out on the transformation of *Brassica napus.* Boulter *et al.*, (1990) produced transgenic rape plants using two different methods. In the first method, an *Agrobacterium tumefaciens* binary system based on Bin 19 was used to transform inflorescence stalks while in the second method, an *Agrobacterium rhizogenes* binary system comprising of pRi1855 and pBin19 plasmids was used to induce co-transformed hairy root formation at the cotyledonary nodes.

1.3.4 RFLP and Its Application in Molecular Genetics

With the extensive use of restriction endonucleases in molecular genetics especially in the area of mapping gene sequences and for studying gene function, and organisation, molecular geneticists have discovered widespread polymorphisms in the locations of restriction enzyme sites amongst individuals of a species. RFLPs

can be defined as the differences observed between genotypes in the fragment lengths of restriction endonuclease-digested DNA (Nienhuis *et al.*, 1987). RFLPs occur as a result of base pair or positional changes in the restriction enzyme recognition sites which flank the chromosomal location of a distinct DNA sequence (eg. a gene)and can be detected by hybridisation with a radioactive-labelled DNA containing sequences homologous to that of the DNA sequence. Burr *et al.* (1983) and Botstein (1980) however argued that alterations producing RFLPs can also occur within a gene. Botstein (1980) suggested that many kinds of genotypic differences are products of either (1) the difference of one or more individual bases resulting in loss of a restriction cleavage site or formation of a new one, or (2) the insertion or deletion of blocks of DNA within a restriction fragment, thus altering its size. According to Burr *et al.* (1983), insertion/deletion polymorphisms can be distinguished from base pair substitution polymorphism by the pleiotropic effect they have when DNA is cleaved with two or more enzymes. Multiple enzymes may show the same difference in the size of restriction fragments from two alleles in the case of insertions or deletions while base pair changes will usually only affect the ability of one enzyme to detect the polymorphism.

Hybridisation using a unique cloned sequence permits the identification of a specific locus. The differences in the restriction fragment lengths revealed after agarose gel electrophoresis, Southern blotting and hybridisation, therefore function as alleles of that locus. Consequently, RFLPs can serve as genetic markers comparable to conventional morphological or biochemical (isozyme) markers except that they are not the product of transcription and translation (Nienhuis *et al.*, 1987).

RFLPs were first used as a tool for genetic analysis in 1974 where the linkage of temperature-sensitive mutations of adenovirus to specific restriction fragment length differences was used to locate the mutations on a physical map of restriction fragments (Grodziker *et al.*, 1974). Botstein *et al.* (1980) were the first researchers to propose the use of RFLPs as genetic markers in human genetics and to use them for the construction of a genetic linkage map. RFLPs have also been used to establish linkages with loci determining many genetic diseases (See review by Gusella, 1986). Other uses of RFLPs in humans include prenatal diagnosis, detection of abnormal mitotic and meiotic divisions, paternity testing, tissue typing for transplantation, and in forensic sciences (Gusella, 1986).

Other species to which RFLPs have been applied include yeast (Petes and Botstein, 1977, Goodman *et al.*, 1977, Cameron *et al.*, 1978) nematodes (Botstein *et al.*, 1980); and in fungi (Kustersvansomeren *et al.*, 1991; Manicom *et al.*, 1987). In fungi, RFLPs have been used in the taxonomy and identification of Aspergilli and Fusarium (Kustersvansomeren *et al.*, 1991; Manicom *et al.*, 1987; respectively).

The successful use of RFLPs, particularly in human genetics, has recently been extended to plant genetics. Its applicability to plant genetics and breeding has been examined and reviewed by several authors (Burr *et al.*, 1983; Beckman and Soller, 1983; Soller and Beckman, 1983; Tanksley, 1983; and Helentjaris *et al.*, 1985). Earlier, probably due to unfamiliarity with the technology, plant geneticists were reluctant to investigate deeper into the potential use of RFLPs for plant breeding. Nowadays, RFLPs has been extensively used in the construction of genetic linkage maps and due to its ability to saturate the linkage map with a large number of mapped positions, the power of RFLP mapping has been widely accepted in the scientific communities. A commercial company (CERES, NPI, 417 Wakara Way, Salt Lake City, UT84108, USA) specialising in the identification and use of RFLPs for plant breeding, has also been established.

The extent of RFLPs brought about by genetic variability has been reported in many plant species as mentioned in the introduction. Genetic studies on selection, identification, and organisation of plant genomes require a detailed genetic linkage map. RFLPs can serve as genetic markers upon which linkage maps can be based. Each change that alters the length of a restriction fragment creates a new allele. To define a 'locus' one only requires a DNA probe which is related or linked to known genes or randomly chosen from a population of clones either in genomic or cDNA libraries (Burr *et al.*, 1983). Botstein *et al.* (1980) further contended that linkage of an arbitrary marker would be just as informative in studying inheritance as would linkage to a morphological or biochemical trait. Subsequent pedigree analyses confirmed that these markers would be inherited as simple Mendelian codominants and could be especially useful as indirect selection criteria for traits with low heritabilities which are difficult or expensive to evaluate phenotypically. Burr *et al.* (1983), for example, found in maize that nearly all random leaf cDNA clones (18 clones) used were able to detect polymorphisms and each cDNA clone

(16 of 18 clones) specified at least two alleles among them. Three other randomly chosen cDNA clones that detect polymorphisms between the parents used in the same cross also showed Mendelian inheritance. It was found that in maize the degree of polymorphisms (i.e. the number) was higher than that demonstrated in humans (Burr *et al.*, 1983; and Rivin *et al.*, 1983). Two studies carried out by Helentjaris (1987) and also Helentjaris *et al.* (1986a) on the construction of a genetic linkage map for maize reported that many of the maize supposedly single copy clones detected more than one fragment. Further studies of these single-copy clones lead to the identification of the genomic locations of duplicate DNA sequences (Helentjaris *et al.* (1988). This is unique to maize where more than one single-copy DNA sequences (multiple gene locations) are present within the maize genome as a result of internal duplication (allopolyploidy) and not a product of hybridisation between two individuals with similar genomic structures (allotetraploidy) (Helentjaris *et al.*, 1988). Moreover, the duplication of low copy number sequences is over fourteen times more numerous in maize as in tomato (Helentjaris *et al.*, 1988). This accounts for the higher degree of polymorphism found in maize.

In other plant species, the degree of intraspecific genetic variability was considerably lower than that shown in maize especially those with a narrow genetic base resulting from agricultural domestication. Helentjaris *et al.* (1986a) and Bernatzky and Tanksley (1986a and 1986b)) confirmed that it was necessary to analyse interspecific (between species) crosses in *Lycopersicon spp.* (tomato) to find sufficient polymorphisms to construct a genetic map, as clones showing polymorphism were rarely identified between cultivated varieties of tomato. Moreover, Helentjaris *et al.* (1985) pointed out that those plant species which are primarily outcrossing in nature yield a higher degree of polymorphism when analysed for RFLPs as compared to the much lower degree of polymorphism exhibited by species that are primarily self-pollinated. Interspecific crosses have also been utilised to study RFLP segregation in lentils (Havey and Muehlbauer,1989a). It is however argued that it is advisable to generate detailed genetic maps using intraspecific crosses if sufficient numbers of markers are detectable (Havey and Muehlbauer, 1989a). Differential fertility, viability or environmental effects complicate the mapping of Quantitative Trait Loci (QTLs) linked to markers showing significant deviations from the expected ratios if interspecific crosses were used in the study

(Havey and Muehlbauer, 1989a and 1989b). QTLs can be defined as one or more segregating loci controlling part of a quantitative character (Geldermann, 1975). Jensen (1989) described the method called maximum likelihood to estimate the recombination frequencies between marker genes and a QTL, as well as the effects and segregation ratio of the linked QTL.

Similar studies on the construction of genetic linkage maps have been carried out in pea, soybean, lettuce, tomato, potato, rice, wheat, barley, lentils, and *Arabidopsis* (Lee *et al.*, 1988; Apuya *et al.*, 1988; Landry *et al.*, 1987a and 1987b; Bernatzky and Tanksley, 1986a and 1986b; Bonierbale *et al.*, 1988; McCough *et al.*, 1988; May and Appels, 1987; Saaghai-Marrof *et al.*, 1984; Havey and Muehlbauer, 1989a, and Nam *et al.*, 1989; respectively).

As has been discussed earlier, the major contribution of RFLP technology is in the construction of genetic linkage maps of plant species. Studies which utilised RFLP markers that can be linked to certain important aspects in agriculture are, however, very limited. RFLP technology can help plant breeders in solving some of their problems, for examples RFLPs have been used in the selection of disease and insect resistance in certain crops. In lettuce (*Lactucca sativa*, Landry *et al.* (1987a) reported RFLP analyses on loci associated with downy mildew resistance genes. In tomato, (*lycopersicon sp.*), Nienhuis *et al.* (1987) reported that RFLP loci on three different linkage groups were found to be correlated with the expression of colorimetric absorbance of 2-tridecanone (2-TD), the principal toxic factor involved in insect resistance in tomato. Nienhuis *et al.* (1987) analysed the colorimetric absorbance of 2-TD of parental, F_1 and F_2 individuals of a cross between a wild, insect resistant tomato species *L. hirsutum* and a domesticated, insect sensitive cultivar, Manapal (*L. esculentum*), as female parent and observed that the average absorbance value of the segregating F_2 individuals were 135 for homozygous (esculentum), and 406 and 588 for heterozygous and homozygous (hirsutum) respectively. These average values were calculated from colorimetric absorbance values of F_2 individuals showing polymorphisms at two RFLP loci on each of the three linkage groups. In another study, Young and Tanksley (1989) used RFLP markers linked to the introgressed segments of DNA flanking the *Tm-2* gene which confers resistance to tobacco mosaic virus (TMV) to monitor the gene's introgression into tomato during backcross breeding. In studies not related to disease and

insect resistance, Martin *et al.* (1989) used RFLP markers linked to genes associated with water-use efficiency in tomato while Tanksley and Hewitt (1988) and Osborne *et al.* (1987) used molecular markers linked to genes controlling content of soluble solids in tomato fruit.

In pea, RFLPs have been utilised to study the organisation, inheritance, and mapping of legumin genes (pea storage protein) (Domoney *et al.,* 1986); *rbcS* (small subunit of chloroplast protein, ribulose-1,5-biphosphate carboxylase) and *cab* (Chlorophyll a/b binding protein) multigene families (Polans *et al.,* 1985); and vicilin (*Vc-5*, pea storage protein) and r_b locus (Lee *et al.,* 1988). Ellis *et al.* (1984) studied the organisation and genetics of ribosomal RNA gene (rDNA) length variations in pea using RFLPs.

Other uses of RFLPs include: studies on phylogenetic relationships in *Brassica sp.* (Song *et al.*, 1990), wheat (Sharp *et al.*, 1989), and lentil (Havey and Muehlbauer, 1989b); estimating the degree of heterosis amongst maize inbred lines (Lee *et al.*, 1989); detection of alien chromosomes in wheat (Sharp *et al.*, 1989); analysis of linkage drag in backcross progenies in tomato (Young and Tanksley, 1989) and studying genetic variation among plants generated from *in vitro* anther cultures of *Solanum chacoense* (Rivard *et al.*, (1989).

1.3.5 Construction of RFLP Genetic Linkage Maps

The construction of an RFLP genetic linkage map is achieved by analysing segregating F_2 progenies resulted from either intraspecific crosses or interspecific crosses using selected probes. There are many methods of analysing the segregation data to determine linkages among the RFLP markers realised by the probes used and eventually organised for the construction of RFLP genetic linkage maps. Almost all of the methods used for this purpose were carried out with the help of computer program developed by several workers (Landers *et al.* (1987), Landers and Green (1987), Suiter *et al.* (1983), Ritter (*in* Gerbhardt *et al.*, 1989) and Lee *et al.* (1988) based on the maximum likelihood formulae for the analysis of linkages proposed by Ott (1985). Among these computer programs the two most commonly used programs are the MAPMAKER (Landers *et al.* (1987), Landers and Green (1987)) and LINKAGE-1 (Suiter *et al.*, 1983). One of these two programs the MAPMAKER, is hereby described. Initially, the RFLP

markers were placed into putative linkage groups based on the lod (likelihood) scores (Ott, 1985) generated from two-point linkage analysis i.e. whether the two markers are linked or unlinked. Lod score is expressed as the Log_{10} of the ratio of the probability that the markers are linked divided by the probability that the markers are unlinked, and is a measure of the deviation from nonlinkage (i.e. a lod score of 3 indicates that the probability that the unlinked markers would generate the observed data is 0.001) (Landers *et al.*, 1987). Pairs of loci were considered linked if the lod score exceeded 3.0 (Chang *et al.*, 1988). Chang *et al.* (1988) contended that a lod score threshold of 3.0 for *Arabidopsis* is quite conservative since the *Arabidopsis* genome is seven times smaller than the human genome which also has the lod score threshold of 3.0. Once the linkage groups were assembled, three-point and subsequently $n-$point linkage analyses were conducted as recommended by Landers *et al.* (1987). The RFLP linkage map obtained thus has the reference point arbitrarily set at Position 0 (0 cM (centimorgan)) at the top of the linkage group depicted as vertical lines and all the RFLP markers were assigned to a specific distance from this point in cM as calculated from the segregation data.

The other biological (morphological) and biochemical (isozyme) markers that were included in the analysis were also assigned to specific positions on the linkage map. These assignments were carried out by evaluating the segregation data of these conventional markers from the same F_2 individuals as that evaluated for RFLP markers. Moreover, the segregation data from the conventional markers were analysed for linkage and assigned to the specific positions in the linkage map in exactly the same manner as the RFLP markers. The constructed genetic linkage map was therefore composed of RFLP markers as well as conventional markers. For plant species which already had a classical genetic map, the orientation of RFLP linkage groups with respect to the classical genetic map and the assignment of the linkage groups to specific chromosomes were determined by virtue of the visible markers (biological and biochemical) with previously-assigned genetic positions that were included in the crosses.

Perhaps the most difficult task in the construction of genetic linkage maps using RFLP markers is to assign RFLP loci or linkage groups to specific chromosomes. There are several ways to achieve this. In tomato, tester lines have been constructed from wide crosses to carry many biochemical (isozyme) markers

(Tanksley *et al.,* 1982). This approach, though very laborious, can be applied to RFLP markers by crossing to stocks containing known RFLP markers and then analysing for linkage. In wheat, mapping has been facilitated by the use of aneuploids (Sears, 1954). Mapping the co-dominantly inherited DNA markers to a chromosome or chromosome arm can be done by pollinating each of the aneuploids with the genetically distinct parents and looking for the presence or absence of the maternal marker among the progeny (Burr *et al.,* 1983). In maize, A/B chromosome translocation and monosomic (lacking one chromosome) analyses may be applied (Burr *et al.,* 1983); Helentjaris *et al.,* 1986b; respectively). In A/B chromosome translocation, assignment of a polymorphic locus to a chromosome arm can be made by preparing DNA from a series of plants which are individually hemizygous for each of the chromosome arms and demonstrating that one of these plants contains only the restriction fragment for the female parent (Burr *et al.,* 1983). In monosomics, the analyses are done on DNA from monosomic lines of every chromosome and probed with cloned DNA sequences. The relative intensity of the RFLP bands obtained were then quantified and compared with each other. Where an allele shows half the relative intensity of the bands, it is assigned to that particular chromosome (Helentjaris *et al.,* 1988). Similarly, in trisomics, the RFLP marker from a cloned sequence can be assigned to a specific chromosome by quantifying the relative intensities of the bands resulted from excess copy number of DNA sequences and assigning the RFLP marker to the chromosome in which the trisomics exhibiting 1.5x relative intensity. (McCough *et al.,* 1988) and Ellis and Cleary (1988). A more rapid chromosomal assignment based on this method was proposed by Young *et al.,* (1987) by which five to eight genomic clones from a single library were pooled and used as above. Trisomy has been used to assign RFLP loci to the chromosomes in crops such as rice and pea (McCough *et al.,* 1988; Ellis and Cleary, 1988, respectively).

1.3.6 Sources of Probes for RFLP Analysis

A critical step towards the establishment of RFLPs is the selection of suitable probes. The best source of probes is one which can efficiently show polymorphisms at high frequency. Burr *et al.* (1983) suggested that probes which hybridise to gene families or to moderately repeated DNA can detect polymorphisms at almost equal efficiency. However, the most useful probes are those which hybridise to

unique sequences in the genome (Burr *et al.*, 1983). Landry and Michelmore (1985) have suggested that RFLP analyses require clones of clustered, or single or low copy number from throughout the genome for use as probes. So far, the use of highly repetitive DNA sequences as probes has not been proposed in any RFLP studies.

To date, the sources of probes used in RFLP analysis include: plasmid vector-based genomic libraries, cDNA libraries, lambda-phage based genomic, cosmid-based genomic libraries, libraries, and characterised clones. The most widely used source of probes is the plasmid-based genomic libraries in which several restriction enzymes have been used to construct these libraries. Examples are: *Pst* I (e.g. Song *et al.* (1990), McCough *et al.* (1988) and Helentjaris *et al.* (1988)), *Mbo* I (e.g. Landry and Michelmore (1985) and Landry *et al.* (1987a)), *Eco* RI (Havey and Muehlbaum (1989a, 1989b) and *Hind* III (Helentjaris *et al.* (1988). cDNA libraries constructed from total RNAs of leaf tissue (Sharp *et al.* (1989); Landry *et al.* (1987a) and Helentjaris *et al.* (1985), (1988)), tubers (Rivard *et al.* (1989) and seedlings (Havey and Muehlbauer (1989a, 1989b) have also been used as source of probe in RFLP studies. Lambda-based genomic clones have been used by Apuya *et al.* (1988) and Chang *et al.* (1988). Characterised clones have also used as probes in many RFLP studies. As examples, Polans *et al.* (1985) used pea cDNA clones corresponding to the small subunit of ribulose-1, 5-biphosphate carboxylase (*rbcS*) and polypeptide 15 of the chlorophyll a/b light-harvesting complex (*cab*) to establish the fact that these two multigene families are each organised in a chromosome in a tightly linked complex. Lee *et al.*, (1988) used genomic clone corresponding to vicilin, a pea storage protein to show linkage between the vicilin *Vc-5* locus and the r_b locus which, besides controlling many other functions, is responsible for the variation in lipid content in pea. Ellis *et al.*, (1984) used plasmid clones containing one full repeat of flax rDNA and also a portion of approximately 100 basepairs of 25S rRNA gene to study rDNA length variants in peas using RFLPs. Cloned chloroplast DNA sequences from *Petunia* and Chinese cabbage were screened for the detection of RFLPs in *Brassica* (Song *et al.*, 1990). Apuya *et al.* (1988) used plasmid clones containing coding regions for glycinin genes G1 and G2, and also lectin gene 1 as probes in an RFLP analysis in soybean. Gebhardt *et al.* (1989) used, among others, potato cDNA clones coding for 4-coumarate:CoA ligase (4-CL), phenylalanine ammonia-lyase (PAL) and a

gene involved in the defense reaction of potato against *Phytopthora infestans* and also lambda-based genomic clones containing genes for actin and patatin in the construction of an RFLP linkage map in potato (*Solanum tuberosum*).

Finally, clones from closely related species can also be used as a source of probes. Bonierbale *et al.*, 1988, used single copy clones (cDNA and genomic) from tomato (*Lycopersicon esculentum*) to study the RFLP segregations in potato (*Solanum tuberosum*) while Tanksley *et al.*, (1988b) used identical clones to study RFLPs in pepper (*Capsicum annuum*). Nucleotide sequences of the tomato clones used in both studies were sufficiently conserved to hybridise with homologous sequences in potato and pepper.

The restriction enzyme used to prepare genomic fragments for library construction is extremely important in realising RFLP s. Genomic libraries constructed using DNA fragments obtained from *Pst* I digests are used more frequently than those obtained from *Mbo* I and *Hind* III digests in RFLP studies. *Pst* I is a methylation-sensitive restriction enzyme and as observed in maize (Burr *et al.*, 1988) and tomato (Tanksley *et al.*, 1988a) undermethylated areas are enriched for single copy sequences. It has been reported that *Pst* I libraries yielded a significantly higher percentage of single-copy genomic clones (McCough *et al.*, 1988; and Helentjaris *et al.*, 1988). Furthermore, it was observed in maize that there is no need for detailed evaluation of the probes for single copy sequences nor to screen for and discard those containing repetitive sequences. In other words, the genomic clones generated from *Pst* I digests can be used directly as probes to detect RFLPs in maize without the necessity of screening for single or low copy sequences. However, the screening might still be necessary for genomic clones constructed from genomic fragments generated from restriction enzymes other than *Pst* I and with different plant species.

The screening for single or low copy sequences in genomic libraries can be carried out by hybridising equal amounts of DNAs from test clones with radiolabelled total genomic DNA using a dot-blot technique (Landry and Michelmore, 1985). Clones showing little or no hybridisation constitute those with single or low copy number sequences while clones showing strong hybridisation contain highly repetitive sequences. After the classification, the clones are used in RFLP analysis.

A comparison of the efficiency of genomic libraries with cDNA libraries for detecting polymorphisms showed that polymorphisms were detected 2.5 times more frequently with cDNA probes than with randomly selected genomic probes (Landry *et al.*, 1987b). It was reported that fewer polymorphisms were detected with cDNA clones homologous to single copy sequences than with cDNA clones homologous to multiple copy DNA sequences although both could efficiently detect polymorphisms (Landry *et al.* (1987b) and Bernatzky and Tanksley (1986b)). In contrast, RFLP data obtained in humans where genomic clones detected polymorphisms more frequently than cDNA clones, showed that only cDNA clones homologous to gene families detected polymorphisms (Helentjaris and Gesteland, 1983) . The reason for these contrasting observations is still unknown. A two-fold increase in the efficiency of cDNA clones for detecting RFLPs over genomic clones was also mentioned by Havey and Muehlbauer (1989a) who suggested that the increase in efficiency may be due to polymorphism in the flanking sequences outside the coding regions, or that the cDNA may encompass a greater length of nuclear DNA when compared with genomic clones which have almost equal average size of inserts, as introns are 'spliced out' of the poly A^+ mRNA. cDNA clones represent the 3' end of transcribed sequences and may therefore be more prone to detect polymorphisms in regions flanking the transcribed sequences (Landry *et al.*, 1987b). The comparatively greater length of nuclear DNA represented by cDNA clones means that they can hybridise to a much wider region(s) in the genome and therefore increases the chance of realising RFLPs associated therewith.

1.3.7 Selection of Restriction Enzymes to detect RFLP

Another parameter which can affect the frequency of detecting RFLPs in an organism is the choice of restriction enzymes used for digesting genomic DNA. The sequence and number of nucleotides in the recognition sites of an enzyme plays a major role in realising RFLPs. Landry *et al.*, (1987) observed in lettuce that enzymes recognising 6 base pairs displayed RFLPs more frequently than 4 base-pair recognising enzymes which occasionally showed no hybridisation at all. This observation was later confirmed by McCough *et al.*, (1988) who found in rice that enzymes having 6 base-pair recognition sites showed the greatest number of RFLPs while those that recognised 4 base-pairs revealed the least. For example, Landry *et al.* (1987b) noticed a significantly higher frequency of polymorphism revealed

by *Eco* RI (GAATTC) and *Xba* I (TCTAGA) compared with *Taq* I (TCGA) and *Pal* I (GGCC). In another study, *Xba* I revealed the highest percentage (40.3 %) polymorphisms with a set of probes while *Taq* I revealed only 21.7 % (McCough *et al.*, (1988). Among the commonly used restriction enzymes in RFLP analysis include *Bam* HI, *Bgl* II, *Dra* I, *Eco* RI, *Eco* RV, *Hind* III, *Rsa* I and *Xba* I.

Landry *et al.* (1987b) carried out another study to determine whether genomic regions rich in cytosine and guanine were more polymorphic than regions rich in adenine and thymine by comparing the frequency of polymorphisms revealed by digestion with *Msp* I (CCGG) and *Pal* I (GGCC) with that of *Dra* I (TTTAAA). It was observed that there was no significant difference in the frequencies of polymorphisms found (Landry *et al.*, 1987b).

Barker *et al.*, (1984) reported high mutation rates associated with CpG dimers in mammalian and prokaryotic systems. High mutation rates are, however, not associated with CpG dimers in crops such as rice (McCough *et al.*, 1988) and lettuce (Landry *et al.*, 1987b). In these crops, higher frequencies of polymorphisms were not detected when enzymes with CpG dimers in their recognition sites were used compared with other non-CpG enzymes (McCough *et al.*, 1988; and Landry *et al.*, 1987).

1.3.8 RFLP studies in Oilseed Rape (*Brassica napus*)

Song *et al.*, (1988a, 1988b, 1990) extensively studied the feasibility of RFLP technology in *Brassica* and provided a new insight into the genome evolution of diploid species in *Brassica* and related genera. Song *et al.* (1990) proposed several hypotheses on the phylogenetic relationships between cultivated *B. oleraceae* and wild n=9 brassicas related to *B. oleraceae*, and the evolutionary pathways within *B. oleraceae* and *B. rapa*. Slocum *et al.* (1990) reported a linkage arrangement of RFLP loci in *B. oleraceae*. Up to the start of, and during the present study, no RFLP studies had been reported on the amphidiploid *B. napus* L..

1.3.9 Random Amplified Polymorphic DNA (RAPD)

Current genetic maps of plant species described so far are based primarily on cloned DNA sequences as probes to reveal RFLP and consequently serve as

RFLP markers in the map. Recently, Olson *et al.* (1989) proposed standardising physical mapping of the human genome using "sequence tagged sites" flanked by short unique nucleotide sequences which can be revealed through the use of PCR using the primers corresponding to the unique sequences ((Saiki *et al.* (1989) and Vosberg (1989)). Another similar approach was proposed by Beckmann and Soller (1990) who stated that specific microsatellite islets contained within a stretch of unique DNA sequence could be individually amplified by means of the PCR reaction using a pair of flanking unique oligonucleotides that pointed inwards towards the bracketted microsatellite motif. The proposal was based on the growing realisation that eucaryote genome is densely interspersed with simple tandemly repeated motifs, termed "microsatellites" that exhibit site-specific length variations (Beckmann and Soller, 1990).

The methods of detecting DNA polymorphisms using PCR as described above are based on some prior knowledge of the target sequences found within the genome. Recently, a PCR-based DNA polymorphism assay called Random Amplified Polymorphic DNA (RAPD) or Arbitrarily Primed PCR (AP-PCR) was described independently by Williams *et al.* (1990) and Welsh and McClelland (1990) respectively. The method is based on the amplification of genomic DNA with single primers of arbitrary nucleotide sequence and is able to detect DNA polymorphisms even in the absence of specific nucleotide sequence information, and with little knowledge of the biochemistry of the species being studied. The actual mechanism of the amplification by these arbitrary primers is still unknown. Welsh and McClelland (1990) proposed a mechanism whereby at a sufficiently low annealing temperature, primers can be expected to anneal to many sequences despite a variety of mismatches. Some of these will be within a few hundreds of base pairs of each other and on the opposite strands and therefore can be PCR-amplified to reproducibly generate discrete bands.

The success of this polymorphism assay relies principally on the nucleotide sequences of the primers used and the annealing temperature during the amplification reactions. The nucleotide sequence of each primer was chosen within the constraints that the primer was 9 or 10 nucleotides in length; had between 50 and 80% G+C content; contained no palindromic sequences (Williams *et al.*, 1990). Williams *et al.* (1990) also contended that even a single nucleotide change

in the primer (and, by inference, the template) can determine whether a given DNA segment will be amplified and therefore alter the overall pattern of amplification products. Annealing temperatures above 40°C in the thermal cycling profile which involve denaturation at 94°C for 1 minute, annealing at 36°C for 1 minute and extension for 2 minutes at 72°C prevented amplification by many of the 10-mer oligonucleotides tested (Williams *et al.*, 1990). Welsh and McClelland (1990) found that at an annealing temperature of 60°C no amplification products were observed while annealing temperatures of between 35 and 50°C is sufficient to generate patterns of species-specific amplification products.

In order to explore the possibility of using the observed DNA polymorphisms as genetic markers for the construction of genetic maps and also to assess whether the assay is reproducible, Williams *et al.* (1990) analysed 66 segregating F_2 individuals from a cross of soybean species *Glycine max* and *Glycine soja*. Each polymorphism observed was scored as a dominant marker and correlated with the segregation data for 430 soybean RFLP markers, derived from the same 66 individuals (manuscript in preparation) (Williams *et al.*, 1990). The RAPD markers were mapped to the existing RFLP maps and were able to increase the saturation of the soybean map by filling in some gaps left by the RFLP markers and also by extending the map in the telomeric direction.

Williams *et al.* (1990) contended that nearly all RAPD markers observed in soybean were dominant, as DNA segments of the same length were amplified from one individual but not from another. It is, however, not possible to distinguish whether a DNA segment is amplified from a locus that is heterozygous (1 copy) or homozygous (2 copies) with dominant RAPD markers. Co-dominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, were detected only rarely, i.e. 4 out of 88 RAPD markers in *N. crassa* genetic map Kubelik *et al.*, 1991). Dominant markers were acceptable for genetic mapping using inbred homozygous parents (Williams *et al.*, 1990). However, wherever it is important to know the zygosity of segregating individuals, tightly linked dominant RAPD markers, each diagnostic for a different parental genotype, could be used in pairs to assess the genotype of the chromosomal region and consequently, to distinguish heterozygotes from homozygotes (Williams *et al.*, 1990).

Besides its utility in the generation of DNA markers suitable for the construction of genetic map, RAPD technology can also be used for DNA fingerprinting as described by Welsh and McClelland (1990, 1991). In this study, Welsh and McClelland (1990) utilised the 20 base pair sequencing primers: Universal M13−20, Reversed M13, and the T7 and T3 promoters,for the production of simple and reproducible DNA fingerprints of a complex genome. The use of RAPD fingerprints in the rapid identification of bacterial strains in the genus *Staphylococcus* and *Streptococcus* has been extensively studied and proved very useful especially for surveying a large number of individuals in a population of bacteria for epidemiological or population studies (Welsh and McClelland, 1990). They also extended the application to other species such as rice and maize and successfully obtained DNA fingerprints for several varieties of these species (Welsh and McClelland, 1990 and 1991).

In their most recent work, Welsh and McClelland (1991) took advantage of the tRNA genes which occur in multiple copies dispersed throughout the genome in many species to obtain 'consensus' tRNA gene primers, and used them singly, or in pairs (where primers facing outward from tRNA genes) in PCR reactions to generate RAPD fingerprints. The consensus tRNA gene primers used amplify the region between tRNA genes and generated PCR fingerprints that were generally invariant between strains or varieties of the same species but often substatially conserved between related species (Welsh and McClelland, 1991). They also suggested that RAPD based on tRNA primers could produce fingerprints from the genomes of a wide variety of organisms and also possibly develop tRNA consensus primers that were targetted preferentially to a particular kingdom or to either the nuclear genome or organelle genomes of eukaryotes.

Finally, Welsh and McClelland (1990, 1991) concluded that it was possible to define conditions that generate common patterns between closely related species by making RAPD less sensitive to sequence divergence for the analysis of phylogenetic relationship, and conversely, to define condition that would increase its ability to detect differences in DNA sequence; thereby distinguish strains or varieties within a species.

1.3.10 RFLP, Repetitive DNA, and Genome Architecture

RFLP analysis has become increasingly important in phylogenetic studies by providing vital information on genetic relatedness at the DNA level. Although in earlier studies RFLP analysis was mainly confined to the study of the heredity of genetic diseases in man (See review by Watkin (1988), it has now been extensively utilised in other species including plants (See review by Tanksley *et al.* (1989)) to study genome homologies among primitive races and wild species and their cultivated counterparts. The analysis can therefore provide important information as to what extent a number of crop plants share taxonomic affinities; for example, maize and sorghum are members of the same tribe in the family *Gramineae* (Smith, 1977), and potato, tomato and pepper belong to the nightshade family *Solanaceae* (Tanksley *et al.*, 1989). Another good example is the genus *Brassica.* Cabbage, turnips, and rape seed all belong to this genus (McNaughton, (1976)). A more extensive study on the phylogeny of the genus *Brassica* and the evolution of its cultivated species was recently published (Song *et al.*, 1990). By using RFLP analysis on the DNAs of many wild and cultivated *Brassica sp.* from various geographical regions, Song *et al.* (1990) was able to construct a phylogenetic tree as well as the species' geographical origins. In a similar study done on the family of *Triticeae* (Sharp *et al.*, 1989), the genome relationship between relatives of wheat revealed by RFLP showed that it is in agreement with the current ideas of evolutionary relationships within the *Triticeae* as suggested by Baum (1978).

During genome evolution or species divergence, sequence amplification, deletion and rearrangement at DNA level are the major sources of genetic variation (Flavell, 1982) all of which can subsequently be detected by RFLP analysis. In order to understand the role of these processes during genome evolution, it is important to discuss genome architecture. The genome of higher eukaryotes contains excess DNA that neither codes for specific proteins, nor has it any demonstrable control function for the transcription of genes adjacent to it (Manuelidis, 1982). Repetitive DNA sequences constitute a large proportion of this so called excess DNA. In cereal genomes, for example, over 75% of the total DNA consists of repetitive sequences (Rimpau *et al.*, 1978, 1980). These repetitive sequences are of many different types which are either scattered in the genome and interspersed

with unique sequences (Walbot and Goldberg, 1979) or arranged in tandem arrays (Flavell, 1980,1982). Some sequence elements are arranged in long tandem arrays; the majority of them are located in heterochromatic regions at telomeres, centromeres, or interspersed chromosomal locations.

The term satellite DNA (stDNA) was first used by Kit (1961) to describe minor bands observed after centrifugation of genomic DNA in neutral CsCl density gradients. This type of stDNA is sometimes referred to as density or patent satellites. Cryptic or hidden satellites refer to those which coincide in density with the major DNA components under neutral CsCl gradients and can only be detected by Hg^{2+} or Ag^{+}-Cs_2SO_4 density gradient or by antibiotics in CsCl gradients. Twin satellites, however, refer to two or more satellites coinciding in density in neutral CsCl gradient. In recent years, the term stDNA has been used to describe tandemly arranged highly repetitive sequences regardless of whether or not they are separable from the major components of the DNA by CsCl gradient ultracentrifugation (Beridze, 1986). This tandemly arranged repetitive DNA sequences generate a characteristic ladder of fragments made up of multiples of the basic repeat unit or monomer when genomic DNA is digested with restriction endonucleases and are sometimes referred to as restriction satellites as opposed to gradient satellites.

With new advances in molecular biology techniques in recent years, much attention has been directed towards the study of restriction satellite DNAs. Earlier studies concentrated mainly on animal stDNA (see review by Singer, 1982). In plants, tandemly repeated stDNA sequences were first sequenced by Peacock *et al.* (1981) of a cloned 185 bp repeat from maize. Later, sequence information was determined from several plant species including rye (Appels *et al.* (1981), *Scilla* (Deumling,1981), melon (352 bp unit; Brennicke and Hemleben, 1983), mustard (172 bp unit; Capesius, 1983), broad bean (59 bp unit; Kato, 1984), *Allium cepa* (Barnes, 1985), radish (172 bp unit; Grellet *et al.*, 1985), *Brassica oleracea* (177 bp unit; Benslimane *et al.*, 1986), *Arabidopsis thaliana* (180 bp unit; Martinez-Zapater *et al.*, 1986; and – 160 bp, 180 bp, and 500 bp units; Simoens *et al.*, 1988), several *Cucurbita sp.* (Leclerk and Siegel, 1987), rice (Wu and Wu, 1987), *Brassica juncea* (174 bp unit; Reddy *et al.*, 1989), *Nicotiana tabacum* (Koukalova *et al.*, 1989) and *Brassica campestris* (177 bp unit; Lakshmikumaran and Ranade, 1990).

1.3.11 Structure and Function of Repetitive Sequences

According to Flavell (1982), repetitive sequences in plants can be classified into 3 categories: (1) tandem arrays of closely related repeating units; (2) dispersed repeat families; and (3) multigene families. The majority of the repetitive sequences in plants fell into the first two categories with the first category representing the highest copy number. The repeat elements can display a wide spectrum of repetition frequency up to more than 10^6 copies per genome, and the basic repeating units can range from a simple dinucleotide pair to several kilobasepairs (kbp) in length (Singer, 1982). Recent comparative data have shown that such repetitive sequences are closely related in primary structure, i.e. up to 50% homology and seemed to be derived from a common ancestor elementary sequence of about 60 base pairs long as suggested by Grellet *et al.* (1985). The suggestion was based on the sequence comparison of satellite DNAs from widely evolutionarily distant plants namely radish, maize, broad bean, mustard, wheat and rye (Grellet *et al., 1985*). Satellite DNAs from other *Cruciferae*, as described later, also contained multiples of 60 basepairs (bp). Whether the phenomenon is true in all species is still not known.

Many of the repeating units found in plants were bordered by *Hind* III restriction enzymes recognition sites (e.g. *Brassica sp,*). Other restriction enzymes which flanked the repeating units were *Alu* I (e.g. cereals), and *Bam* HI (e.g. *Nicotiana tabacum*).

Apart from a small fraction which encodes important functions such as ribosomal RNAs (Flavell, 1982; Wu and Wu, 1987) and tRNAs (Benslimane, 1986; Wu and Wu, 1987), the biological significance of the repetitive DNA remains controversial. It has been proposed to be involved in chromosome-folding and pairing, determination of nuclear size, gene regulation or to play a role in speciation processes and genome evolution (Flavell, 1982). Rees *et al.* (1982) concluded that the supplementary, non-coding DNA, which is largely repetitive sequences affected chromosome pairing and chiasma formation at meiosis and further added that the divergence and evolution of flowering plants, were often accompanied by massive quantitative changes of this DNA. However, Simoens *et al.* (1988) argued that the repetitive sequences may have no function at all and further added that it

might reside and accumulate in the genome because there was no phenotypical disadvantage. The question of why some species have large genomes while others have very little (e.g. *Arabidopsis*) was, so far, not addressed. Flavell (1982) suggested that species with large genomes, i.e. with large proportions of non-coding DNA were probably tolerate more macromutations and therefore diverged more rapidly between populations than the genomes of organisms with low DNA contents. Bennett (1972) found that the duration of the mitotic cell cycle increased linearly with increase in nuclear DNA, even though the composition of the DNA and its quality, was vastly different for different species. *Arabidopsis* has a comparatively short life cycle. It is still not known whether or not *Arabidopsis* was naturally selected for small genome to maintain short life cycle; or whether the genome is still undergoing the slow process of accumulating the non-coding DNA, thus increasing the genome size and eventually will have a major effect on its life cycle.

CHAPTER II

MATERIALS AND METHODS

2.1 GLASSWARE AND PLASTICWARE

All glassware and plasticware used in experiments involving nucleic acids were siliconised and autoclaved at 121°C, 15 p.s.i., for 20 minutes. Siliconisation was carried out by rinsing inner surfaces of glassware with 5% (v/v) dimethyldichlorosilane/chloroform, inverted on a few layers of paper towel to remove excess liquid and left to dry under fumehood for 30 minutes. For small items (tips, pipettes, microcentrifuge tubes and test-tubes), siliconisation was carried by placing them under vacuum inside a large vacuum dessicator with a small beaker containing 1 ml dimethydichlorosilane for 2 hours. After siliconisation, glassware and plasticware were thoroughly rinsed with distilled water before autoclaving.

For work involving RNA samples, glassware was baked at 110°C for at least 4 hours and plasticware was treated with 0.01% v/v diethyl pyrocarbonate (DEPC) after siliconisation, followed by autoclaving.

2.2 CHEMICALS AND BIOLOGICAL REAGENTS

Unless otherwise indicated, all chemicals and reagents used were purchased from BDH Chemicals Ltd., Poole, Dorset, UK and were of analytical grade (AnaLar). Other materials were obtained from the designated sources as listed below.

Bovine serum albumin (BSA), polyvinyl pyrolidone (PVP), herring sperm DNA, sodium iodide, xylene cyanol, bromophenol blue, ethidium bromide (EtBr), dithiothreitol (DTT), ATP, RNase A, pronase, proteinase K, lysozyme, sodium dodecyl sulphate (SDS), sodium lauryl sarcosinate, ampicillin, isoamyl alcohol, dimethyl formamide, dimethyl sulfoxide (DMSO), and diethyl pyrocarbonate (DEPC) were from Sigma Chemical Company, Poole, Dorset, UK.

Restriction endonucleases and DNA modifying enzymes were purchased from Northumbrian Biologicals Ltd., Cramlington, UK., Boehringer Manheim GmbH, Manheim, W. Germany, and New England Biolabs, Beverly, Maryland, USA. cDNA Synthesis Kits, oligo-dT-cellulose, caesium chloride (CsCl), 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal), isopropylthiogalactoside (IPTG), dNTPs and random hexanucleotides (random primers) were also purchased from Boehringer Manheim.

Taq DNA polymerase with 10x reaction buffer for PCR reactions was purchased from Promega, Madison, Wisconsin, USA.

3MM paper, filter papers and *GF/C* glassfiber discs were from Whatman Ltd, Maidstone, Kent, UK. Nitrocellulose filters (Grade BA-85) were from Schliecher and Schuell GmbH, Dassel, W.Germany. 'Hybond-N' nylon filters, radioactive ^{32}P-dATP and ^{32}P-dCTP were from Amersham International plc., Amersham, Oxon, UK. X-Ray films (Fuji RX100) were from Fujimex, Swindon, Wiltshire, UK.

Ficoll 400, Sepharose CL-4B, Sephadex G-50, *Eco RI* linkers (phosphorylated and non-phosphorylated) and glycogen were purchased from Pharmacia Fine Chemicals, Milton Keynes, Bucks., UK.

Guanidium thiocyanate and guanidium hydrochloride were from Fluka Chemical Ltd., Glossop, Derbyshire, UK.

Agarose, low melting agarose, high gelling temperature agarose and maximum efficiency DH-5α competent cells were from BRL Ltd., Uxbridge, Middlesex, UK.

Yeast extract was from Biolife S.r.l., Milan, Italy. Bacto-agar was from Difco, Detroit, Michigan, USA. Bactotryptone was from Oxoid Ltd., Basingstoke, Hants, UK.

MS Salts medium (Murashige and Skoog) and 7x detergent were from Flow Laboratories, Rickmansworth, Hertfordshire, UK.

2.3 BACTERIAL STRAINS AND CLONING VECTORS

Bacterial strains used were derivatives of *E. coli* K12. Cloning vectors used were either plasmids or bacteriophages. Below is the list of bacterial strains and

cloning vectors used (including genotypes wherever applicable).

2.3.1 Bacterial strains

DH5α – *supE*44,Δ*lac*U169, (ϕ80*lacZ*ΔM15), *hsd*R17, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *rel*A1. (Bethesda Research Laboratories, 1986)

JM101 – *supE thi*,Δ(*lac-proAB*), F'[*traD*36, *proAB*$^+$ *lacI* ^{q}*laZ*ΔM15]. (Messing, 1979)

TG2 – *supE,hsd*Δ,5*thi*,Δ(*lac-proAB*),Δ(*srl-recA*)306::Tn*10* (*tet*^{r}) (Sambrook, 1989)

2.3.2 Cloning Vectors

Plasmids

pUC18/19 – amp^r, *lacZ* cloning region from M13 mp18/19 (Yanish-Perron *et al.* (1985).

pBR322 – amp^r, tet^r, ColE1 replicon, *bom*$^+$ (Bolivar *et al.* (1977).

Bacteriophage M13

mp18/19 –*lac Z* cloning regions

2.4 BACTERIAL CULTURE MEDIA

For the growth and maintenance of bacterial cultures, the following media were used.

YT-medium: 3 g bactotryptone, 5 g yeast extract, 2.5 g NaCl per 1 litre.

YT-medium (2x) : 6 g bactotryptone, 10 g yeast extract, 2.5 g NaCl per 1 litre.

LB-medium (Luria-Bertani) : 10 g bactotryptone, 5 g yeast extract, 2.5g NaCl, 0.5 g D-Glucose per 1 litre.

For agar plates, 15 g of bactoagar was added per litre of medium before

autoclaving (121°C, 15 p.s.i., 20 minutes). Before adding antibiotic (50 μg/ml ampicillin from 10 mg/ml stock in 70% ethanol), the medium was allowed to cool to about 55°C. For overnight bacterial cultures of ampicillin resistant strains, LB-medium was supplemented with 50 μg/ml ampicillin. For the detection of functional β-galactosidase activity in transformed *E. coli* (DH-5α) clones, the YT-agar (55°C) was supplemented with 40 μg/ml X-gal.

2.5 PLANT MATERIALS

Commercial varieties of oil seed rape (*Brassica napus*) were obtained from various seed companies and individuals in the United Kingdom, Poland, Sweden and West Germany (See Acknowledgements). Seventeen winter varieties and twelve spring varieties were obtained from these companies. These were Ariana, Astec, Bien Venu, Bolko, Cobra, Doublol, Jantar, Jet Neuf, Libravo, Lictor, Liradona, Lirawell, Mikado, Pasha, Rafal, Score, and Tapidor for the winter varieties; and Bronowski, Brutor, Comet, Drakkar, Elin, Global, Hanna, Optima, S.V. Karat, Topas, Westar and Willi for the spring varieties. Most of these varieties are available to farmers in their respective countries either as recommended varieties or as trial varieties. With respect to the erucic acid and glucosinolate contents of these varieties, Ariana, Bronowski, Cobra, Doublol, Drakkar, Libravo, Lictor, Liradonna, Lirawell, Optima, Score, Tapidor and Topas are 'Double Zero' (i.e. Low in erucic acid and low in glucosinolate content) while Astec, Bien Venue, Brutor, Jet Neuf, Mikado, Pasha, Rafal, and Willi are 'Single Zero' (i.e. Low in erucic acid but high in glucosinolate contents). One double high variety (no varietal name; designated as double high (DH)) was also obtained. The status of the other varieties with respect to glucosinolate content are still unknown as the suppliers were unable to supply this information. As these varieties or indeed any other cultivars not listed here come from a narrow genetic background due to their relatively recent breeding programme, a very small number of scorable genetic characteristics are present in these cultivars. One of these is flower colour in which variety Pasha has orange flower as opposed to yellow in other varieties. Other characteristics include petal, leaf and pod shapes (Variety Brutor has elongated petals), bristleless (Variety Mikado), height, node length, early/late flowering, seed colour, etc.. The genetic analysis of these characters are very subjective and is greatly dependent

on growth conditions. The information about each variety used in this research programme is summarised in Table 3.1.

Seeds from each variety were planted in 6-inch pots containing sterile peat-based compost and perlite mixture (1:1 volume) and grown in a growth room with 16-hour lighting and temperature range between 18°C (dark) to 22°C(light). Plants were watered daily and fed with 1/2 concentration of Murashige/Skoog (MS) salts once a week. Young leaves were harvested into liquid air and were either stored at −80°C or used immediately for DNA extraction.

2.6 CROSSES BETWEEN VARIETIES

Crosses were made between rape varieties to study the inheritance of certain characters. Unopened flowers were selected so as to ensure that pollen was still inside the anthers. The anthers were removed (emasculated) and pollen taken from a different variety was used to pollinate the ovules. Great care was taken to ensure that all six anthers were removed. Pollination was carried out either by taking the whole open anthers containing the pollen using a pair of tweezers or pollen extracted from newly opened anthers using sterile cocktail stick and applying the pollen generously onto the stigmas. Whole anthers from several flowers of the same varieties were used in the pollination to ensure a good and unbiased representation of pollen from a particular variety. Anthers from newly opened flowers were carefully selected and used in the pollination to ensure that the pollen was still viable.

Pollinated flowers were covered with polythene bags and occasional inspections were carried out to check the development of pods. Flowers which emerged later were removed to avoid the risk of pollen contamination. The pods containing the seeds were allowed to develop for about 8 to 10 weeks post-anthesis with regular feeding as described earlier in Section 2.5. At this stage, the feeding was stopped to allow mature seeds and pods to dry out for about 2 to 4 weeks before harvesting. Harvested seeds were either immediately replanted for selfing to obtain F_2-seeds or stored at 4°C until used.

Harvested F_1-seeds obtained from crosses between varieties were planted and allowed to flower. The inflorescence was allowed to develop with daily inspection

and was covered with polybags just before the flowers opened up. After anthesis had started, the inflorescence was gently shaken everyday inside the polybag so as to facilitate self-polination. Special care must be employed to ensure that the polybag remained in place to avoid pollen contamination from other plants. Pods were allowed to develop and plants maintained as described earlier in this section.

2.7 VERNALISATION OF WINTER VARIETIES

Vernalisation was necessary to induce winter varieties of oilseed rape to flower. Vernalisation was carried out as follows. 3 to 4 weeks after planting, when the plants reached about 20 cm high and had 5 to 6 mature leaves, the plants were transferred to a growth room at a temperature of about 8 to 10°C to acclimatise the plants. After 1 to 2 weeks at this temperature, the plants were transferred to a 4°C cold room with 9-hour daylight lighting. After 6 to 8 weeks, the plants were transferred back to the 8 to 10°C growth room for about 1 to 2 weeks before returning them to the original growth room. Flowering usually start after about 2 weeks. Acclimatisation steps before and after the vernalisation were essential to ensure a gradual temperature change and avoid temperature shock.

2.8 STORAGE OF BACTERIA

Bacteria grown on plates were stored, inverted and sealed with Nescofilm for up to 6 weeks at 4°C. For continued storage at 4°C, the bacterial colonies were subcultured onto fresh plates. For long-term storage, bacterial lawns grown from single clones on selective agar plates were transferred into sterile vials containing 1 ml of LB-medium and 50 μg/ml ampicillin, and mixed thoroughly by vortexing. One ml of sterile 80% glycerol was added to each vial, vortex-mixed and finally stored at −80°C.

The storage of cDNA clones was carried out in 96-well, U-bottomed, microtitre plates which contained 50 μl LB-medium, 50 μg/ml of ampicilin and 40% glycerol in each well. Using sterile cocktail sticks, white transformants were selected and inoculated into individual wells of the microtitre plates. After all the wells were inoculated, the microtitre plates were sealed with Nescofilm and incubated overnight at 37°C on a plate-shaker. An equal volume (50 μl) of sterile 80% glycerol was added into each well by using a multipipettor and the plates

were gently shaken for 10 minutes until the glycerol and the cell suspension were thoroughly mixed. The microtitre plates were carefully labelled, resealed with Nescofilm and stored at $-30°C$. Duplicate plates were made, to avoid accidental loss of the clones, using a specially designed replicating fork. This was simply an aluminium block with a handle and 48 equal-length steel spikes, where each spike was positioned on the block to exactly correspond to the bottom-centre of each of the 48 wells (one-half) of the 96-well microtitre plate. The replicating fork was used to facilitate the transfer of bacterial inoculi from each of the stock bacterial-glycerol suspensions stored in the wells of the microtitre plates. It is also used to transfer bacterial inoculi from the microtitre plates to nitrocellulose filters for *in situ* colony hybridisation and screening as shall be discussed later in Section 2.10.4.

2.9 BIOCHEMICAL TECHNIQUES

2.9.1 Centrifugation

Centrifugation steps were carried out in some of the procedures employed. For manipulations involving the use of eppendorf tubes, centrifugation was carried out in an MSE Micro-Centaur microcentrifuge at 13,000 rpm (12,000 x g). For manipulations involving the use of Corex tubes, centrifugation was carried out in a Beckman Model J2-21 high-speed centrifuge using a Beckman Swing-out rotor Model JS13.1 at 10,000 rpm (25,000 x g). Centrifugation times and temperatures were indicated at each centrifugation step.

Ultracentrifugation of caesium chloride/EtBr gradients was carried out in a Sorvall OTD65B Ultracentrifuge using either a Beckman Type 70Ti Anglehead rotor at 50,000 rpm (300,000 x g) or a Beckman Type VTi 50 Vertical rotor at 50,000 rpm (250,000 x g) for gradient volumes up to 36 ml and a Beckman Type VTi 65 Vertical rotor at 45, 000 rpm (250,000 x g) for 5 ml gradient volume. All ultracentrifugation runs were carried out using Beckman Quick-Seal polyallomer centrifuge tubes.

2.9.2 DNA Precipitation Using Alcohol

DNA samples down to concentrations of $1\mu g/ml$ were precipitated by first

adding 0.1 volume of 3M sodium acetate pH 5.2 to the DNA solution and mixing gently, followed by two volumes of cold 100% ethanol, again with gentle mixing, and the mixture incubated at $-20°C$ for 2 to 3 hours or overnight. For precipitating DNA samples of concentrations lower than 200 ng/ml, 1 to 2μg glycogen per 10 ml total volume was added as carrier. Precipitation was occasionally performed for 30 minutes at $-70°C$. Gentle mixing was employed when handling high molecular weight DNA to avoid shearing. DNA precipitated in an eppendorf tube was pelleted by centrifuging for 10 minutes at 13,000 rpm (12,000 x g). Larger samples up to 10 ml volume were centrifuged for 30 to 60 minutes at 4°C depending on the size and concentration of the DNA, at 10,000 rpm (25,000 x g) using 15ml or 30 ml Corex tubes. DNA was washed three times with cold 70 % ethanol, centrifuged as before for 5 minutes each time, and then dried in a vacuum dessicator for 5 to 10 minutes. The dry DNA pellet was resuspended in sterile TE buffer (10mM Tris-HCl, 1mM EDTA, pH 7.2) or in sterile H_2O to the required concentration. Occasionally, 0.6 to 1.0 volume of isopropanol was used instead of ethanol where the final volume of the solution was to be minimised when precipitating DNA of low concentration.

2.9.3 Deproteinisation of Nucleic Acid Samples Using Phenol

Phenol for deproteinising nucleic acids was obtained by the distillation of phenol crystals, collection and storage under TE buffer (pH 7.2) and nitrogen at $-20°C$.

Solutions containing nucleic acids were deproteinised by extraction with phenol-chloroform-isoamyl alcohol (25:24:1 v/v respectively) henceforth referred to as "phenol" only. 0.5 to 1.0 volume of phenol was added to the sample of at least 100 μl volume, gently mixed by inverting the tube and then centrifuged briefly for 30 seconds in a microcentrifuge. Larger samples were deproteinised in Corex tubes and then centrifuged for 5 minutes at 10,000 rpm. The aqueous phase (upper layer) was carefully removed and transferred to a fresh tube and the extraction repeated for at least three times. When extracting small amounts of nucleic acids (less than 100 ng), the phenolic phase was back-extracted by adding an equal volume of TE buffer and the resulting aqueous phase pooled with the original aqueous phase. After phenol extractions, the sample was extracted 3

times with equal volumes of chloroform-isoamyl alcohol (24:1 v/v) to remove the remaining traces of phenol. DNA samples for sequencing were finally extracted 3 times with ether (upper layer). Excess ether was removed by centrifuging DNA samples at 5000 rpm (12500 x g) in a vacuum centrifuge for 30 minutes. DNA was then recovered by ethanol precipitation as described in the preceeding section (Section 2.9.2).

2.9.4 Analysis of Nucleic Acids Using Spectrophotometry

In order to determine the concentration of nucleic acid in DNA or RNA solutions, a small sample (1 μl to 5μl) was added to 1 cm quartz cells containing 1 ml TE buffer or water, the solution was thoroughly mixed, and the optical density (OD) was measured at a wavelength of 260 nm against a blank of TE buffer or water. An OD at 260 nm of 0.020 corresponds to a DNA concentration of 1 μg/ml while an OD of 0.025 corresponds to an RNA concentration of 1 μg/ml. The concentration of oligonucleotides used as primers in Polymerase Chain Reaction (PCR) was determined by measuring the OD_{260} against a blank of distilled water as described above where an OD value of 0.05 corresponds to an oligonucleotide of 1 μg/ml.

Analysis of the purity of the nucleic acids was estimated from a spectrophotometric scan from 200 to 300 nm by calculating the OD ratio of 260/280 nm whereby ratios for pure DNA and RNA were 1.8 and 2.0 respectively.

2.9.5 Fluorimetric Analysis of DNA using Diaminobenzoic acid

Estimation of DNA concentration was determined by performing fluorimetric analysis using diaminobenzoic acid (DABA) as described by Thomas and Farquhar (1978). 1 to 5 μl samples of DNA solutions to be analysed were placed in 1.5 ml eppendorf tubes. Samples of volumes greater than 5 μl were first dried under vacuum. 20 μl DABA solution (400 mg/ml in H_2O) was added to each tube, mixed, centrifuged briefly, and incubated at 60°C for 30 minutes. 1.6 ml of 1M HCl was added to each tube, mixed thoroughly, and the relative fluorescence was measured on a Baird-Atomic *Flouripoint* fluorimeter at 405 nm excitation wavelength and 505 nm emission wavelength. A standard curve was constructed using λ-DNA

standards containing 20 ng to 1 μg of DNA. The concentration of DNA was then estimated using the standard curve.

2.9.6 Extraction of High Molecular Weight DNA from Leaf Tissues

Total genomic DNAs were isolated from plant tissue according to the procedure described by Shirsat (1984). Briefly, 5 g of leaf tissues were frozen in liquid air and ground in a cold mortar and pestle to a fine powder. 5 ml homogenising buffer (0.1M NaCl, 0.025M EDTA, 2% SDS), 2.5 ml 5M sodium perchlorate, 5 ml redistilled phenol, and 5 ml 1% (v/v) octanol/chloroform mixture were added sequentially with thorough mixing. The resulting suspension was transferred to a sterile 100 ml conical flask and was shaken on a rotary shaker (100 rpm) for 1.5 hours at 4°C. The suspension was then centrifuged in Corex tubes for 10 minutes at 10,000 rpm at 4°C to separate the phases. The aqueous (upper) phase was removed and extracted once with an equal volume of octanol/ chloroform mixture. The DNA was then recovered by spooling using two to three volumes of cold (−20°C) 100% ethanol gently layered over the supernatant. The spooled DNA was gently resuspended in 1 ml resuspension buffer (TE buffer, pH 7.2) overnight at 4°C. 25 μl of stock pronase (25 mg/ml; previously self-digested for 2 hours at 37°C) was added and the mixture was incubated at 37°C waterbath for two hours. The DNA was then purified by centrifuging twice on caesium chloride/ethidium bromide gradients using a Sorvall OTD65B Ultracentrifuge. The band containing the DNA was collected, extracted with caesium chloride-saturated isoamyl alcohol to remove the ethidium bromide and dialysed against TE buffer for 48 hours at 4°Cwith several buffer changes. Dialysed DNA was then concentrated by ethanol precipitation (overnight at −20°C), pelleted, redissolved in TE buffer (pH 7.2), and the concentration and purity determined using a spectrophotometer as described in Section 2.9.4.

All the extracted DNAs were tested for digestibility using various restriction enzymes followed by fragment analysis on agarose gels. Unrestricted samples of the DNAs were also analysed in the experiment to check the size and quality of the intact DNA.

2.9.7 Isolation of Total RNA from Plant Tissue

Total RNA samples were isolated from plant tissues using a combination of methods by Chomcynski and Saachi (1987) and Logemann *et al.* (1987). 1 g of frozen plant material was placed in a 35 ml MSE tube or similar container. 4.5 ml of cold guanidinium thiocyanate extraction buffer was quickly added and the tissue was allowed to thaw for several seconds before being homogenised with a polytron blender for 20 seconds at speed 10. Guanidinium thiocyanate extraction buffer was 25 g guanidinium thiocyanate, 29.3 ml H_2O, 1.76 ml 0.75M sodium citrate pH 7.0, 2.64 ml 10% sarcosyl (previously heat-treated at 65°C for 2 hours), and 36 ul β-mercaptoethanol for every 5 ml buffer, added just before use. The homogenate was centrifuged for 10 minutes at 10,000 rpm (25,000 x g), the clear supernatant poured off into 15 ml Corex tubes, extracted twice with 5 ml phenol, followed by extraction 3 times with chloroform/isoamylalcohol (24:1 v/v) and centrifuged again for 20 minutes at 10,000 rpm. The final aqueous phase was collected, 1/10 th volume of 2M sodium acetate pH 4.0 was added, and gently mixed followed by 2 volume of cold 100% ethanol, gently mixed again and the RNA finally allowed to precipitate overnight at −20°C. The RNA was pelleted by centrifuging for 30 minutes at 10,000 rpm. The pellet was washed 3 times with 5 ml of 3M sodium acetate (pH 5.2) with centrifugation for 5 minutes at 10,000 rpm at each step. The pellet was finally washed with 70% ethanol made with sterile, DEPC-treated H_2O and resuspended in sterile, DEPC-treated H_2O for about 10 minutes in an icebath. RNA samples were reprecipitated by adding 0.2 volume of 1M acetic acid and 0.7 volume of 100% ethanol as described by Logemann *et al.* (1987) to remove any contaminating DNA in the RNA samples. All RNA samples were stored under liquid air.

2.9.8 Analysis of Total RNA

Samples of total RNA were analysed by scanning on a spectrophotometer between wavelengths of 200 nm and 300 nm. The purity was assessed by calculating the ratio of OD's at 260 nm over 280 nm and the concentrations of the RNA samples were determined as described in Section 2.9.4. Further analysis of the RNA samples was done by running a denaturing gel electrophoresis as described in Section 2.9.12.

2.9.9 Rapid Minipreparation of plasmid DNA

The modified method of Birboim and Doly (1979) was used in the preparation of small quantities of plasmid DNA. 10 ml aliquots of LB-medium with added ampicillin (50μg/ml) in universal bottles were each inoculated with bacteria from single colonies using of a sterile toothpick, and grown overnight at 37°C on a rotary shaker. The cells were pelleted by centrifugation at 4200 rpm (600 g) for 10 minutes and resuspended in 200 ul of lysis buffer (50mM glucose, 25mM Tris-HCl pH 8.0, and 10mM EDTA pH 8.0) containing 4 mg/ml lysozyme (freshly added to the buffer). The cell suspension was transferred to a 1.5 ml eppendorf tube and placed on ice for about 30 minutes. 400 ul of freshly prepared alkaline-SDS (1 % SDS, 0.2M NaOH) was added to the tube, mixed gently by inverting the tube several times and kept on ice for a further 5 minutes. 300 ul of ice cold potassium acetate solution (3M potassium, 5M acetate) pH 4.8 was added and the solution was thoroughly mixed before incubating on ice for a further 15 minutes with occasional mixing. The sample was then centrifuged in a microcentrifuge at 13000 rpm (12000 x g) for 10 minutes. The supernatant was phenol extracted once, chloroform extracted twice and precipitated with ethanol. The pellet was washed with 70% ethanol, vacuum dried and resuspended in TE buffer (pH 7.2). The plasmid DNA was stored at -80°C.

1 to 2 μl RNase A (10 mg/ml, previously heat-treated by boiling for 20 minutes to inactivate DNase) was normally added during restriction digestions of the plasmid DNAs isolated using this method.

2.9.10 DNA Restriction Endonuclease Analysis

Genomic DNA was digested with restriction enzymes according to the manufacturer's recommendations. Buffers for the restriction enzymes were either supplied by the manufacturers or prepared as described in Table 2.1. The buffers described in Table 2.1 were used either for restriction enzymes not supplied with the manufacturer's recommended buffers or for simultaneous digestions as recommended by Sambrook *et al.* (1989).

Digestion was carried out for 2 to 4 hours or overnight (12 to 16 hours) using enzyme concentrations of 3 U/μg DNA unless otherwise indicated, and incubated

Table 2.1 — Buffers for Restriction Endonucleases

Buffer	Components (mM)			
	Tris/HCl (pH 7.5)	$MgCl_2$	DTT	NaCl
Low Salt	10	10	1.0	None
Medium Salt	10	10	1.0	50
High Salt	10	10	1.0	100

at the temperatures recommended by the manufacturers. Most of the enzymes used have been shown to work adequately at relatively wide range of NaCl concentrations and hence multiple digestions were performed simultaneously in the same buffer. For combinations of enzymes in which digestions could not be performed simultaneously, the enzyme requiring a lower concentration of NaCl was added first and incubated for 1 to 2 hours. After adjusting the NaCl concentration by adding an appropriate amount of NaCl, the second enzyme was added to the mixture and the reaction mixture was further incubated. Where necessary, the concentration of glycerol in the mixture was also adjusted by dilution to avoid star activity.

2.9.11 Agarose Gel Electrophoresis of DNA

Analysis of DNA fragments using agarose gel electrophoresis was done according to the method described in Sambrook *et al.* (1989). Briefly, the required amount of agarose [0.5% to 1.5% (w/v)] depending on the size of DNA fragments to be resolved), was weighed and added to Tris-acetate EDTA (TAE) buffer (40mM Tris-acetate pH 7.7, 2mM EDTA). The mixture was boiled in a microwave oven (medium-heat setting) with occasional swirling until the agarose was completely dissolved. The solution was cooled under cold running tap water with constant gentle swirling until the temperature was about 55°C before EtBr was added to a final concentration of 0.5 μg/ml and the solution was poured into a perspex mould fixed to a glass plate by silicone grease and fitted with an appropriate well-forming comb. For routine electrophoretic analysis, the perspex mould used can cast a gel measuring 180x150x8 mm which requires 200 ml of agarose solution. For electrophoresing larger numbers of DNA samples for RFLP analysis, a specially designed perspex mould was used which can cast a gel measuring 240x200x10 mm and requires 450 ml of agarose solution. The comb and the mould were carefully removed after the gel had completely set. The gel was then transferred to an electrophoresis tank containing enough TAE buffer to cover the gel with about 5 mm of buffer. Electrophoresis buffer also contained 0.5 μg/ml of EtBr. The DNA samples to be electrophoresed were prepared by adding 1/10 th volume of 10x loading dye (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 50% (v/v) glycerol, 10 mM EDTA). The samples were gently mixed and briefly centrifuged (2 to 3 seconds) to collect the solution in the bottom of the tube. The samples were then carefully loaded into the wells and the electrophoresis was performed

either for 2 to 3 hours at 120 V (150 mA) or overnight at 20 to 30 V (40 mA). The DNA bands were visualised on a UV-transilluminator (300 nm) and photographed with a Polaroid MP-4 Land camera through a Kodak 23A Wrattan filter, using Polaroid film type 667 with an exposure of 1 to 2 seconds at f8.0. Occasionally, the gel was destained in distilled water for 2 to 3 hours in the dark with several water changes in order to reduce background EtBr staining prior to photography.

Minigel electrophoresis was used to monitor the progress of restriction digests and to estimate DNA concentrations, due to its rapidity and ability to detect small amounts of DNA. The procedure was essentially the same as that for the normal size gels already described above with the exception that a minigel apparatus (UniScience, Cambridge) was used are in which a 100x80x5 mm agarose gel could be casted directly into the tank. Tris-borate EDTA (TBE) buffer (0.089M Tris-borate, 0.089M boric acid, 2mM EDTA) was used instead of TAE buffer. Minigels were run for 30 minutes to 1 hour at 50V (40mA).

2.9.12 Denaturing Agarose Gel Electrophoresis of RNA

RNA samples were electrophoresed in a formaldehyde denaturing agarose gel using the method described by Miller (1987). Electrophoresis was performed on a 1.0% (w/v), 150x180x5 mm agarose gel prepared by dissolving 2.0 g of high gelling temperature agarose in 164 ml buffer containing 0.05M morpholinopropanesulfonic acid (MOPS) and 1mM EDTA, pH 7.0. After the solution was allowed to cool to 60°C, 36 ml 37% (v/v) formaldehyde solution was added, mixed, poured into the gel mould and allowed to set for at least 1 hour. RNA samples and DNA size markers were prepared by adding 2.2 μl Buffer A (294 μl 10x MOPS/EDTA buffer in 706 μl H_2O) and 4.8 μl formaldehyde/formamide solution (final concentration of 2.2M formaldehyde and 50% formamide), heating at 70°C for 10 minutes and cooling on ice until used. The gel was pre-electrophoresed at 60 V for 3 minutes before loading. After the addition of 1.5 μl loading buffer (322 μl Buffer A, 5 mg xylene cyanol, 5 mg bromophenol blue, 400 mg sucrose, 17.8 μl formaldehyde, 500 μl formamide), samples were loaded and electrophoresed at 100 V (60mA) for 2 to 3 hours.

After the electrophoresis was complete, the gel was stained for 5 minutes in 5 μg/ml EtBr in H_2O, destained in the dark for 1 to 2 hours with several changes

of distilled water and photographed as described in Section 2.9.11.

All the manipulations throughout the course of this procedure were performed under a fume hood.

2.9.13 Southern Transfer of DNA Fragments onto Filters

Southern transfers were carried out essentially according to the procedure of Southern (1975). Briefly, the gel was treated by submerging in denaturing solution (0.4M NaOH, 0.8M NaCl) for 1 hour, neutralising solution (0.5M Tris.HCl, 1.5M NaCl, pH 7.6) for 1 hour, and 20x SSC (0.3M sodium citrate, 3.0M NaCl) for 30 minutes or until blotted. For genomic DNA digests, the gel was first treated with 1% HCl (v/v) for 15 to 30 minutes pior to denaturation to cleave large fragments of DNA to facilitate their transfer to nitrocellulose or nylon filters. The DNA fragments were transferred in 20x SSC onto nylon filters " Hybond-N" (Amersham) either by capillary blotting overnight or by vacuum blotting for 2 to 3 hours using a Hybaid "Vacu Aid" vacuum blotter (HyBaid Ltd, Teddington, Middlesex, U.K.), both at room temperature. Capillary blotting was performed by placing the gel on an apparatus consisting of a tray containing a glass plate on a platform about 3 cm high overlaid with 2 layers of oversize 3MM paper acting as wicks on all sides of the glass plate and filled up to a level 1 cm below the top of the glass plate with 20x SSC. A Nylon filter (Amersham International, Oxon) presoaked in 20x SSC was carefully placed over the gel and several layers of presoaked 3MM papers were overlaid on top of the filter. Finally, 4 layers of disposable nappies cut to an appropriate size were placed on the 3MM papers followed by a 1 kg weight on a glass plate. After the transfer, the position of the wells and the orientation of the gel were marked on the filter before lifting it off the gel and placed onto clean 3MM paper.

DNA was fixed to the nylon filters either by crosslinking for 2 minutes (DNA side down) on a UV-transilluminator or baking for 1 hour at 80°C in a vacuum oven sandwiched between two glass plates.

2.9.14 Northern Blotting of RNA

RNA electrophoresed in a denaturing agarose gel was transferred onto ny-

lon filters by the method recommended by Sambrook *et al.* (1989). After electrophoresis, the RNA was directly transferred onto a capillary blotting apparatus and subsequent manipulations were exactly as that described for Southern blotting of DNA fragments (Section 2.9.13) except that 10x SSC was used as the transfer buffer .

2.9.15 Recovery of DNA from Agarose Gels by Electroelution

The DNA was digested with the appropriate enzymes, electrophoresed and a small piece of gel containing the band of DNA as visualised on a UV-transilluminator was excised. The gel piece was put into dialysis tubing secured at one end with a dialysis tubing clip. 0.5 ml of sterile TBE electrophoresis buffer was added, and after securing the open end of the tubing with another clip, the DNA fragments were electroeluted in TBE buffer using a minigel tank by electrophoresing at 60 V (50mA) for about 30 minutes or until the DNA was visible as a thin line on the side of the tubing under UV-illumination. The polarity of the current was then reversed and electrophoresis was continued for about 1 minute to detach the eluted DNA from the dialysis tubing. The gel piece was carefully removed from the tubing and the buffer containing the eluted DNA fragment was transferred to a 1.5 ml Eppendorf tube. The DNA was extracted with phenol and chloroform and finally precipitated with ethanol.

2.9.16 Radiolabelling of DNA Probes by Random Priming

DNA fragments for use as hybridisation probes were isolated from agarose gels by electroelution. Radiolabelling of DNA fragments was performed using the oligonucleotide random priming method as outlined by Feinberg (1984). Purified DNA fragments (10 to 100 ng) were first denatured at 100°C for 10 minutes and immediately put into an icebath until required. Random priming buffer [Buffer O (prepared by mixing 1 ml of solution containing 1.25M Tris-HCl, 0.125M $MgCl_2$(pH 8.0), 18 μl β-mercaptoethanol, and 5 μl each of commercially prepared dATP, dGTP, and dTTP], Buffer L (2M HEPES (pH 6.6)), and Buffer B (Hexanucleotides (90 OD units/ml in TE buffer (pH 7.2))), in a ratio of 20:50:30 (v:v:v)), nuclease-free bovine serum albumin, ^{32}P-dCTP (50 to 100 μCi from stock of specific activity greater than 400 Ci/mmole) and DNA polymerase (Klenow enzyme; 2

units) were added sequentially and the reaction mixture (50 μl volume) was left at room temperature (22°C) for 2 to 3 hours or overnight depending on the amount of DNA to be labelled. Labelled probe was separated from unincorporated ^{32}P dCTP using a Sephadex-G50 (fine) column equilibrated with 150mM NaCl, 10mM EDTA, 5mM Tris-HCl (pH 7.5) and 10% SDS and 0.5 ml fractions collected. 1 μl aliquots of each fraction were subjected to scintillation counting in aqueous scintillation cocktail (Ecoscint, Koch Laboratories). Specific activities of between 5 x 10 ^{8} cpm/μg DNA and 1 x 10 ^{9} cpm/μg DNA were routinely prepared by this method.

2.9.17 Hybridisation of Probes to Filter-Bound DNA

Southern blots on nylon filters were processed by washing in solutions of 3x SSC for 1 hour , 6x SSC plus 5x Denhardts solution (0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin, 0.2% Ficoll) for 1 hour, and finally in 20 ml prehybridising solution (6x SSC, 5x Denhardts solution, 100 μg/ml sheared, denatured herring sperm DNA) for at least 2 hours at 65°C in a shaking waterbath or a Techne Hybridiser HB-1 hybridisation oven (Techne Cambridge Ltd, Duxford, Cambridge, UK). All prewashes and hybridisations were performed in double-sealed polythene bags or Techne siliconised glass hybridisation vessels. The isolated probe was denatured by boiling for 10 minutes, cooled in an ice bath and then added to the prehybridising solution and left overnight for hybridisation at 65°C. After hybridisation, the probe was recovered for later use and stored at -20°C. The filters were washed once each in solutions of 3x SSC for 30 minutes, 1x SSC for 30 minutes at 65°C and for the highest stringency, twice in 0.1x SSC for 15 minutes each, at 65°C. The filters were blot-dried between 2 layers of 3MM papers, sealed in clingfilm to keep them moist during exposure and autoradiographed.

2.9.18 Hybridisation of Probes to RNA Blots (Northern Blots)

Filters containing bound RNA were prehybridised in a solution containing 5x SSC, 5x Denhardt's solution, 50% deionised formamide, and 100 μg/ml of sheared and denatured herring sperm DNA for at least 2 hours at 42°C either in a polythene bag with constant shaking, or in a Techne hybridisation oven. Hybridisation was carried out essentially under similar conditions to prehybridisation with the

exception that 2x Denhardt's solution and 200 μg/ml herring sperm DNA was used. Probes were denatured by boiling for 5 minutes followed by rapid cooling in an icebath and then added to the hybridisation solution. Hybridisation was carried out at 42°C for 12 to 24 hours. Filters were sequentially washed with two washes of 10 minutes each in 3x SSC, 1x SSC, and 0.1x SSC containing 0.1% SDS at 60°C. The filters were sandwiched between two 3MM papers to remove excess washing buffer, sealed with clingfilm and autoradiographed.

2.9.19 Autoradiography

The approximate radioactivity (in cpm) bound to filters was determined using a Geiger counter to estimate the exposure time. Generally, radioactivity of about 600 cpm needed an exposure of 2 weeks. Pre-exposed X-ray film (Fuji RX-100, flashed twice with a calibrated photographic flashgun) was exposed to the filters in Kodak cassettes fitted with Kodak X-O-Matic (Regular) intensifying screens. All autoradiography was performed at −80°C.

2.9.20 Removal of Probes and Re-Use of DNA Blots

Nitrocellulose and nylon filters were kept moist by sealing them in clingfilm during exposure. Nitrocellulose filters for re-use were treated by submerging them in a solution of hot 0.1% SDS, boiled for 30 minutes and allowed to cool to room temperature. The filters were checked for residual radioactivity with a Geiger Counter. The above treatment was repeated until all detectable radioactivity was removed.

For HyBond nylon filters, an alternative treatment was employed to remove the bound radioactive probes. Filters were incubated at 45°C in 0.4M NaOH for 30 minutes, and then transferred to a solution containing 0.1x SSC, 0.1% SDS, and 0.2M Tris-HCl (pH 7.5) for a further 30 minutes of incubation at 65°C. The filters were checked for residual radioactivity and the treatment was repeated when necessary.

Once the probes were completely removed, the filters were prehybridised and hybridised with the new probes as described previously in Section 2.9.17.

2.9.21 Removal of Probes and Re-Use of RNA Blots

For RNA blots, probes were removed from the nylon filters by washing blots at 65°C for 2 hours in a solution containing 5mM Tris-HCl (pH 8.0), 2mM EDTA, and 0.1% Denhardt's solution. After the probe was removed, the filters were prehybridised, and hybridised as described in Section 2.9.17.

2.10 RECOMBINANT DNA TECHNIQUES

2.10.1 DNA Ligation

10 μg of plasmid pUC18 was digested with 4U/μg of appropriate restriction enzyme(s) for 3 to 6 hours to ensure complete digestion. A small aliquot was analysed on a minigel to monitor the digestion. Removal of 5'-phosphate groups from linearised plasmids with calf intestinal phosphatase was carried out according to Sambrook *et al.* (1989) to minimise self-religation. The reaction mixture was subjected to 2 phenol extractions, 3 chloroform extractions and finally ethanol precipitation. The pellet was resuspended in sterile distilled water for about 10 minutes in an icebath and immediately used or stored at −20°C.

DNA inserts with either compatible or blunt ends were ligated to the linearised plasmid DNA using T4 DNA ligase in a minimal volume of 10 to 20 μl with ligase buffer (20mM Tris-HCl (pH 7.6), 10mM $MgCl_2$ and 10mM, and 0.6mM freshly made ATP). The ligation reaction was carried out overnight at 16°C using 1U to 2U of enzyme per μg DNA. More enzyme was added for blunt-end ligations and for ligation of *EcoRI* -linkers to blunt-ended cDNA. A fresh aliquot of T4 DNA ligase and fresh ATP were added after 15 hours and incubation was continued at room temperature (22°C) for up to 6 to 12 hours.

2.10.2 Preparation of Competent Cells

For the construction of a cDNA library, commercially prepared, maximum efficiency DH-5α competent cells [greater than 10^9 transformants/ μg pUC18 (BRL Ltd)] were used for the transformations. In other transformation experiments, competent cells were prepared according to one of the two methods described below.

1. Polyethyleneglycol/Dimethysulfoxide Method

The method described was essentially that of Chung *et al.*, (1989). Briefly, 10 ml of overnight culture of DH-5α or other required strains was made by inoculating 10 ml of LB-medium with a loopful of the stock bacterial culture stored in glycerol as described in Section 2.8. 1 ml of the overnight culture was then used to inoculate 50 ml LB medium in a baffled flask and the bacteria allowed to grow to the exponential stage at 37°C with vigorous shaking. The OD of the bacterial suspension was checked every 30 minutes after about 1 hour and incubation was continued until the OD reached 0.2 at 650 nm on a spectrophotometer. The cells were pelleted by centrifuging at 4,200 rpm (900 x g) for 5 minutes. The supernatant was discarded and the pellet resuspended in 1/10 th of the original volume of the culture using ice-cold sterile transformation and storage solution (TSS) buffer pH 6.5 (LB-medium containing 10% (w/v) PEG 6000, 50mM $MgCl_2$ and 5% (v/v) DMSO; pH to 6.5). The dimethylsulfoxide (DMSO) was added after autoclaving. 100 to 200 μl of competent cells were used for every transformation reaction.

2. $CaCl_2$ Method

100 ml of LB-medium was inoculated with 1 ml of overnight culture of the desired strain of bacteria and shaken at 37°C until an OD_{650} of 0.2 to 0.3 was reached. Cells were then centrifuged at 4,000 x g for 10 minutes at 4°C and the pellets were resuspended in 0.5 volume of 0.05M ice-cold solution of $CaCl_2$. Tubes were left on ice for 10 minutes and recentrifuged at 4,000 x g for 20 minutes at 4°C. Pelleted bacterial cells were gently resuspended in 1/15 th of the original volume of ice-cold 0.05M $CaCl_2$ and left on ice for at least 30 minutes prior to use. 100 to 200μl of competent cells were used in the transformation.

2.10.3 Transformation of Competent *E. coli* cells with DNA

Ligation mixtures were diluted in sterile distilled water to desired concentrations and test transformations were carried out with several dilutions. Competent cells were added to tubes containing between 10 and 20 μl of the diluted ligation mixtures and stored in ice for 30 minutes. The cells were heat-shocked at 42°C for 4 minutes, then 1 ml of LB-medium added , mixed thoroughly, and incubated for

1 hour at 37°C with occasional mixing. 100 µl of the transformation suspension was spread on YT-agar plates containing 50 µg/ml ampicillin and 40 µg/ml X-gal and incubated overnight at 37°C. As controls, unligated plasmid DNA of known concentration and religated plasmid DNA were also transformed. An aliquot of the competent cells was also spread on YT-plates with and without 50 µg/ml of ampicillin to check viability and ampicillin sensitivity.

2.10.4 Colony Hybridisation

The procedure used was that described by Sambrook *et al.* (1989). Bacterial colonies were replica-plated onto two gridded nitrocellulose discs on selective YT-plates and grown overnight. The master plate was sealed with Nescofilm and stored inverted at 4°C. For cDNA clones stored in microtiter plates, a specially designed replicator fork was used to sample the bacterial suspensions in the wells and then gently applied onto the nitrocellulose filters (20 cm X 20 cm) on large square petri dishes containing selective YT-agar.

The filters containing the bacterial colonies were processed by placing them, colony-side-up, for 5 minutes each, on several layers of 3MM paper saturated with the following solutions: i) 10% (w/v) SDS, ii) 0.5M NaOH, 1.5M NaCl, iii) 1M Tris-HCl pH 7.5, iv) 3x SSC (0.045M Na-Citrate, 0.45M NaCl pH 7.0). The filters were air dried for at least 2 hours followed by baking at 80°C for 1 hour in a vacuum oven. The filters were then hybridised with suitable radioactive probes and positive clones were identified. The corresponding colonies from the master plate were subcultured onto fresh media and stored in glycerol as described in Section 2.8.

2.11 CONSTRUCTION OF A cDNA LIBRARY

2.11.1 Isolation of Polyadenylated (Poly-A^+) RNA

Isolation of polyadenylated (poly-A^+) rich RNA (mRNA) was carried out using a batch method utilising oligo-dT-cellulose exactly as described by Maniatis *et al.* (1982). Occassionally, a similar method of isolating poly-A^+ RNA using a spin–column (Boehringer Mannheim, W. Germany) was also used. In this method, a total RNA sample was treated in the same way as for the column

chromatography method as described by Maniatis *et al.* (1982). After the initial sample preparation, the RNA sample in a loading buffer (20 mM Tris-HCl (pH 7.6) 0.5 M NaCl, 1 mM EDTA, 0.1% SDS), was loaded onto the spin-column and was allowed to flow freely. The flowthrough was collected and reapplied to the column. All the subsequent washing steps were done by putting the column into a sterile, DEPC-treated 15 ml Corex tube centrifuge tube and the whole assembly was centrifuged for 2 minutes at 150 x g in a bench centrifuge. The column was then washed twice with 1 ml loading buffer, twice with 1 ml low-salt buffer (as loading buffer but with 0.1M NaCl), and finally the poly-A^+ RNA was eluted with 4 times 0.25 ml hot (65°C) elution buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.05% SDS). The eluant containing enriched poly-A^+ RNA was then repurified on a regenerated column by the same procedure. The poly-A^+ RNA eluants were pooled and ethanol-precipitated at -20°C overnight with the addition of 1/10th volume 3M sodium acetate (pH 5.2) and 1 μg glycogen as a carrier. The poly-A^+ RNA was then recovered by centrifuging at 10,000 rpm at 4°C and resuspended in sterile DEPC-treated water. The poly-A^+ mRNA was stored in liquid nitrogen until used.

2.11.2 Synthesis of Double-Stranded cDNAs

cDNA synthesis was done using a cDNA synthesis kit (Boehringer Mannheim) according to the manufacturer's instructions. A small amount (5 μCi) of radioactive ^{32}P-dCTP was added to an aliquot of the reaction mixtures for monitoring the first and the second-strand syntheses. Aliquots of the radioactive cDNA were subjected to trichloroacetic acid (TCA) precipitation according to Maniatis *et al.* (1982) to determine the amount of ^{32}P-dCTP incorporation, hence the estimation of the cDNA synthesised. In the second-strand synthesis, ^{32}P-dCTP was also added to the whole of the reaction mixture so as to facilitate the monitoring of the cDNA during subsequent manipulations. After the final step of the cDNA synthesis, the cDNA sample was phenol extracted twice, chloroform extracted 3 times, ethanol precipitated and washed. The phenolic phases were back-extracted with an equal volume of H_2O and the aqueous phases were pooled before chloroform extraction. The pellet was resuspended in sterile distilled water. The technique produced a blunt-ended cDNA. An aliquot of the radioactive cDNA was analysed

on a 1.5% (w/v) agarose gel and autoradiographed directly by placing the gel on a rigid support in a polybag and placing X-ray film directly on the wrapped gel.

2.11.3 Addition of Synthetic *Eco* RI Linkers

0.01U of phosphorylated *Eco* RI linkers (Pharmacia) were ligated to the blunt-ended cDNA using 2U of T4 DNA ligase at 15°C for 12 hours or overnight. The reaction mixture was then heat-treated at 70°C for 10 minutes to inactivate the T4 DNA ligase. After adjusting the buffer conditions the sample was subsequently digested with 120U of high-concentration *Eco* RI for at least 6 hours at 37°C. 2 ul of 0.5M EDTA pH 8.0 was added to stop the reaction and the sample extracted once with phenol and 3 times with chloroform. The sample was then applied to a Sepharose CL-4B column to separate the linkered cDNA from the digested linkers.

2.11.4 Sepharose CL-4B Gel Filtration Chromatography

Sepharose CL-4B, equilibrated in TE buffer (pH 8.0) containing 0.1M NaCl and 0.1% SDS was autoclaved and then loaded into a sterile 10 ml disposable pipette plugged with a small amount of siliconised glass wool. The column was first equilibrated with several column volumes of sterile TE buffer pH 7.6 containing 0.3M NaCl. The sample was loaded onto the column and 24 fractions of 0.5ml each were collected. The aliquots were then sampled and counted in a scintillation counter. Fractions containing the radioactive cDNA were pooled and ethanol-precipitated overnight at $-20°C$ with the addition of 1 μg glycogen.

2.11.5 Ligation of Linkered cDNA to pUC18 and Transformation

Eco RI-linkered cDNA was ligated to pUC18 linearised with *Eco* RI, using T4 DNA ligase as described in Section 2.10.1. After the ligation was completed, several dilutions of the ligation mixture were made. 20 ul of each ligation mixture dilution were used to transform DH-5α competent cells as described previously (Section 2.10.3). After the optimum dilution (lowest dilution giving the most transformants) was determined, the original ligation mixture was diluted accordingly and used in subsequent transformations. Maximum efficiency DH-5α (10^9transformants/μg plasmid DNA) was used in the transformation. The transformed bacterial suspension was spread onto selective agar plates and incubated

overnight at 37°C.

2.11.6 Selection and Storage of cDNA Clones

After incubation at 37°C, the plates containing the transformed bacterial colonies were then transferred to a 4°C refrigerator and further incubated for several hours or overnight to increase the intensity of the blue colonies, hence facilitating the selection of white cDNA clones. White colonies were then replated onto fresh selective plates, grown overnight at 37°C and reselected to minimise false positives before storage. The selected cDNA clones were stored in microtitre plates in glycerol as described in Section 2.8. A total of 1536 cDNA clones were stored in 16 microtitre plates. After subsequent screening, selected cDNA clones were also stored as described.

2.12 BACTERIOPHAGE M13 SUBCLONING

2.12.1 Ligation of DNA fragments to M13 vectors

DNA fragments to be cloned were isolated and purified as described in Section 2.9.15. The DNA fragments were resuspended in 10 μl TE buffer and used in the ligation. 250 ng of M13-mp18 or M13-mp19 cloning vector was digested with the appropriate restriction enzymes to generate compatible cohesive ends, purified by phenol-chloroform extractions, ethanol precipitated with the addition of 1 μg glycogen as carrier, and resuspended in 10 μl TE buffer. The DNA fragments were ligated to the M13 cloning vector according to the procedure outlined in Section 2.10.1.

2.12.2 Transformation of Bacteriophage M13 containing DNA inserts

Competent cells for the transformation of bacteriophage M13 was prepared as described in Section 2.10.2 using F' episome-containing *E. coli* strains (JM 101 or TG2). 1/50th and 1/100th dilutions of the ligation mixture were prepared in separate eppendorfs to which 50 to 100 μl competent cells were added, incubated on ice for 30 minutes and finally heat-shocked at 42°C for 2 minutes. 200μl of freshly prepared log-phase plating bacteria (JM 101 or TG2), 10 μl of stock IPTG (100 mM), and 20 μl of 10% (w/v) X-Gal in dimethylformamide (DMF)

were added to the diluted ligation mix and gently mixed. The whole mixture was then transferred to 3 ml of liquified top agar (YT-medium with 0.75% bactoagar) at 42°Cmixed and immediately poured onto LB-agar plate. The plate was then gently swirled around to evenly spread the top agar mixture, left to set at room temperature for about 10 minutes and incubated overnight at 37°C.

2.12.3 Minipreparation of Single-Stranded M13 DNA

50 μl of fresh overnight culture of either JM 101 or TG2 was inoculated into 24 ml of 2x YT-medium, mixed and aliquoted in 2 ml volumes into each of 12 McCartney bottles. Well isolated, white plaques were picked from several plates with sterile cocktail sticks and transfering them into McCartney bottles. The bottles were then incubated at 37°C for 6 to 8 hours. 1.5 ml of the culture was transferred into an eppendorf tube, and centrifuged for 5 minutes at 10,000 x g. The supernatant was transferred into a second tube, recentrifuged for a further 5 minutes, and finally transferred into a third tube. 200μl of solution containing 20% (v/v) PEG 6000 and 2.5M NaCl was added to 1 ml of the supernatant. The solution was mixed gently by inverting the tube several times and then incubated at room temperature (22°C) for 45 minutes. After the PEG precipitation step, the tube was centrifuged for 15 minutes at 10,000 x g, the supernatant was discarded and the pellet resuspended in 100 μl TE buffer. The sample was phenol-chloroform extracted several times and finally extracted with ether to remove traces of phenol and chloroform. The ether was removed by centrifuging the tubes containing the samples for 30 minutes in a vacuum centrifuge. The DNA was then precipitated with 100% ethanol as described in Section 2.9.2. and finally resuspended in 20μl of TE buffer.

1 μl of each single-stranded DNA (ssDNA) samples was analysed on a 1% (w/v)agarose gel as described in Section 2.9.11, with non-recombinant M13 standard as control.

2.12.4 Determination of the Orientation of M13 ssDNA

The orientation of the ssDNA samples isolated from M13 subcloning was carried out according to the procedure outlined in Sambrook *et al.* (1989) with modifications. Briefly, one recombinant was chosen at random and designated as

the reference recombinant. 1 μl each of a selection of other recombinant ssDNAs were transferred into separate eppendorf tubes containing 20 μl H_2O. Each tube then received 1 μl of the reference recombinant ssDNA and 2 μl of 2% (w/v) SDS. After vortexing, the tubes were incubated at 65°C for 5 minutes. 4 μl of 0.5M NaCl was added to each tube and the incubation continued for another 1 hour at 65°C. The samples were then electrophoresed as described in Section 2.9.11. After electrophoresis, the gel was examined for the presence of complementary forms of the DNA sequence as judged by the "figure-eight" forms which migrated slower than the unhybridised ssDNA (Sambrook *et al.* (1989).

2.13 DNA SEQUENCING

DNA sequencing was performed according to Sanger's dideoxy chain termination method (Sanger *et al.* , 1980) using an Applied Biosystem 373A Automatic DNA Sequencer using fluorescent M13 Universal primers (Courtesy of Miss J. Bryden). The automatic DNA sequencer was based on flourescent rather than autoradiographic detection of electrophoretically separated DNA fragment (Connell *et al.* (1987). Four dye-labeled M13 Universal primers, each paired with one of the four dideoxynucleoside triphosphate (ddNTPs) chain terminators, are used in the sequencing reaction. Each of the four ddNTPs are paired with different dyes so that the separate A,C,G and T reactions can be combined , co-electrophoresed in a single lane, and distinguished during electrophoresis by the colour of their fluorescence detected with a laser scanner. For single-stranded M13 DNA, complementary DNA samples (both +/− strands)were sequenced and as for double-stranded plasmid DNA the sequencing was carried out using M13 Universal as well as Reverse primers.

2.14 POLYMERASE CHAIN REACTION - PCR

2.14.1 Production of PCR Primers

All primers used for PCR reactions were synthesised using an Applied Biosystem 381A DNA Synthesiser (courtesy of Mr. J. Gilroy). Oligonucleotides synthesised with β-cyanoethylphosphoramidites were simultaneously decyanoethylated and cleaved from the support using ammonia. After removing the base-protecting

groups with fresh concentrated ammonium hydroxide at 55°C overnight, the solution containing the oligonucleotide was dried using a UNIVAP vacuum centrifuge (Uniscience), redissolved in distilled water, dried again, and finally stored dessicated at −20°C until used. Primers were then dissolved in sterile H_2O and its concentration was determined by spectrophotometry as described in Section 2.9.4.

2.14.2 DNA Amplification by PCR using Non-Homologous Primers

Non-homologous primers for *Vicia faba* L. unknown seed protein (courtesy of Dr. R. Bassüner) were used in PCR DNA amplification using *Brassica napus* L. genomic DNA as template. Amplification was performed in 100 μl volume containing reaction buffer (50 mM KCl, 5 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.01 mM EDTA, 0.1 mM DTT, 20 g/ml gelatin and 0.1% (v/v) Triton X-100), 1.25 mM each of dATP, dGTP, dTTP, and dCTP, 10 μg each of 5' and 3' primers, and 0.5μg template DNA. All the components except the template DNA were added into an eppendorf, mixed, spun for 10 seconds to collect the solution in the bottom of tube, and overlaid with mineral oil. Template DNA was then added to the reaction mixture under mineral oil and the PCR reaction initiated. Control reaction with no template DNA was also included.

Amplification was performed in a HyBaid Intelligent Heating Block (Hybaid Ltd, Teddington, Middlesex, UK) programmed for 29 cycles of 1 minute at 94°C, 1 minute at 55°C, 2 minutes at 72°C, followed by the final cycle of 2 minutes at 55°C and 9 minutes at 72°C. The amplification products were analysed on a 1.5% (w/v) agarose gel and the desired band was cut out, electroeluted, purified and finally ethanol precipitated as described in Section 2.9.2.

2.14.3 DNA Amplification by PCR using Arbitrary Primers

Amplification reactions were performed essentially as suggested by Williams *et al.* (1990) with minor modifications. The reaction buffer used was that supplied with the *Taq* DNA Polymerase (Promega) and its components described in the preceeding section (Section 2.14.3). The reaction was carried out in volumes of 50 μl with the addition of 100 μM each of dATP, dCTP, dGTP, and dTTP (Boehringer Mannheim), 0.2 μM primer, 25 ng of oilseed rape genomic DNA, and 0.5 to 1 unit of *Taq* DNA polymerase (Promega) in a HyBaid Intelligent Heating

Block programmed for 45 cycles of 1 minute at 94°C, 1 minute at 36°C, and 2 minutes at 72°C. Amplification products were analysed by electrophoresis in 1.4% (w/v) agarose gels and stained with EtBr. The gel was destained by submerging in several volumes of H_2O for 2 to 3 hours in the dark at 4°C before photographing.

Genomic DNAs from several phylogenetically related *Brassica sp.* and high and low glucosinolate varieties of oilseed rape were used as templates in the PCR reactions. A control reaction with no template DNA was also included in every PCR reaction run. In a typical experiment, a master reaction mixture containing all of the components except for the template DNA was made, appropriate volumes were aliquoted into eppendorf tubes and the reaction mixture overlaid with 25 μl of mineral oil. Template DNA samples of appropriate concentration (5 ng/μl) were separately added to the reaction mixture under the oil overlay and gently mixed by drawing the reaction mixture in and out of the pipette tips. This careful manipulation was carried out to minimise cross-contamination of template DNA. A control reaction, in which an equal volume of distilled water was used instead of template DNA, was also included in every set of amplifications. Silicone grease was applied to the outer surface of the eppendorf tubes to ensure efficient heat exchanges between the tubes and the heating block.

2.15 CALCULATION OF GENE COPY EQUIVALENTS

Gene copy equivalents for oilseed rape was calculated in order to determine the amount of plasmid DNA equivalent to 1 gene copy per μg oilseed rape genomic DNA. 1 mole of single copy gene is contained in the gram equivalent weight of the haploid genome of *Brassica napus*; i.e. 1.6 x 10^9 base pairs.

1 mole = 1.6 x 10^9 x 625g where MW_{ave} of 1 base pair = 625g

6 x 10^{23} sequences = 1.6 x 10^9 x 625 g where 6 x 10^{23} is Avo. No..

Weight of 1 sequence $= \frac{1.6 \times 10^9 \times 625}{6 \times 10^{23}}$

in 1 μg genomic DNA $= \frac{6 \times 10^{23} \times 10^{-6}}{1.6 \times 10^9 \times 625}$ **sequences**

For plasmid pRS1, size = 2686 + 180 = 2866

$$\textbf{As above, weight of 1 sequence} = \frac{2866 \times 625}{6 \times 10^{23}}$$

$$\frac{6 \times 10^{23} \times 10^{-6}}{1.6 \times 10^9 \times 625} \textbf{ sequences} = \frac{2866 \times 625}{6 \times 10^{23}} \times \frac{6 \times 10^{23} \times 10^{-6}}{1.6 \times 10^9 \times 625} \textbf{ g}$$

$$= 1.79 \text{ x } 10^{-12} \text{ g} = 1.79 \text{ pg}$$

Thus, 1.79 pg of plasmid pRS1 should contain the equivalent of 1 gene copy per 1 μg of genomic DNA.

CHAPTER III

RESULTS AND DISCUSSION

3.1 DNA ISOLATION FROM RAPE LEAVES

3.1.1 Collection of Seed Materials

Seeds of commercial varieties of oilseed rape were generously supplied by a number of seed companies and individuals, together with some details of their genetic/phenotypic characteristics; most importantly on their erucic acid and glucosinolate contents. Except with variety Pasha, which has orange flowers, most other lines do not have distinct morphological characteristics which can be used as markers. During experiments with growing plant material for DNA extractions, a close inspection of the floral morphologies in different varieties revealed distinct oblong shaped petals in the variety Brutor compared with the rounder petals found in other varieties. Moreover, in the variety Mikado, the young plants lacked bristles on their leaf stalks and stems. In these cases, further observations are needed with other plants to confirm these characters. The characters so far mentioned are in themselves insignificant if they cannot be genetically linked with useful characters of agronomic importance such as glucosinolate content of the varieties. Table 3.1 summarises the oilseed rape varieties used in the study with the available information about their genetic characteristics.

Based on the information gathered so far, variety Mikado is the only true breeding homozygous diploid line (HDL) which was bred from a cross of F_1 parents and produced by using colchicine to double the chromosome number of a naturally occuring haploid in the F_2 (Hughes, 1988). However, such extensive breeding manipulation to produce a line or variety may not be applied to other varieties. Some other varieties may be produced by the pedigree methods where plants with desirable characters were selected and inbred. Varieties which were produced by the pedigree method of breeding may still be partly heterozygous due to limited inbreeding. However, Thompson and Hughes (1988) stated that

Table 3.1 — Oilseed rape varieties and some of their genetic characteristics

Variety	Sowing[a]	Erucic/Glucosin. Cont.[b]	Other Characteristics
Ariana	Autumn	Double low	N/A[c]
Aztec	"	Single low	"
Bienvenu	"	"	"
Bolko	"	N/A	"
Bronowski	Spring	Single low	Low glucosin. gene
Brutor	"	"	Oblong petal
Comet	"	N/A	N/A
Drakkar	"	Double Low	"
Elin	"	N/A	"
Global	"	"	"
Hanna	"	"	"
Jantar	Autumn	"	"
Jet Neuf	"	Single low	"
Karat	Spring	Double low	"
Libravo	Autumn	"	"
Lictor	"	"	"
Liradona	"	"	"
Lirawell	"	"	"
Mikado	"	Single low	Leaf bristles, HDL[d]
Optima	Spring	Double low	N/A
Pasha	Autumn	Single low	Orange petal
Rafal	"	"	N/A
Tapidor	"	Double low	"
Topas	Spring	"	"
Willi	"	Single low	"

[a]Winter varieties (autumn sown) require vernalisation to induce flowering; no vernalisation required for spring varieties.

[b]Single low (0) denotes low erucic acid and high glucosinolate content in the seeds; double low (00) denotes low in both.

[c]N/A denotes information not available or no distinguishing features.

[d]HDL denotes homozygous diploid line.

lines obtained by this pedigree method will be fairly uniform in character by F_6 or F_7 generations. U.K. government regulations require breeders to ensure that the seeds for commercial purposes were obtained from a plot where oilseed rape has not been grown for at least 10 years and 'isolated' from other oilseed rape plots by a distance of at least 200 meters. This is to ensure the highest degree of homogeneity of the seed materials.

The question of how homogenous the seed materials are must be considered. It is extremely important in this study that the homogeneity of the seeds should be confirmed by DNA analyses from individual plants so that any polymorphisms observed are a true representation of a particular genetic line and are not due to underlying heterogeneity in its seed material. In this study, extraction of DNA for each variety was always carried out using leaf material collected from at least 10 individual rape plants of the same varieties. As observed in later experiments involving DNA extractions from different samples of leaf materials (eg. Sections 3.2.1 and later 3.2.4) the RFLP patterns observed using genomic DNAs from these varieties were consistent and reproducible for each variety. This therefore indicated that the seeds were relatively homogenous as was guaranteed by the seed suppliers who stated that the seed homogeneity was greater than 99%. As such the seed materials were therefore deemed suitable for the studies undertaken.

Several genetic crosses were made between varieties having phenotypic characters namely glucosinolate content (e.g. Topas (low) *vs* Willi (high)), flower colour (e.g. Ariana (yellow) *vs* Pasha (orange)), petal shape (e.g. Topas (round) *vs* Brutor (oblong)), and bristles (e.g. Libravo (absent) *vs* Mikado (present)). Except with crosses involving spring varieties (Topas, Brutor and Willi) all other crosses carried out involved two vernalisation steps; first for the parental varieties, and secondly for the F_1 generation; a process which took a long time to complete. Normally, the time taken from sowing the parental stocks, crossing, and harvesting dry, mature F_1 seeds was at least 6 months for spring varieties and 8 months (6 months and 2 months vernalisation) for winter varieties. So, the total process of obtaining F_2 seeds was 1 year for spring varieties and 1 year and 4 months for the winter varieties. Reciprocal crosses were carried out in all of the crossing experiments. It is extremely important to avoid cross contamination of progeny seeds by covering

the developing rape pods containing seeds and by continuous, careful inspection for the emergence of newly opened flowers after completing the crossing procedure.

During the course of the crossing experiment it was observed that F_1 seeds collected from each pod involved in the hand pollination were comparatively smaller in number when compared with naturally pollinated pods, i.e. flowers allowed to develop naturally. The manipulation involved during the hand pollination which involved forcing open flowers and emasculating developing anthers might somehow physically damage the developing pod or interfere with the natural pollination process and therefore cause lower numbesr of developing seeds. It is also possible that the pollination was carried out only once in hand pollination while in natural pollination, the pistils were pollinated almost constantly during the pollination period.

In earlier RFLP studies in plants and animals, segregation analysis of RFLP markers was routinely used in the construction of RFLP maps. The crosses chosen for the segregation analysis always include some scorable morphological or biochemical markers so as to facilitate the analysis. Therefore, RFLP studies in plant species require either breeding materials which involved multiple crosses with a pair of scoreable phenotypes as mentioned above or multiple samples of two particular phenotypes that were closely related as with high and low glucosinolate contents of rape seeds. Since oilseed rape is only a recently developed crop species, very few genetic characters were known and consequently their inheritance - not fully studied. At the start of this research programme, no classical genetic linkage maps were available for oilseed rape. Alternatively, the construction of molecular linkage map based on RFLP markers could certainly contribute to better understanding of oilseed rape genetics.

The Rape breeders' main objective today is to breed for low glucosinolate varieties and since these new varieties are of commercial values, the breeders, who are mainly affilliated to seed companies are not prepared to give breeding materials even for experimental purposes. They are, however, able to send seed samples of the released commercial varieties. As a result, a collection of high and low glucosinolate varieties were used in this research project to explore the possibility of finding RFLP marker(s) linked to glucosinolate content while studying the

incidence of RFLPs associated with rape extensin and self-incompatibility genes.

3.1.2 DNA Isolation and Analysis

The DNA isolation procedure used in this study gave relatively high molecular weight genomic DNA of between 40 and 50 kbp from leaf tissues of rape plants. The yield was however low compared with the yield from tobacco and averaged around 200 to 500 μg DNA per 5 g of leaf tissue. It was apparent that young rape leaf tissue gave a significantly higher yield of DNA compared with that of older leaf tissue. Though in general it is normal to have a lower yield with older leaf tissues, in oilseed rape plants, this problem is significant. In the isolation procedure, the genomic DNA was purified by ultracentrifugation in EtBr-CsCl gradient. During this purification, the EtBr was removed from the DNA samples using CsCl-saturated isoamyl alchohol instead of the more usual isopropanol because it was found that isoamyl alchohol did not remove as much water from the DNA solution thereby eliminating crystallisation of the CsCl during the extraction. The complete removal of the EtBr and CsCl from the DNA solution was necessary as it was found that the restrictibility was affected if the DNA samples were still contaminated with EtBr or CsCl.

The purity of the isolated genomic DNA was assessed firstly by spectrophotometry in which the A_{260}:A_{280} ratio was determined, and then secondly by its digestibility with restriction endonucleases. Such tests revealed that most rape DNA samples isolated by this method were clean and could be successfully restricted with a wide range of restriction enzymes. Occasionally, some of the extracted DNA samples showed a lower degree of restrictibility and these were re-extracted with phenol. After this treatment, the DNA samples were usually fully restrictable. Although estimates of the concentrations of the DNA samples were initially made using spectroscopy, these estimates were sometimes not sufficiently accurate, possibly due to contaminating RNA or other substances. The concentrations of DNA samples were therefore adjusted accordingly, based on the visual inspection of the restricted DNA samples on gels. In those experiments where it was necessary to know the exact amount of DNA present, the DABA assay was employed and found to be very accurate in determining the concentration of the

DNA samples. Great care was taken to ensure complete resuspension high molecular weight DNAs before sampling, so as to avoid inaccurate DNA estimations. Complete dispersion of such DNA usually took about 15 to 30 min if aided by a very gentle mixing to avoid shearing the DNA samples.

Rape genomic DNA digested with the restriction enzymes *Kpn* I and *Pst* I apparently showed a small degree of incomplete digestion as judged by a dense smear of high molecular weight DNA close to the origin of the electrophoresed DNA 'tracks' on agarose gels. When the digested DNA was transferred to nylon filters and probed, hybridisation to the DNA in this region was observed (Figure 3.2). No additional treatment of the DNA was successfull in removing this artifact. However, no bands were visible within this smear and the restrictable proportion of the DNA was apparently completely restrictable as judged by the absence of discrete partial digestion products. Furthermore, the results showed that polymorphisms could still be detected in spite of the smear of undigested DNA, confirming that these enzymes could still be used in the study. It was also observed that the variety-specific polymorphic patterns were consistent and reproducible. The consistency and reproducibility of the polymorphic patterns shown by *Kpn* I digestion will be discussed in Section 3.2.6.

All other enzymes used in the RFLP analyses were also found to generate consistent and reproducible polymorphic patterns for each variety of rape.

3.2 RFLP Analysis Using Characterised Rape cDNA clones

As an approach to identifying RFLPs and choosing enzymes most likely to reveal RFLPs, DNA samples from several rape varieties were each digested with the following restriction enzymes: *Bam* HI, *Eco* RI, *Eco* RV, *Hind* III, *Kpn* I, *Pst* I, *Xba* I, and *Xho* I. Digestions were carried out under conditions established to give complete digestion i.e 3U/μg of restriction enzyme for at least 6 hours or overnight with occasional gentle mixing. Restriction digests were electrophoresed on agarose gels and the DNA fragments were transferred to nylon filters.

A range of available rape cDNA clones were chosen to test the system and to reveal any RFLPs associated with these sequences. In the experiments, the following characterised, cloned rape cDNA sequences pRR566, pBOS2, pRR275A,

pN2, and pC1 were used as ^{32}P-labelled probes for the RFLP analysis. pRR566 and pRR275A are rape root-specific cDNA clones (generously supplied by Dr. I.M. Evans and L. Gatehouse) while pBOS2 is a cDNA sequence encoding the *S5* incompatibility allele in *B. oleraceae* (kindly supplied by Dr. C. Scutt). pN2 and pC1 are cDNA clones encoding the rape storage proteins napin and cruciferin respectively (generously supplied by Dr. M. Crouch). The characteristics associated with these clones are summarised in Table 3.2.

3.2.1 Details of Characterised Rape cDNA clones

3.2.1.1 Clone pRR566

The clone pRR566 is a cDNA which encodes a rape extensin protein found in abundance in root tissues (Evans *et al.*, 1990). Extensins are cell wall proteins rich in hydroxyproline which have been implicated in the extension and direction of cell wall growth. The 630 bp coding sequence was GC rich and arranged in repeats which encoded protein comprises of a highly repeated amino acid sequence mainly of two types:

Ser-$(Pro)_4$-Val-Tyr-Lys/His and

Ser-$(Pro)_4$-Lys-Lys-His-Tyr-Glu-Tyr-Lys/Asn

This highly repetitive nature of the extensin protein encoded by *extA* was the main feature of the clone. The presence of such repeated sequences in a gene allowed further variation in terms of number of repeats and variability within repeats, to be very readily generated (Evans *et al.*, 1990). Furthermore, this sequence variability might then be advantageous in generating different proteins which were adapted for various structural and storage roles (Evans *et al.*, 1990)

The highly repetitive nature of pRR566 which could generate variation within a gene was therefore chosen as a candidate for RFLP probes which had the potential of realising RFLPs amongst rape varieties.

3.2.1.2 Clone pBOS2

The clone pBOS2 is a stigma-specific cDNA clone from *Brassica oleraceae*

Table 3.2 — Summary of the characterised Brassica cDNA clones used in RFLP analysis

Clone Designation	Insert Size	Vector	Site	Specificity	Species
pRR566[a]	700 bp	pUC18	*Eco* RI	Root extensin, *extA*	*B. napus*
pBOS2[b]	1600 bp	pUC18	*Eco* RI	Self-incompatibility	*B. oleraceae*
pRR275A[c]	250 bp	pUC18	*Eco* RI	Unknown root specific cDNA	*B. napus*
pN2[d]	739 bp	pBR322	*Pst* I	Seed storage protein, napin	*B. napus*
pC1[e]	1620 bp	pBR322	*Pst* I	Seed storage protein, cruciferin	*B. napus*

[a] Evans *et al.*, 1990.

[b] Scutt , 1990.

[c] L. Gatehouse. Unpublished.

[d] Simons *et al.*, 1985.

[e] Crouch *et al.*, 1983.

which encodes a protein partially homologous to the known S-locus-specific glycoproteins (SLSGs) associated with self-incompatibility in *Brassica* (Scutt, 1990) . The so-called S-like sequence contained in pBOS2 was possibly a novel form of the SLSG gene corresponding to the S_5 allele (Scutt, 1990). The amino acid sequence of pBOS2 sequences contains 10 cysteine residues towards the C-terminus and also has putative N-glycosylation as well as O-glycosylation sites along the length of the protein. Cysteine residues present in the amino acid sequence could presumably be involved in disulphide bond formation or bonding between putative SLSG molecules to form covalently bound dimers or polymers (Scutt, 1990).

The Self-incompatibility (SI) system in *Brassica* is determined by a single polyallelic S-locus comprising of over fifty known alleles (Scutt, 1990). Scutt (1990) further added that the SLSG gene was S-linked and showed substantial sequence divergence between lines of different S-specificity and has demonstrated that RFLPs between two *B. oleraceae* varieties can be identified.

Due to the polyallelic nature of the S-locus and its ability to identify between two *B. oleraceae* lines of different S-alleles using RFLPs, pBOS2 was chosen for the RFLP studies in *B. napus*.

3.2.1.3 Other rape cDNA clones

Besides the two cDNA clones mentioned above, several other available rape cDNA clones were also used in the study. The first clone was pRR275A which was isolated from a rape root cDNA library. pRR275A showed comparatively higher expression in rape root than in any other rape tissues and hybridised to an mRNA of about 3 kb (L. Gatehouse, pers.comm.). The other two clones were storage protein cDNAs, pC1 and pN2 which encode cruciferin and napin respectively. Cruciferin is an 11-12S globulin composed of six 50 kilodaltons (kd) subunits, each subunit containing an interpeptide disulfide-bonded 26-30 kd (α) and 20-21 kd (β) polypeptide (Simon *et al.*, 1985). pC1 encodes one member of a family of cruciferin precursor polypeptides and contains a hydrophobic signal sequence, followed by the α and β polypeptides (Simon *et al.*, 1985). The other storage protein cDNA clone used was pN2 which encodes a small (1.7S), basic water-soluble protein called napin (Crouch *et al.*, 1983). Napins have an average molecular weight of 13 kd and are composed of a large and a small subunit (9 kd

and 4 kd, respectively) linked by disulfide bonds, which are derived from precursors that are cleaved post-translationally (Crouch *et al.*, 1983). Unlike legumin-like storage proteins, napins have extensive regions other than the signal peptides, which are removed prior to the formation of the mature protein, i.e. about 35% of the polypeptide precursor is not included in the final product (Crouch *et al.*, 1983).

3.2.2 Production of RFLP probe using PCR

An attempt has been made to isolate a rape clone based on PCR using non-homologous primers from *Vicia faba* which is specific to unidentified seed protein (Dr R. Bassuner). DNA fragment of approximately 700 bp long was successfully amplified from rape genomic DNA template. The fragment was subsequently cloned into pUC18 at *Eco* RI/*Hind* III sites as specified by the primers. Sequence analysis was carried out by subcloning the 700 bp insert of the clone, designated as pUSP1 (unidentified seed protein), into M13. The DNA sequence was then compared with the sequence of the USP clone of *Vicia faba*. No homology was found between the two sequences. There was however a short DNA sequence which had 100% homology to the primers used suggesting that the cloned insert was from a non-related DNA sequence with the priming sites (Data not presented). The clone was not used for subsequent RFLP analysis.

3.2.3 Preliminary Experiments on RFLPs

From preliminary experiments, it was found that pRR566 and pBOS2 were able to satisfactorily reveal RFLPs in a small number of DNA samples restricted with certain enzymes. Other clones used failed to give satisfactory RFLPs results (Data not shown). Clones pC1 and pN2 gave identical RFLP patterns in all the varieties while pRR275 gave high background hybridisation which impedes RFLP analysis when used as probes for the preliminary Southern blots. As a result, it was decided to investigate only the RFLPs associated with extensin, as revealed by probe pRR566, and also with *Brassica* self-incompatibility as revealed by probe pBOS2. The other clones which did not give satisfactory RFLPs results were not studied further. DNA samples from a total of 19 rape varieties were extracted and used in the experiments to explore further the feasibility of using RFLPs as revealed

by these two clones. Some of the results to demonstrate the RFLPs associated with each of these two clones are presented in the following sections (Sections 3.2.4 and 3.2.5). Throughout this chapter, the number in brackets immediately after the name of rape variety corresponds to the lane number assignment in the corresponding figure of EtBr-stained gel or the autoradiograph.

3.2.4 RFLPs Associated with Extensin Gene

When pRR566 was used as a probe, DNA samples from varieties digested with *Xba* I and *Kpn* I showed extensive RFLPs. These are shown in Figures 3.1, and 3.2 respectively. As indicated by the numbered arrows, numerous fragment length differences were observed at the positions shown. At each polymorphic band position indicated, the DNA fragment was present in some varieties but absent in others. A good example is shown at Position 8 in *Xba* I-digests (Figure 3.1) where varieties Comet (4), Elin (7), Lirawell (12), Optima (14), SV Karat (17), and Willi (19) contained a 6.6 kbp band but this was absent in all the other varieties.

Although *Xba* I-digested DNA samples probed with pRR566 were highly polymorphic, the patterns were very complex involving many scoreable polymorphic band positions. This made it difficult for the polymorphic patterns to be grouped for easy analysis. However, when the same DNA samples were digested with *Kpn* I and probed with pRR566, it was found that a comparatively simple polymorphic pattern was obtained. Figure 3.2 depicts the autoradiograph of *Kpn* I-digested DNA samples after probing with pRR566. As indicated by the arrows, only two polymorphic positions (Position 1 and 2) were found with this enzyme and hence it was easier for the patterns to be grouped. Upon close analysis of the *Kpn* I-produced RFLPs, it was observed that the patterns could be assigned to 4 groups: (1) varieties that contained a band at Position 1; eg. Lanes 3, 14, and 18, (2) varieties that contain a band at Position 2; eg. Lanes 1, 2, 5, 6, and 8., (3) varieties that contained both bands at Position 1 and 2, and (4) no band at either positions. Furthermore, it was interesting that the polymorphic patterns observed mimicked a situation of segregation analysis where the pattern in Lane 12 or possibly Lane 7 could have been produced by a cross between individuals from Groups 1 and 2, and hence these were heterozygotes whereas the rest of the

Figure 3.1 — RFLP patterns of Rape DNA samples digested with *Xba* I and probed with clone pRR566. Oilseed rape varieties Ariana (1), Bienvenu (2), Brutor (3), Comet (4), Cobra (5), Drakkar (6), Elin (7), Global (8), Hanna (9), Jet Neuf (10), Libravo (11), Lirawell (12), Mikado (13), Optima (14), Pasha (15), Rafal (16), SV Karat (17), Topas (18), and Willi (19) (Lane numbers indicated in brackets) were digested with *Xba* I (3U/μg), electrophoresed in 0.75% agarose gel, blotted, and probed with ^{32}P-labelled pRR566, a root-specific cDNA clones. The numbered arrows indicate polymorphic band positions. L, H, and U denote low, high and unknown glucosinolate contents in the seeds. Examples of RFLP patterns are shown by varieties Topas (Lane 18) and Willi (Lane 19) at Positions 3, 4, 5 and 6. A band at Position 5 is absent in variety Willi. The restriction fragment at Position 8 showes a good example of scorable RFLP band.

Figure 3.2 — RFLP pattern of rape DNA samples digested with *Kpn* I and probed with clone pRR566. The varieties and order of lanes are as indicated in Figure 3.1. L, H, and U denote low, high and unknown glucosinolate contents in seeds. The number of polymorphic band positions in this case is only two (Positions 1 and 2). The polymorphic bands can be classified into 4 RFLP patterns: (1) a band at Position 1 (eg. Lanes 14 and 18); (2) a band at Position 2 (Lanes 16 and 19); (3) two bands, one from each position (eg. Lanes 7 and 12); and (4) no band at both positions.

Figure 3.1

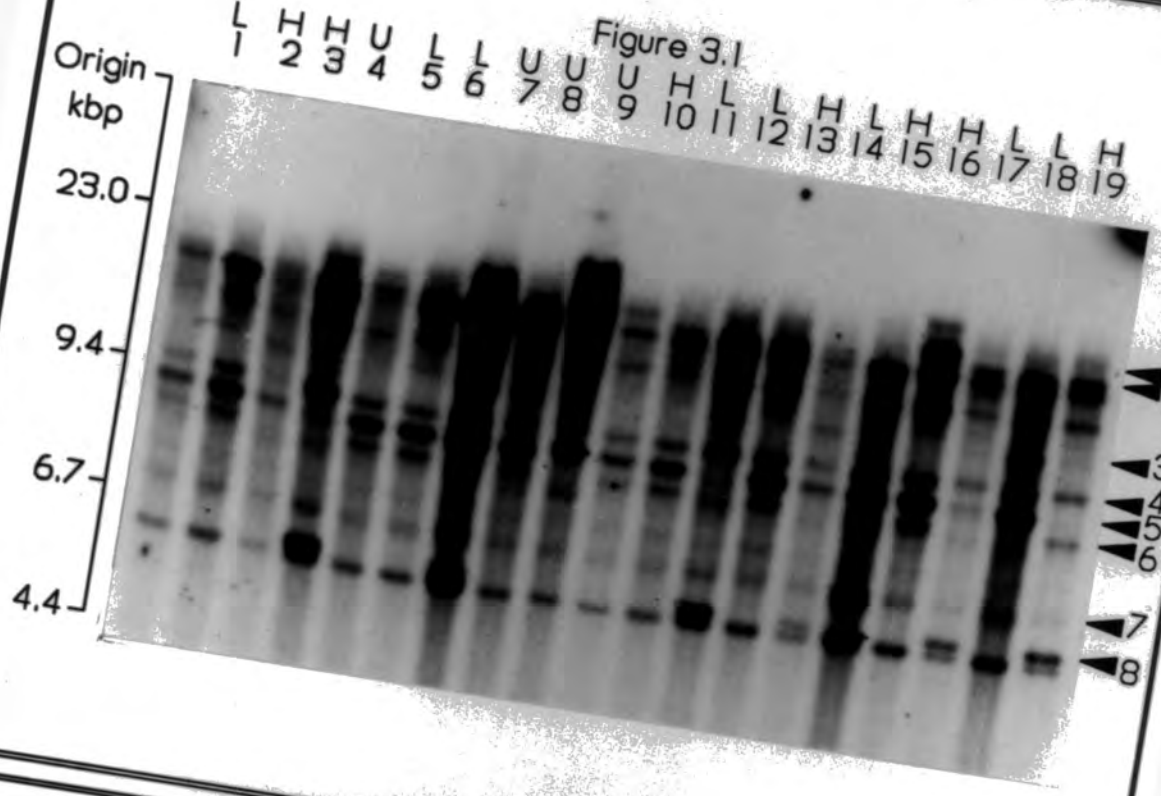
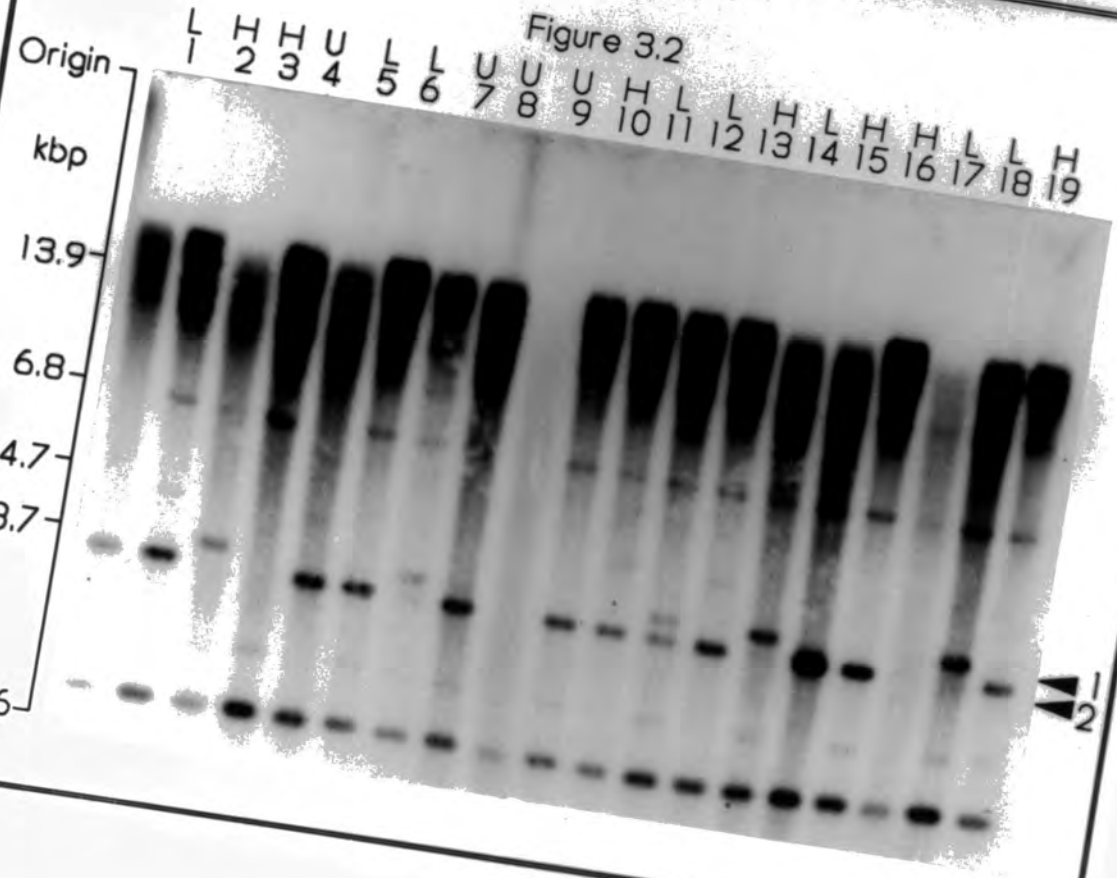


Figure 3.2



varieties were homozygotes of one or the other parental types. However, varieties that did not contain any band at these positions i.e. belonging to Group 4, eg. Lanes 4 and 17 could not be assigned to any of these categories. As suggested by Botstein *et al.* (1980), due to the lower number of polymorphic band patterns shown by the *Kpn* I-digested DNA samples, this enzyme/probe combination was suitable for RFLP analysis of segregating populations simplifying the analysis of F_1 and F_2 individuals. On the other hand, the highly polymorphic band patterns of the *Xba* I digests showing differences at many positions, may be useful for analyses in DNA fingerprinting (Botstein *et al.* (1980).

Upon closer examination of the polymorphic band patterns of each variety digested with *Xba* I and *Kpn* I, it was found that the group pattern in *Xba* I-digested DNA's showing the absence of a band (8.9 kbp) at Position 6 could be correlated with the Group 2 band pattern in *Kpn* I-digested DNA's that contained only the band (3.5 kbp) at Position 1. The presence of band patterns which could be correlated between the two digestions was probably due to the fact that the probe used (pRR566) was hybridising with identical genes in both cases and that the genes were on the same stretch of DNA in closely related varieties.

The absence of the prominent 'smear' at about 13.9 kbp region, characteristic of *Kpn* I-digested DNA and also other bands (except the 1.7 kbp) in Lane 9 was probably due to extensively degraded DNA samples possibly caused by DNase contamination.

While *Xba* I and *Kpn* I were able to generate RFLP band patterns when pRR566 was used as a probe, other enzymes used (*Eco* RI, *Eco* RV, *Xho* I, and *Pst* I) either failed to generate meaningful polymorphic band patterns or showed highly conserved band patterns between varieties. Examples of these analyses are shown in Figures 3.3 and 3.4 in which *Eco* RI and *Xho* I were used for the restrictions, respectively. In Figure 3.3, it was extremely difficult to analyse the *Eco* RI-generated band patterns as most of the bands were not discrete but were superimposed on a high background hybridisation which appeared as a smear.

In Figure 3.4, although the *Xho* I-generated band patterns (major bands) were very much simpler and hence this combination was an excellent potential candidate for RFLP analysis, they were, in general, highly conserved within most

varieties with the exception of three varieties which showed the absence of the 6.8 kbp band (i.e. Bienvenu (2), Pasha (15), and Rafal (16)) and three which showed the absence of the about 2.0 kbp band (i.e. Ariana (1), Cobra (5), and Hanna (9)). Faint minor bands that showed RFLPs were also observed in the region between 6.8 kbp and 13.9 kbp of the size marker especially in Lane 9 (variety Hanna) which had a more prominent minor band of size about 12.5 kbp and could be considered as variety-specific. Another good example of variety-specific band is shown by variety Pasha, a variety which has an orange flower instead of the normal yellow colour, which showed unique bands indicated by X and Y in Figure 3.3 and A in Figure 3.4.

3.2.5 RFLPs Associated with Self-Incompatibility Gene

A second probe that gave highly polymorphic band patterns was pBOS2. The best example of the use of pBOS2 to survey the extent of RFLPs in oilseed rape varieties was shown by enzyme *Bam* HI. Figure 3.5 shows the autoradiograph of the *Bam* HI-digested oilseed rape DNA samples from the 19 varieties, probed with pBOS2. As indicated by arrows, there were at least seven scorable positions found in the polymorphic band patterns. Two major polymorphic bands were observed at Positions 4 and 6. As was encountered earlier with the pRR566 probe used to analyse *Xba* I digested DNA samples, so with the pBOS2 probe it was extremely difficult to classify the polymorphic band patterns into groups so as to facilitate an RFLP analysis. Bands, unique only to some varieties, was observed at Position 3 and to a lesser extent at Position 2. It was found that the bands at these positions were present only in the five varieties, Pasha (15), Rafal (16), SV Karat (17), Topas (18), and Willi (19), respectively (Figure 3.5). There are no known genetic similarities shared between these varieties particularly in connection with self-incompatibility or floral development which could therefore be correlated with these unique bands. Oilseed rape is a self-compatible species. Close analysis of the RFLP patterns generated by *Bam* HI-digested DNA's also revealed no unique pattern associated with the glucosinolate content of any variety. Another enzyme that generated highly polymorphic band patterns when probed with pBOS2 was *Eco* RV. Figure 3.6 showed the autoradiograph of *Eco* RV-digested DNA from rape varieties grouped according to their known glucosinolate contents – Lanes 1 to 8 contained DNA samples from low glucosinolate varieties including the variety

Figure 3.3 — RFLP Patterns of Rape DNA Samples Digested with *Eco* RI and Probed with pRR566. Oilseed rape varieties Ariana (1), Bienvenu (2), Brutor (3), Comet (4), Cobra (5), Drakkar (6), Elin (7), Global (8), Hanna (9), Jet Neuf (10), Libravo (11), Lirawell (12), Mikado (13), Optima (14) , Pasha (15), Rafal (16), SV Karat (17) , Topas (18), and Willi (19) (Lane numbers indicated by brackets) were digested with *Eco* RI (3U/μg), electrophoresed in 0.75% agarose gel, Southern blotted, and probed with ^{32}P-labelled pRR566, a root-specific cDNA clones. *Eco* RI gave a complex RFLP pattern and as such made RFLP analysis difficult. However, the variety Pasha (Lane 15) showed unique bands not found in any other varieties as indicated in Figure 3.3 by 'X' and 'Y'.

Figure 3.4 — RFLP Patterns of Rape DNA Samples Digested with *Xho* I and probed with pRR566. The varieties and order of the lanes are exactly as described in Figure 3.3 (above) but without variety SV Karat. The major RFLP patterns generated are conserved in almost all varieties indicating that the enzyme *Xho* I is less suitable for RFLP study using pRR566 as probe. However, as with *Eco* RI (Figure 3.3, above), *Xho* I generated a unique band pattern in Pasha as indicated by 'A'.

Figure 3.3

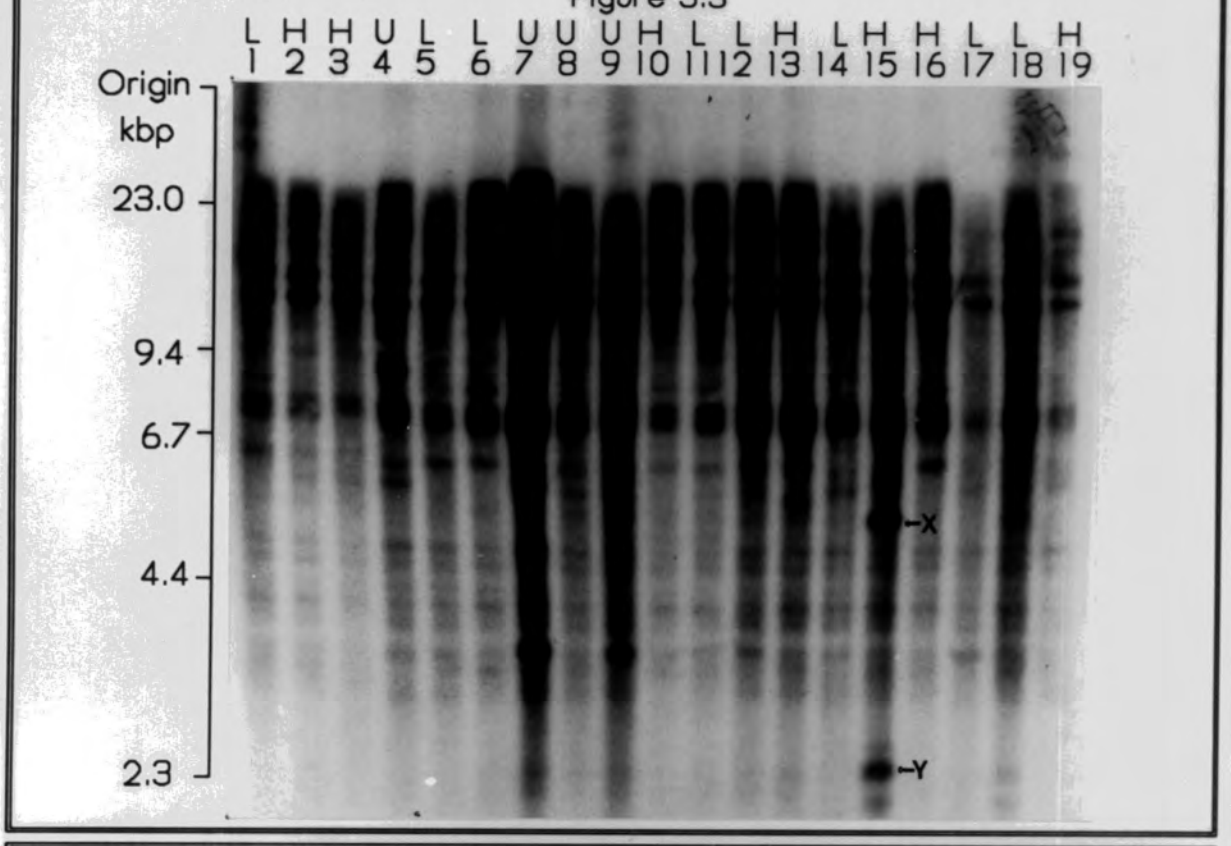
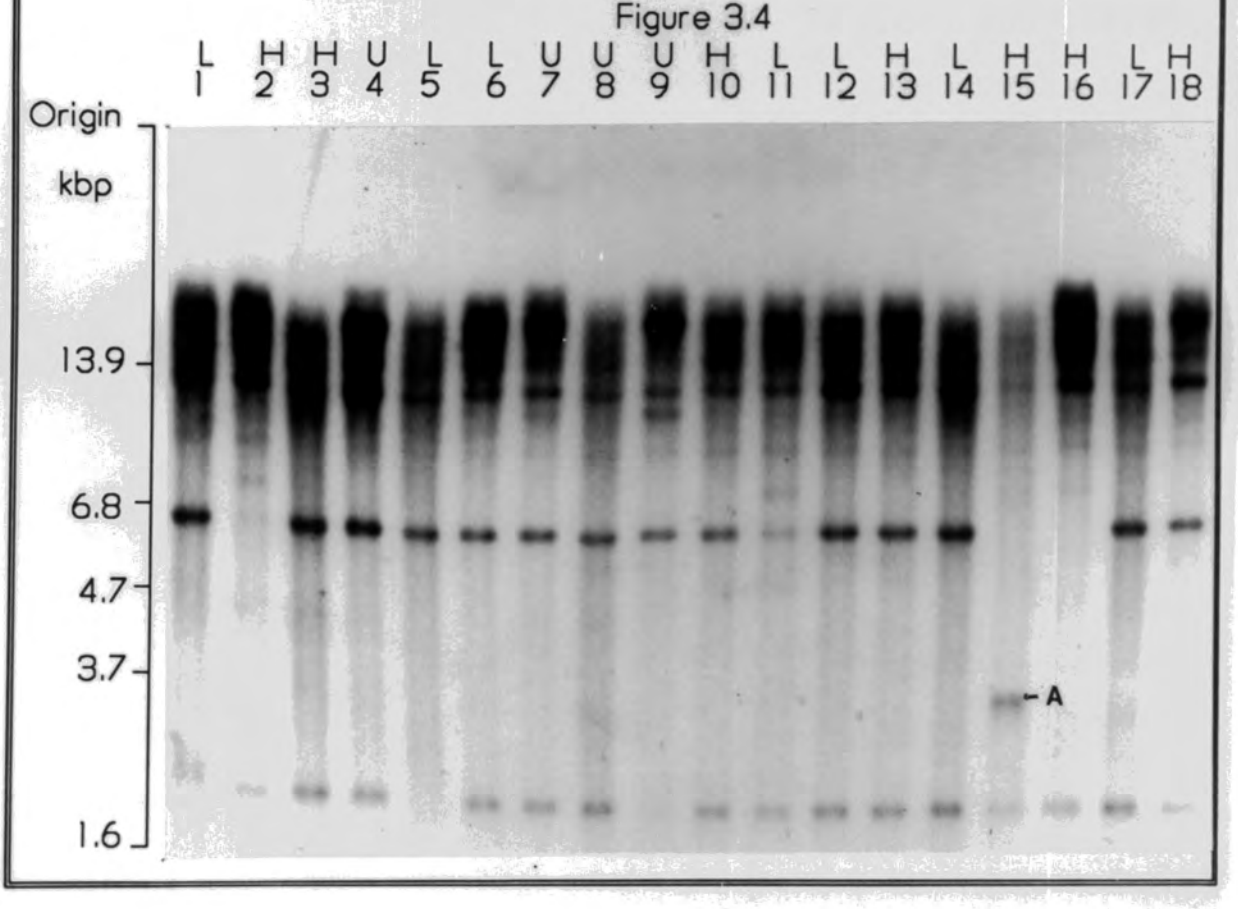


Figure 3.4



Bronowski (8), while Lanes 9 to 15 contained DNA samples from high glucosinolate varieties. Other rape varieties with no available information about their glucosinolate contents were not included in this particular analysis, which was to facilitate the comparative analysis of polymorphic patterns between high and low glucosinolate varieties. When the polymorphic band patterns were analysed, it was found that there no unique pattern was discernable which could be linked to either the high or low glucosinolate varieties. As was found with the *Bam* HI-digested DNA samples, *Eco* RV-digested DNA samples also generated band patterns which were polymorphic (Positions 2 and 4) and also produced variety-specific fragments (Positions 1, 3 and 5) as indicated by the arrows (Figure 3.6). An interesting band pattern was shown by the variety Bronowski (Lane 8, Figure 3.6), which is the single progenitor of all low glucosinolate lines of rape. All the bands shown in the other varieties, with the exception of the absence of band at Position 5 , were represented in the pattern generated by variety Bronowski.

Similar results were observed with other enzymes (*Eco* RI, *Kpn* I, and *Xba* I) which also showed polymorphic band pattern after probing with pBOS2. While pBOS2 showed highly polymorphic band patterns with certain enzymes, other enzymes (*Pst* I, *Hind* III, and *Xho* I) only gave very limited polymorphisms. An example of this latter type of enzyme analysis was *Xho* I as shown in Figure 3.7. Almost identical band patterns were observed in every variety analysed, indicating that *Xho* I was unsuitable for use in RFLP analyse s using pBOS2 as a probe.

From the other analyses using pBOS2 to detect RFLPs, it is concluded that due to the highly polymorphic band patterns generated by some of the enzymes, pBOS2 is more suitable for use as a probe for DNA fingerprinting using RFLPs than for segregation analysis for the RFLP molecular linkage map.

3.2.6 Assessment of the RFLP Analysis

This study has successfully showed extensin- and self-incompatibility-related RFLPs as revealed by their respective cDNA clones namely pRR566 (coding for extensin) and pBOS2 (coding for self-incompatibility allele) but failed to demonstrate unique band patterns which could be correlated with the glucosinolate content of oilseed rape. The failure of these RFLP analyses to demonstrate glucosinolate-linked band patterns was quite expected since the chances of having a charac-

Figure 3.5 — RFLP patterns of rape DNA samples digested with *Bam* HI and probed with pBOS2. Oilseed rape varieties Ariana (1), Bienvenu (2), Brutor (3), Comet (4), Cobra (5), Drakkar (6), Elin (7), Global (8), Hanna (9), Jet Neuf (10), Libravo (11), Lirawell (12), Mikado (13), Optima (14), Pasha (15), Rafal (16), SV Karat (17), Topas (18), and Willi (19) (lane numbers indicated in brackets) were digested with *Bam* HI (3U/μg), electrophoresed in 0.75% agarose gel, Southern blotted, and probed with ^{32}P-labelled pBOS2, a self-incompatibility cDNA clone from *B. oleracea*. The best example of polymorphic bands are shown at Positions 1, 4 and 6. Polymorphic bands at Positions 2 and 3 are only found in varieties Pasha (15), Rafal (16), SV Karat (17), Topas (18), and Willi (19) (Lanes 15 to 19).

Figure 3.6 — RFLP patterns of rape DNA samples digested with *Eco* RV and probed with pBOS2. The oilseed rape varieties are in this case Ariana (1), Cobra (2), Drakkar (3), Libravo (4), Lictor (5), Lirawell (6), Optima (7), and Bronowski (8) for the low glucosinolate varieties and Astec (9), Bienvenu (10), Brutor (11), Jet Neuf (12), Mikado (13), Pasha (14), and Rafal (15) for the high glucosinolate varieties. The varieties were grouped together in this analysis according to the glucosinolate contents to facilitate RFLP patterns comparison. No unique RFLP patterns could be correlated with glucosinolate content.

Figure 3.5

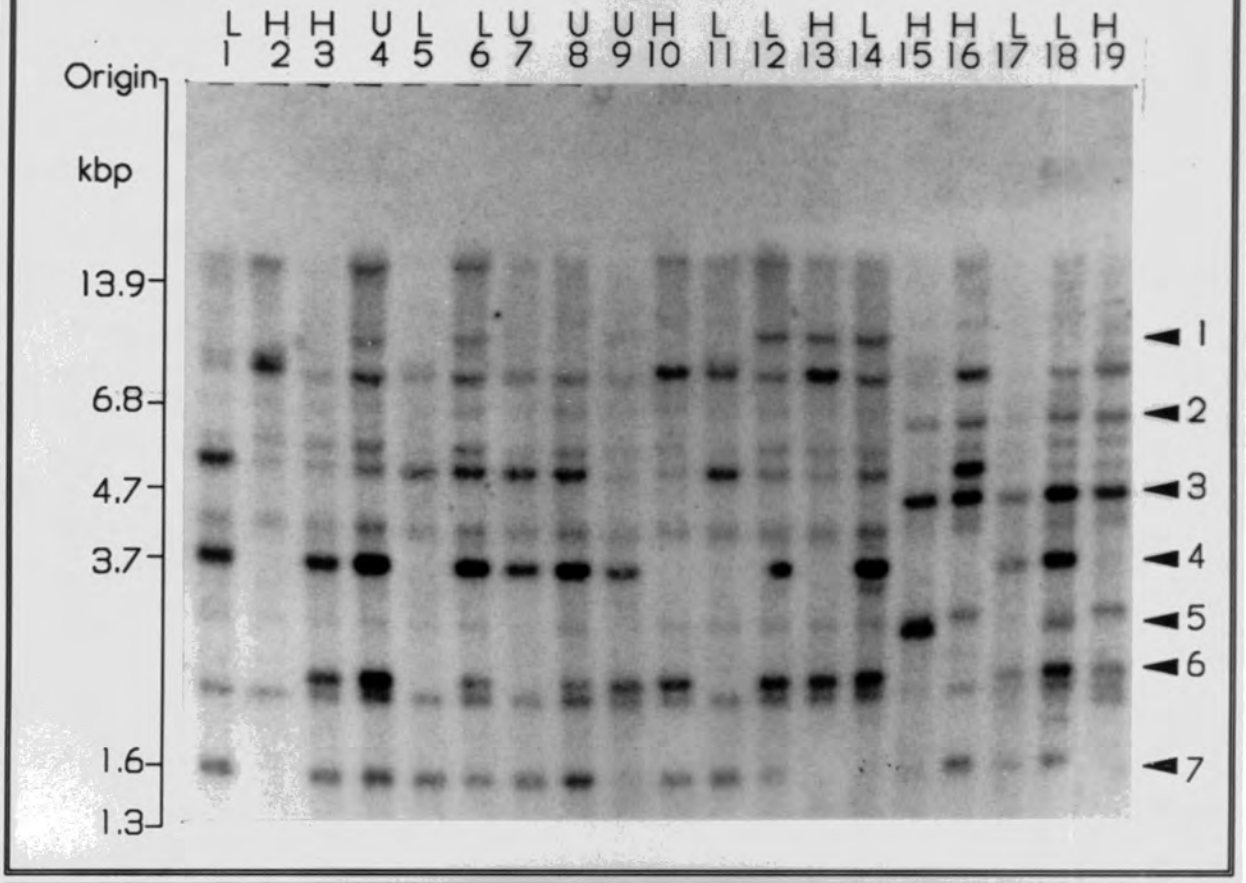


Figure 3.6

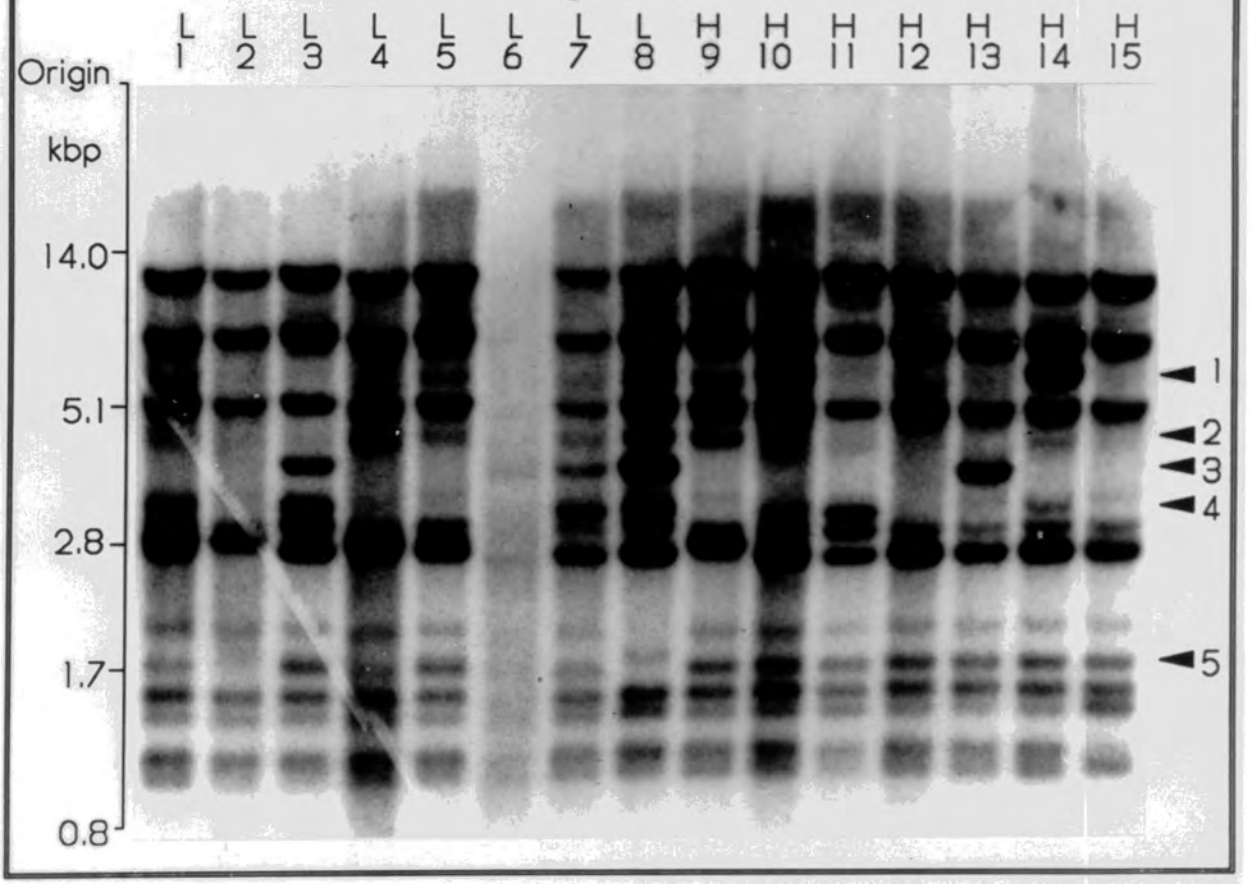


Figure 3.7 — RFLP patterns of rape DNA samples digested with *Xho* I and probed with pBOS2. Oilseed rape varieties Ariana (1), Cobra (2), Drakkar (3), Libravo (4), Lictor (5), Lirawell (6), Optima (7), Bronowski (8), for the low glucosinolate varieties and Astec (9), Bienvenu (10), Brutor (11), Jet Neuf (12), Mikado (13), Pasha (14), and Rafal (15) for the high glucosinolate varieties (lanes indicated by brackets) were digested with *Xho* I (3U/μg), electrophoresed in 0.75% agarose gel, blotted, and probed with ^{32}P-labelled pBOS2, a self-incompatibility gene in *B. oleraceae.* In contrast to the band patterns generated by *Bam* HI and *Eco* RV, shown in Figures 3.5 and 3.6, the *Xho* I-generated patterns were highly conserved.

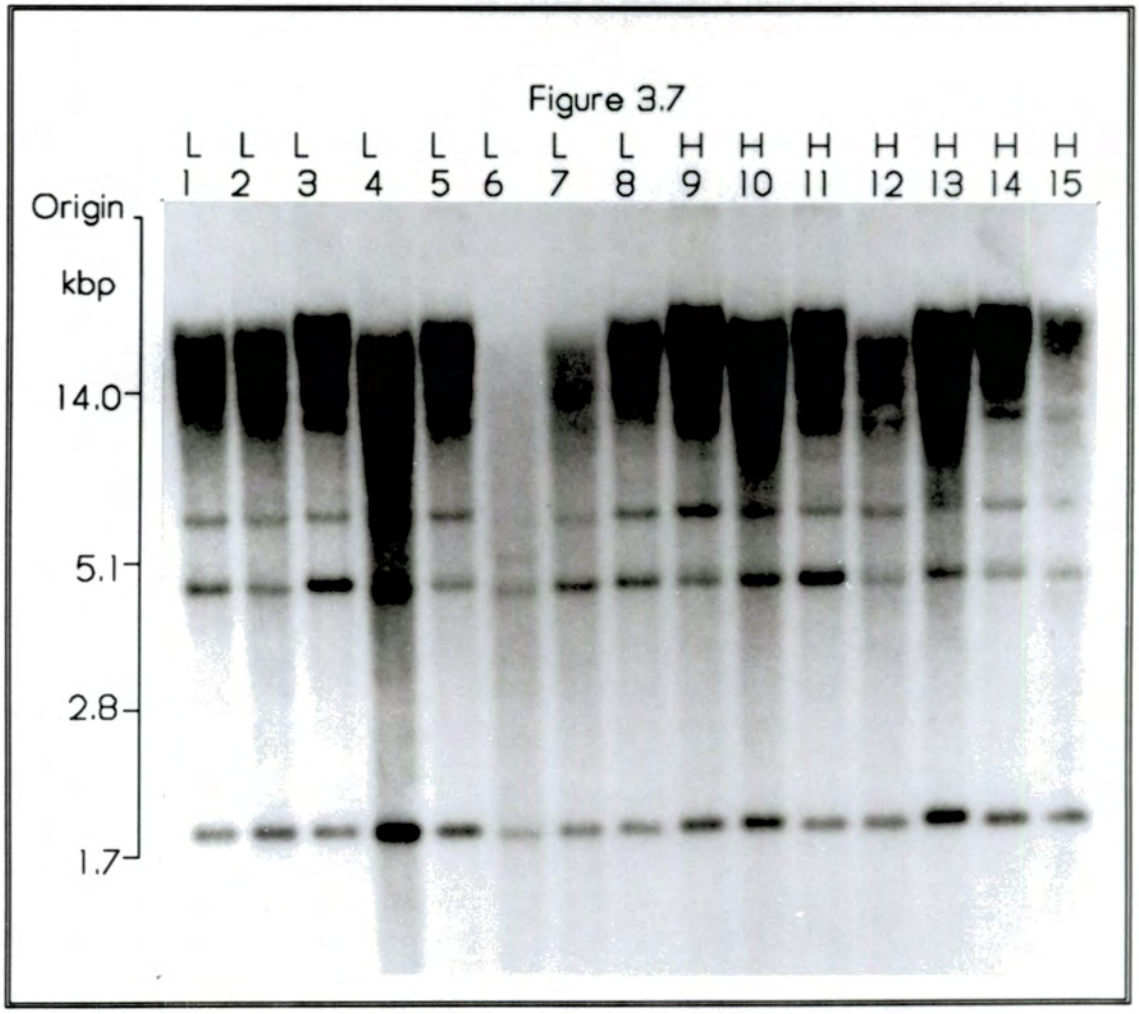


Figure 3.7
L 1
L 2
L 3
L 4
L 5
L 6
L 7
L 8
H 9
H 10
H 11
H 12
H 13
H 14
H 15
Origin
kbp
14.0
5.1
2.8
1.7

terised single copy gene linked to the genes controlling glucosinolate content in rape is extremely remote. Additionally, one of the cDNA clones chosen were cell wall -related and expressed mainly in roots (pRR566) while the other was re self-incompatibility-related and expressed in the stigma (pBOS2) and as such their associations with glucosinolate content in the seeds with complex biosynthesis are extremely unlikely.

There were, however, RFLP patterns which were unique to a particular variety which could be used in variety identification. An example of this unique, variety-specific band patterns was revealed by the probe pRR566 on *Xba* I-digested rape DNA's where variety Willi has bands at Positions 4 and 6 but none at Position 5 which almost all other varieties had (Figure 3.1). This pattern of bands was found in none of the other 18 varieties. The most striking example of a pattern unique to a particular genetic line was shown by the variety Pasha. Restrictions of Pasha DNA with all enzyme systems, except *Hind* III consistently showed either one or more extra bands with strong hybridisation when probed with pRR566 (See Figures 3.1 to 3.4, Lane 15). These observations were interesting as Pasha was the only rape variety analysed which has an orange petal colour as opposed to the usual yellow colour in other varieties. Perhaps a correlation between the unique patterns observed and the orange petal colour of variety Pasha may be revealed after suitable crossing studies. However, such a study while of genetic interest, would have limited or no agronomic importance and was not attempted in this programme.

From this study using pRR566 and pBOS2 as probes in the survey of the incidence of RFLPs in oilseed rape varieties, it can be concluded that with certain enzymes both probes could detect RFLPs of varying degree of polymorphism with pBOS2 showing a greater tendency to reveal polymorphisms. However, with other enzyme selected in the analysis, the polymorphisms were not so apparent or sometimes band patterns were highly conserved in most varieties. Therefore it can be concluded that selection of the best combination of probes and restriction enzymes was extremely important in RFLP studies.

3.2.7 Segregation Study of RFLPs Linked to Rape Extensin Genes

In an attempt to analyse the behavior of certain RFLP markers in a segrega-

tion study, two varieties, namely Topas and Willi were chosen as parental varieties in a cross. These two varieties showed different band patterns when digested with *Kpn* I and *Xba* I and probed with pRR566 (Figures 3.8 and 3.9, respectively). *Kpn* I was chosen because it has been shown that it can generate simple RFLP band patterns at two positions when probed with pRR566, while *Xba* I was chosen because of its moderately complex pattern when the same probe was used; thus enabling the analysis on the reproducibility of a moderately complex RFLP pattern in the segregation study. RFLP patterns generated by other enzymes and probed by pRR566 or pBOS2 were thought to be too complex to be analysed in a segregating polulation.

Due to the lack of appropriate breeding material, all crosses with the rape varieties had to be performed "in house". All crosses were performed in controlled environment growth rooms. The difficulties and the time taken to obtain breeding material has been discussed earlier in Section 3.1.1. Straight and reciprocal crosses were made between the two varieties using the methods described in Section 2.6 of the Materials and Methods. Straight crosses were carried out by taking pollen from Topas flowers and pollinating Willi flowers and the reciprocal cross was made by taking pollen from Willi flowers and pollinating the Topas flowers. F_1 seeds resulting from the crosses, were collected, sowed and their flowers were self-pollinated for the production of F_2 seeds. Each cross therefore produced two separate lots of F_1 and F_2 seeds. Collected F_2 seeds were sown and DNA was extracted from leaf materials of 12 F_2 individuals, 6 of which were from straight crosses and another 6 were from reciprocal crosses. DNA samples were also extracted from the parental varieties and also the self-pollinated F_1 individuals during the production of F_2 seeds. The DNA samples were then digested with *Kpn* I or *Xba* I, blotted, and probed with pRR566 and the autoradiographs are shown in Figures 3.8 and 3.9. In Figure 3.8, Lanes 1 and 2 represent parental varieties Topas and Willi respectively, Lanes 3 and 4 represent F_1 individuals from straight and reciprocal crosses respectively, and finally Lanes 5 to 10 and 11 to 16 represent F_2 individuals from straight and reciprocal crosses respectively.

As shown in Figure 3.8 (Lanes 1 and 2), the polymorphic patterns observed earlier (Section 3.2.1, Figure 3.2, Lanes 18 and 19) for the two parental varieties, Topas and Willi, were fully reproduced in this experiment. This result further

Figure 3.8 — Segregation of extensin RFLP patterns in crosses between two oilseed rape varieties revealed with the enzyme *Kpn* I. Straight and reciprocal crosses were carried out with rape varieties Topas and Willi. The two parents (Lanes 1 and 2 respectively), F_1 individuals from straight and reciprocal crosses (Lanes 3 and 4 respectively), F_2 individuals from straight crosses (Lanes 5 to 10), and reciprocal crosses (Lane 11 to 16) were digested with *Kpn* I and probed with pRR566. T, W, and H denote parental varieties Topas and Willi and "Heterozygous" (additive pattern of the two parents respectively). The RFLP patterns of F_2 individuals can be assigned to either one of the two parents (e.g. Topas (Lane 15) and Willi (Lane 14)) or as "heterozygous" (e.g. Lanes 5 and 9).

Figure 3.9 — Segregation of moderately complex extensin RFLP patterns in crosses between two rape varieties revealed with the enzyme *Xba* I . Lane assignment is exactly as described in Figure 3.8 above. Although the RFLP patterns are moderately complex, it is still possible to assign the genotype of the F_2 individuals by limiting the assessment of polymorphic bands to Positions 1, 2, and 3 as indicated by arrows. Only one individual (denoted by 'U' = Unknown (Lane 10)) could not be assigned to any genotype. This individual was assigned to Topas based on the above analysis (Figure 3.8).

Figure 3.8

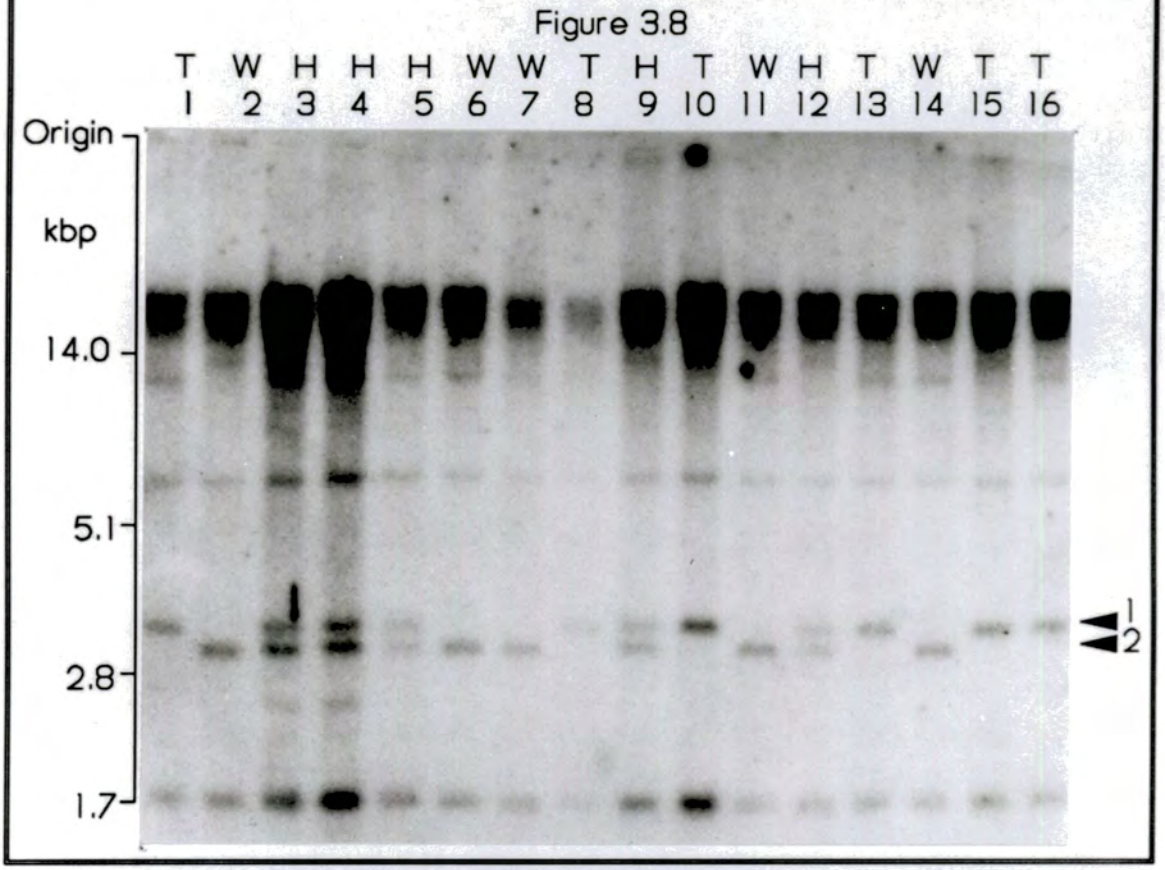
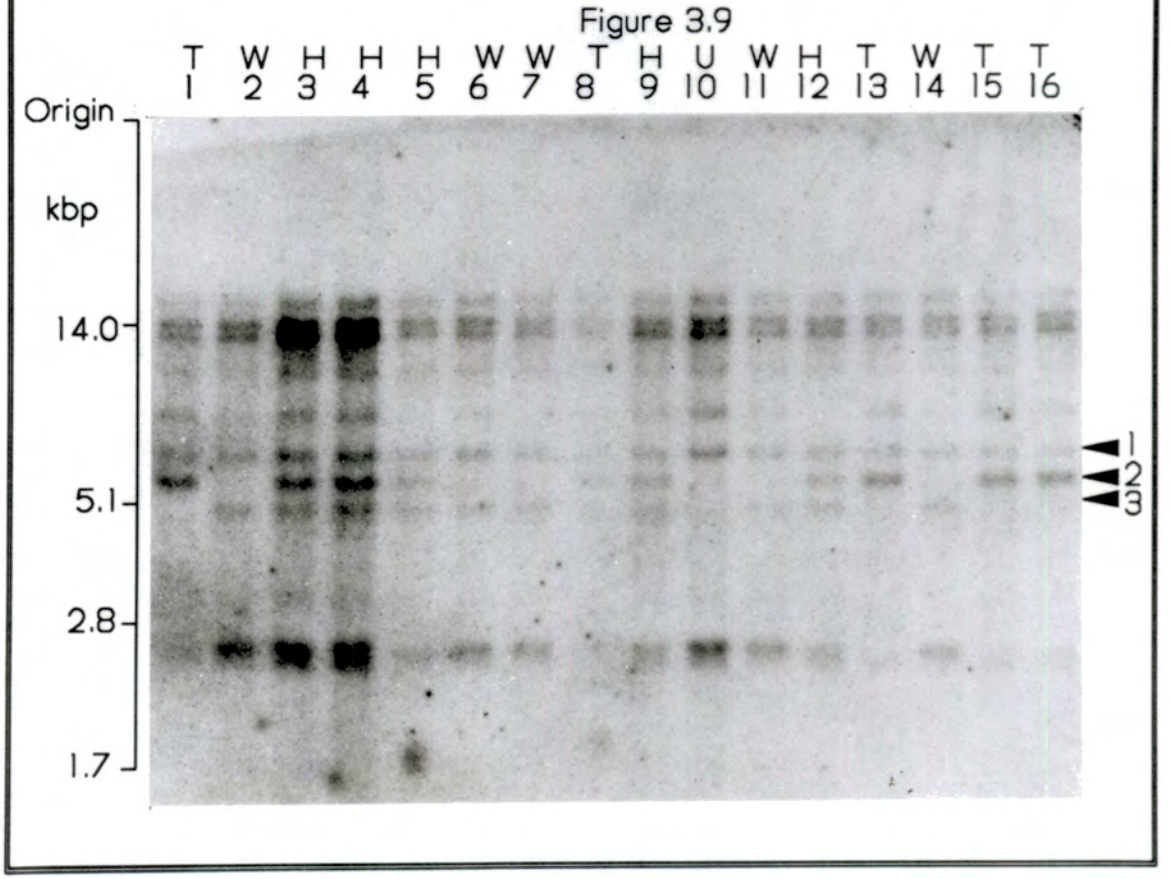


Figure 3.9



confirmed that the pattern generated by *Kpn* I digestion of DNA from these varieties was consistent and reproducible despite the apparent incomplete digestion of the DNA samples. When the polymorphic band patterns of the F_1 individuals were analysed (Lanes 3 and 4), it was found that the patterns for both individuals were identical and represented the additive band pattern ("heterozygous")of the parental varieties i.e. two bands, the top band (Position 1), was from Topas (Lane 1) and the bottom band was from Willi (Lane 2). Polymorphic band patterns from the two parental varieties and their additive pattern were observed in the F_2 individuals. There were five, four and three F_2 individuals which had patterns identical to Topas, Willi and "heterozygous" respectively. The number of F_2 individuals analysed were too small to be able to draw any conclusion about the ratio of homozygous and heterozygous individuals.

DNA from the same individuals from this segregating population, were also analysed using the enzyme *Xba* I, and probed with pRR566, a combination which reveals moderately complex RFLP patterns (Figure 3.9). *Xba*-I RFLP patterns observed were moderately complex compared with that of *Kpn* I; there were at least 7 major bands and 3 minor bands in the *Xba* I pattern as opposed to only 4 major bands and 2 minor bands in *Kpn* I digests. Analysis of polymorphic bands was limited to the 3 positions indicated by arrows for the purpose of simplification. Again, consistent with the previous results, polymorphic band patterns generated by the parental varieties were identical to those observed earlier in Section 3.2.4 (Figure 3.1, Lanes 18 and 19). When the F_1 individuals were analysed (Lanes 3 and 4), the expected additive pattern of the two parental varieties was also observed. When F_2 individuals were analysed (Lanes 5 to 16), it was found that all of them gave polymorphic band patterns which could be assigned to either of the two parental varieties or the additive pattern of heterozygous individuals with the exception of 1 individual (Lane 10). The assignment of the F_2 individuals to their respective identities corresponded exactly with that assigned earlier when *Kpn* I digests were analysed. The F_2 individual in Lane 10 that could not be identified fully on the basis of its *Xba* I band pattern was tentatively assigned to the parental variety Topas based on the pattern revealed in *Kpn* I digest which corresponded to that of Topas. One possible explanation for the different pattern of bands would be the loss of *Xba* I restriction site(s) that flank(s) the DNA fragments responsible for giving the RFLP band at Position 2 (Figure 3.9). As a result,

the DNA fragment probably remained attached to one of the higher molecular weight DNA fragments. In Lane 8 (Figure 3.9), the polymorphic pattern was very faint; however, upon prolonged re-exposure (result not presented), it was shown that the pattern was identical to that of Topas as found in *Kpn* I analysis. From these experiments it was shown that to a certain extent, complex polymorphic band patterns such as revealed by *Xba* I and pRR566 can still be useful in RFLP analysis if the scoring of polymorphic bands was selective and limited to only a few clearly defined positions; in this particular case, three positions were used.

3.2.8 Homozygosity and Heterozygosity in RFLP Analysis

The terms "homozygous" and "heterozygous" were used based on the polymorphic bands that were present in an individual. "Homozygous" F_2 individuals were those that possessed the parental-type polymorphic band patterns while "heterozygous" F_2 individuals were those that possessed the additive pattern of bands of t he two parents. "Homozygosity" and "heterozygosity" of the F_2 individuals in this case only applies to the RFLP patterns or markers generated by the enzymes *Kpn* I and *Xba* I when probed with pRR566. Since the number of F_2 individuals analysed for each cross was small, it was not possible to analyse the segregating population for Mendelian inheritance to give a true representation of the segregation of these markers. In a Mendelian inheritance in an independent segregation of the F_2 individuals, the ratio of homozygotes to heterozygotes is 1:1 derived from the 1:2:1 ratio corresponding to homozygous dominant:heterozygous:homozygous recessive. RFLP markers always behaved as simple Mendelian codominants in a segregating population (Botstein *et al.*, 1980).

3.2.9 Concluding Remarks

This study on the incidence of RFLPs in oilseed rape varieties has revealed that it was indeed possible to detect polymorphic band patterns amongst the different varieties using well characterised cDNA clones. Some of the polymorphic band patterns generated were simple with 1 or 2 polymorphic band positions while some were highly polymorphic and complex with at least 6 polymorphic positions. The observed RFLP patterns were greatly dependent on the choice of probes and restriction enzymes. One of the classic observations of RFLP analyses is when

a combination of probe and enzyme used in the analysis reveals DNA fragments of a specified length in one genetic line but one or more DNA fragments of a different length in another genetic line. For example, a single fragment revealed in one genetic line but two or more bands observed in another genetic line whose combined lengths equal the size of the fragment in the first line or *vice versa*, indicating the presence (acquisition) or absence (loss) of a restriction enzyme site within the sequence probed. This phenomenon, however, was not observed in any of the RFLP analyses using the characterised cDNA clones indicating that sites for the enzymes used lay outside the sequence probed.

As discussed earlier, this study was able to demonstrate RFLPs associated with the extensin gene family as revealed by pRR566 and the set of genes associated with self-incompatibility as revealed by pBOS2. The success of the cDNA clones, pRR566 in realising RFLPs in oilseed rape varieties was probably due to the involvement of extensin in many of the morphological characters such as pod length, height, node length, leaf shape, etc.. If the changes in the extensin genes in a variety are significant enough to cause variation in the morphological characters mentioned as may be selected in a breeding programme, then, these changes can also be detected in an RFLP study using pRR566.

The RFLPs observed when pBOS2 was used in the study can probably be explained by the fact that there are at least 50 self-incompatibility alleles in *Brassica sp.*. The presence of individual or combinations of allele(s) in a variety will generate unique RFLP patterns which will be revealed by pBOS2. Moreover, since the varieties used in this study came from different geographical regions and breeders, the variation with respect to self-incompatibility sequences is likely to be significant. Furthermore, there are several groups of genes that are highly homologous in a very complex gene system; all of which can be detected with pBOS2.

3.2.9.1 Selection of Probes and Enzymes

As has been shown in this study, the success or failure in revealing RFLPs greatly depended on the choice of the probe-enzyme combination in revealing RFLP among rape varieties. For example, pRR566 showed extensive hybridisation to DNA fragments digested with *Eco* RI and *Eco* RV; however, it was impossible

to extract clear , meaningful information on polymorphisms because of high background hybridisation even after washing to high stringency. Moreover, as shown in the autoradiographs, *Xba* I-digested samples (Figure 3.1) had more hybridising bands than in the *Kpn* I-digested samples (Figure 3.2) although the same probe (pRR566) was used in both. A similar situation was also observed when *Bam* HI- and *Eco* RV-digested DNA (Figures 3.5 and 3.6 respectively) gave more hybridising bands than *Xho* I-digested DNA. So, the best probe:enzyme combination is those that can give clear, scoreable polymorphic bands with low to medium pattern complexity (e.g. pRR566:*Kpn* I)

McCough *et al.*, (1988) suggested that the enzymes having 6 base-pair recognition site showed polymorphisms more frequently, than those having 4 base-pair recognition sites. The results from this study indicate that almost all the enzymes used, which have 6 base-pair recognition sites, can detect polymorphisms. Restriction enzymes recognising 4 base-pair restriction sites were not used in this study due to predicted highly complex patterns. It is possible that this observation was based on the difference in the frequency of cleavage of the 6 base-pair enzymes versus the 4 base-pair enzymes in the genomic DNA. Generally, 6 base-pair enzymes will most likely cleave the DNA to give longer lengths of fragment, and in many cases outside the gene; therefore these gave simpler band patterns. On the other hand, 4 base-pair enzymes are likely to cleave inside the gene; thus diminishing any size differences and most likely give complex patterns. However, further experiments comparing the 6 base-pair enzyme results with those using the 4 base-pair enzyme will be necessary to confirm McCough *et al.* (1988) suggestions.

3.2.9.2 Advantages of Using Nylon Filters

The use of nylon filters as the support for Southern blots for RFLP analyses has several advantages over conventional nitrocellulose filters. Firstly, the nylon is much stronger and non-brittle when dry, and it can therefore withstand repeated removal of probes for reprobing the blots; the analyses performed in this study involved the repeated use of the Southern blots. Secondly, the nylon is much more resistant to chemical attack and therefore dilute NaOH (0.4N) can be used to remove probes efficiently. It is necessary, however, to ensure that hybridised nylon filters are never allowed to dry out completely before the removal of a hybridised

probe. This is important to ensure a more complete removal of probes. Finally, the use of nylon filters, besides its ease of handling, is also essential to cut down the costs of screening, since in RFLP studies , several hundred clones may need to be screened for the required polymorphissm.

3.2.10 Problems Encountered and Their Possible Solutions

In the preliminary experiments, it was occasionally observed that there seemed to be an uneven hybridisation on all of the Southern blots although precautionary steps were taken to ensure an even distribution of the liquid containing the probe. Residual SDS from the solutions used to remove the probe was suspected of interfing re with the hybridisation of a second probe. So, repeated washing of the blots for at least three times with 3X SSC was carried prior to reprobing in an effort to remove any residual SDS on the blots; this apprared to alleviate the problem.

Another problem encountered during the course of this study was the long period of growth required for oilseed rape plants to reach the flowering stage especially with the winter varieties where vernalisation for at least one month was is required. It was found that if grown at higher number of plants per pot i.e 10 plants per 6-inch pot, rape plants flowered in a shorter time. This may be due to the competition for limited nutrients, which reduced the vegetative growth and thus forcing the plants to mature and flower earlier. So, it may be possible to reduce the flowering time using this method so as to facilitate crossing experiments. However, the process of "forcing" rape plants to mature and flower earlier using this method resulted in smaller, mature rape plants which produced only a small quantity of breeding material.

3.3 cDNA LIBRARY AS SOURCE OF RFLP PROBES

3.3.1 Rationale

As discussed in the introduction, the biosynthetic pathway of glucosinolates in oilseed rape is not fully understood. There is, however, a proposed biosynthetic pathway for glucosinolates, as suggested by Schnug (1990). The lack of information on the glucosinolate pathway is due to its extreme complexity involving many different enzymes and intermediary products. One of the objectives of this project was to try to find cDNA clones which gave unique RFLP patterns and to identify any which could be related to the glucosinolate content of different varieties. Strategically, cDNA clones from a cDNA library were chosen because these expressed sequences were found to be at least 2.5 times better in detecting RFLPs over sequences from a genomic library (Landry *et al.*, 1987). The strategy was based on the possibility that through differential screening, certain mRNAs coding for enzyme(s) responsible for the production of high glucosinolate levels in the high glucosinolate varieties could be identified and used as potential RFLP probes. As suggested by Josefsson (1973), and later confirmed by Underhill (1990), the low glucosinolate content in low glucosinolate varieties appeared to be induced by a metabolic blockage before the synthesis of 5-methylthiopentaldoximes in the pathway. Therefore, it can be speculated that enzymes catalysing the production of glucosinolates are present in high glucosinolate variety but due to the metabolic blockage one or more of the enzymes catalysing the subsequent steps after the synthesis of aldoximes, may be absent in the low glucosinolate variety. As for glucosinolates present in rapeseed, Lein (1972) and Austin *et al.* (1988) contended that alkenylglucosinolates (the major glucosinolates in seeds) which are accumulated in the seeds, are not synthesized *in situ* in the seeds but are instead made in the pod tissue and then transferred later to the seed during seed development. Based on this information, a cDNA library was constructed from mRNA isolated from pod tissues (including the seeds) of a high glucosinolate variety thereby maximising the chances of selecting high glucosinolate-specific sequences. The differential screening was performed firstly, by screening the library with total cDNA from the high glucosinolate variety, and secondly, screening with total cDNA from poly-A^+ mRNA isolated from pod tissues of a low glucosinolate variety at an identical stage of development.

3.3.2 Isolation of Total RNA from Rape Pod and Seed Tissues

Although it has been suggested that glucosinolates are synthesised in the pod tissues and later transferred into the seeds, this has not been confirmed yet. Consequently, to maximise the chances of selecting high glucosinolate-specific sequences, and also due to the difficulty of separating the seeds from the pods, especially those of the early stages of pod development, the seeds were included in the extraction of total RNA. Pod materials including developing seeds were collected from the high glucosinolate variety "Jet Neuf", at two-week intervals after anthesis until maturity but before the stage where the embryo has started to change colour from dark green to yellow (about 6-8 weeks). The pod materials were collected in liquid air and were stored, wrapped in aluminium foil, at $-80°C$. Equal weights of pod material from each stage of development were pooled and total RNA was extracted using a modified method as described in Section 2.9.7 of the Materials and Methods. Several RNA extraction methods were tested initially but none worked satisfactorily with rape pod material. The "hot-SDS" method (D. Bown, pers. comm.) was used but although this method gave high quality undegraded total RNA, the procedure was too time-consuming and the yield too low. A method by Logemann *et al.* (1987) which uses guanidinium hydrochloride was also tested and was found to extract a white gelatinous material, believed to be polysaccharide which coprecipitated with the RNA and which hindered its resuspension after ethanol precipitation and centrifugation. Another method of extraction described by Chomcynski and Saachi (1987) which uses guanidinium thiocyanate in the extraction buffer was also tested but also had the same problem of contaminating polysaccharide. Closer inspection of the developing seeds revealed that besides the developing embryo, a gelatinous material was found enclosed within the testa suggesting that the polysaccharide contamination may have come mainly from these developing seeds. During the extraction, a thin layer of fatty acids was also consistently observed on the surface of the extraction mixture. Therefore, the synthesis and accumulation of fatty acids in the developing rape seeds may also contribute to the difficulty in extracting total RNA from the pod tissues containing the seeds.

The quality of the RNA samples extracted from all the methods tested was high and although contaminated with polysaccharides, the RNA samples from the

two latter methods were of comparable quality when analysed on a denaturing formaldehyde agarose gel. However, it was found subsequently that RNA samples contaminated with polysaccharides gave poor quality, degraded poly-A^+ rich RNA. After several modifications and experimentations, a modified method based on that of Chomcynski and Saachi (1987) combined with that of Logemann *et al.* (1987) was found to work satisfactorily with rape pod material. It was also found later that the modified method could be used to extract total RNA from developing rape seeds satisfactorily. The method is described fully in the Materials and Methods. In order for this modified method to work satisfactorily, the volume of the extraction buffer (assuming 1 g/ml buffer) must be at least twice the weight of the tissue material. The ratio of material:buffer (by weight) should be 1:2.

3.3.3 Construction of cDNA library

Poly-A^+ RNA was isolated from total RNA extracted from pod material using an oligo-dT-cellulose (Boehringer Manheim) batch method as described in Maniatis *et al.* (1982). This method was chosen over the column methods for poly-A^+ RNA isolation because it requires minimal equipment and glassware and is a more rapid method (avoiding prolonged exposure of the RNA to room temperature), thus decreasing the chance of RNAse contamination. This method was used to isolate poly-A^+ RNA for the cDNA library. Subsequently, poly-A^+ RNA isolations were carried out with a spin-column method as suggested by Sambrook *et al.* (1989) which gave a much better quality and purer poly-A^+ RNA in a much shorter time than the batch method. This spin-column method was later modified by using a hot (65°C) elution buffer during the elution step (I.M. Evan, pers. comm.) which increased the yield of poly-A^+ rich RNA considerably.

As suggested by Maniatis *et al.* (1982), at least 1 to 2 μg of poly-A^+ RNA was needed to construct cDNA library. The proportion of poly A^+ RNA in a sample of total RNA was only about 1% (Sambrook *et al.* 1989). From the experiment using the batch method, the percentage of recovery of poly-A^+ rich RNA was only about 60% of the expected total poly-A^+ contained in a sample of total RNA (i.e. assuming 1% of the total RNA). Therefore, about 500 μg to 1 mg of total RNA was routinely used in the poly-A^+ RNA isolation which yielded about 3 to 6 ug (60% of 5 to 10 μg of the expected total poly-A^+ RNA contained in 500μg to 1

mg of total RNA). 1 to 2 μg of glycogen was always added as a carrier during ethanol percipitation. The isolated poly-A^+ RNA was analysed by scanning on a spectrophotometer to estimate the concentration and to determine the quality and finally stored under liquid air until used.

A cDNA synthesis kit (Boehringer Mannheim) was used for the synthesis of double stranded, blunt-end cDNA. ^{32}P-dCTP was added to the reaction mixture during the first- and second-strand synthesis to monitor the progress of the syntheses and purification of the cDNA. In the first-strand synthesis, radioactive dCTP was added to a small aliquot (1/10th volume) of the reaction mixture in a separate tube and incubated together with the main reaction mixture (remaining 9/10th volume). After completion of the first-strand synthesis, the small aliquot containing ^{32}P-dCTP was used to estimate the percentage of incorporation. The main first-strand reaction mixture were subsequently used for the second-strand synthesis. Similarly, radioactive dCTP was also added to the total reaction mixture to estimate the percentage of incorporation of the second-strand synthesis and to facilitate the monitoring of the synthesised cDNA during subsequent manipulations. Control mRNA supplied by the manufacturer was also used in parallel during the synthesis as a control for the cDNA synthesis. The percentage of incorporation of the ^{32}P-CTP was determined for both first- and second-strand synthesis by Trichloroacetic acid (TCA) precipitation as described in Maniatis *et al.* (1982). Calculation of the yield of the first- and second-strand syntheses was performed as described by the manufacturer. The yield of total double-stranded cDNA synthesised in a reaction can be estimated as equivalent to the yield of the second-strand cDNA, and calculated as the percentage of transcription of the poly-A^+ RNA into double-stranded cDNA. The percentage of transcription routinely obtained was between 20 to 25%. In the construction of the cDNA library, 0.26 μg of cDNA was sucessfully synthesised from 1 μg of poly-A^+ giving a yield percentage of 26%. According to the manufacturer, about 15 to 30% cDNA was obtained for each reaction depending on the quality of the poly-A^+ used in the synthesis. The synthesized cDNA was phenol extracted and ethanol precipitated twice; again with added 1 to 2 μg of glycogen during ethanol precipitation. A small aliquot of the cDNA was saved for the preparation of probes for differential screening. 1/10th volume of the synthesized cDNA was electrophoresed on a thin, ordinary, non-denaturing 1.5% agarose gel and directly autoradiographed. The

autoradiograph is shown in Figure 3.10. The cDNA appears as a smear of up to 2.0 kbp in length. Unincorporated ^{32}P-dCTP appears as a huge smudge at the low molecular weight area of the track.

Phosphorylated *Eco* RI linkers (Pharmacia) were ligated to the blunt-end cDNA with excess T-4 DNA ligase (3U/10 μl reaction mixture) to ensure good ligation of the linkers to both end of the cDNA. The linkered cDNA was subsequently digested with *Eco* RI and the digested linkers were separated from the cDNA in a Sepharose CL-4B column inside a 4°C coldroom. It was found that small air bubbles tended to form in the column if the separation was done at room temperature.

Eco RI digested pUC18 was first dephosphorylated to minimise self-ligation. The linkered cDNA was then cloned into pUC18. Commercially prepared Maximum Efficiency DH5α (BRL) was used in the transformation of the cloned pUC18 with first etablishing the optimum dilution of the ligation mixture. The calculated cloning efficiency was about 6 x $10^8/\mu$g of the cloned pUC18 which was close to that reported by the manufacturer ($10^9/\mu$g pUC18). White transformants were selected and replated again to minimise false positive clones before being reselected and stored in the 96-well microtiter plates. A total of 1536 cDNA clones were selected and stored in 16 microtiter plates at -30°C freezer. Microtiter plates were used as a convenient way of storing individual cDNA clones of the cDNA library in glycerol and also to facilitate the sampling of the cDNA clones using a specially designed replicating forks for rapid production of replica plates or for inoculating sets of identical nitrocellulose filters for colony hybridisation. The replicating fork and its use has been described earlier in the Materials and Methods. The use of microtiter plates and replicating forks also has an added advantage of inoculating bacterial colony of each cDNA clones in a regular "grid" (see Figure 3.11) on each sets of duplicate nitrocellulose filters thus greatly facilitating the analysis of data of each clone from differential screening.

Two clones from each plate were randomly selected and used in the plasmid minipreparations. Plasmids isolated from these clones were digested with *Eco*RI and electrophoresed on a 0.75% agarose gel. Figure 3.11 shows one of the gels with 18 cDNA clones. As exemplified in Figure 3.11, almost all clones contained inserts

which ranged in size from 0.2 to 1 kbp. The average size of the cDNA inserts was 0.8 kbp (800 bp). Three of the 32 cDNA clones analysed at random did not have any inserts while one of the minipreparations did not contain any plasmid. Therefore, it can be estimated that about 87% of the clones in the cDNA library contained inserts. During the analysis of the cDNA clones, an extra "DNA band" (with varying sizes of between 1.5 to 1.7 kbp) was observed in each of Lanes 2, 3, 6, 7, 9, 10 and 11. These clones were subjected to restriction with another single-cutting enzyme, *Hind* III, and electrophoresed, together with unrestricted plasmids from each clone in an attempt to determine the actual size of the insert (Data not shown). The results showed that the insert size of each clone was identical to that previously determined using *Eco* RI, the cloning enzyme. Therefore, the extra "DNA band" observed in these cDNA clones was an experimental artifact which probably resulted, from poor plasmid preparations and not resulted from the presence of double *Eco* RI inserts.

3.3.4 Analysis of the cDNA Library

The cDNA clones for colony hybridization were prepared by growing on nitrocellulose filters. Several sets of identical filters containing the clones were prepared. In order to confirm that the inserts were from the synthesized cDNA and at the same time survey their relative abundancy, the filters containing the cDNA clones were probed with total cDNA aliquot saved after the synthesis. 1536 cDNA clones (whole library) were screened. A representative autoradiograph of the filters from the first screening is shown in Figure 3.12. The cDNA synthesised from RNA isolated from pod tissue of a high glucosinolate variety hybridized to the cDNA clones and showed a wide range of relative abundance (a, b, and c; Figure 3.12). Some of the clones were highly abundant while others appeared to show no hybridization and therefore probably contained inserts with very low relative abundance (Evans *et al.*, 1990) or no inserts at all. Examples of clones with their relative abundance classification are indicated in Figure 3.12. Relative abundance of each cDNA clone compared with those having the lowest hybridisation signal ("background hybridisation") in each autoradiograph was scored by assigning symbols (+++) for the high abundance and (+) for the low abundance. The scoring was carried out in this manner where each clone was scored relative to the overall background hybridization so as to increase the accuracy of assigning

Figure 3.10 — Autoradiograph of ^{32}P-labelled cDNA after its synthesis. An aliquot of the newly synthesised ^{32}P-labelled cDNA was run on an ordinary, non-denaturing, 1.5% (w/v) agarose gel and was autoradiographed by directly placing X-Ray film onto the frozen gel wrapped in Saran wrap. Unincorporated ^{32}P-dCTP is indicated by an arrow.

Figure 3.11 — cDNA clones from the pod cDNA library digested with *Eco* RI. The size of the cloned inserts ranged from 0.2 kb to 1.0 kb and averaged about 0.8 kb. Lane 1 is linearised pUC18 while Lanes 2 to 19 are the cDNA clones. No inserts were found in Lane 5 while no plasmids were found in Lane 12.

Figure 3.12 — Primary screening of the cDNA library with pod cDNA as probe. Plates A and B represent typical autoradiographs of the filters after colony hybridisation of cDNA clones from two independent 96-well microtiter plates and inoculated onto the filters using the specially designed replicating fork to give a uniform pattern of bacterial colonies. The pod cDNA used as probe for the screening was identical to that used in the construction of the library. Relative abundancy of the clones was assessed by intensity of hybridisation of probe as indicated by a, b and c which correspond to highly abundant (+++), moderately abundant (++) and low abundant (+) relative to the clones having the lowest hybridisation (background) as indicated by w, y, and z.

Figure 3.10

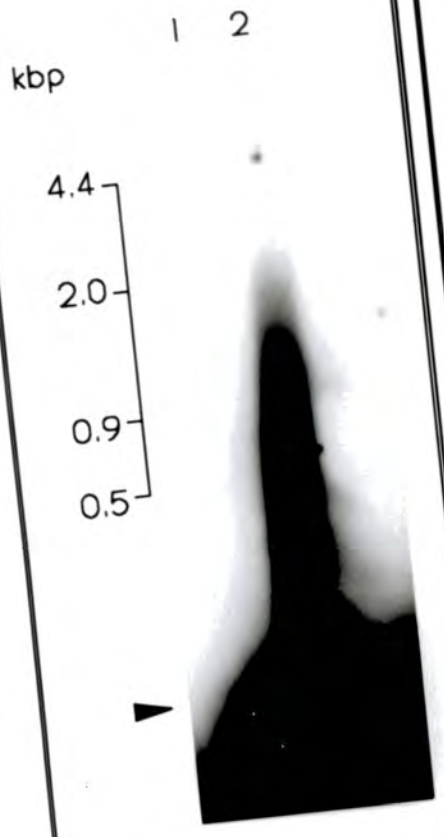


Figure 3.12

PROBE-Pod cDNA (High)

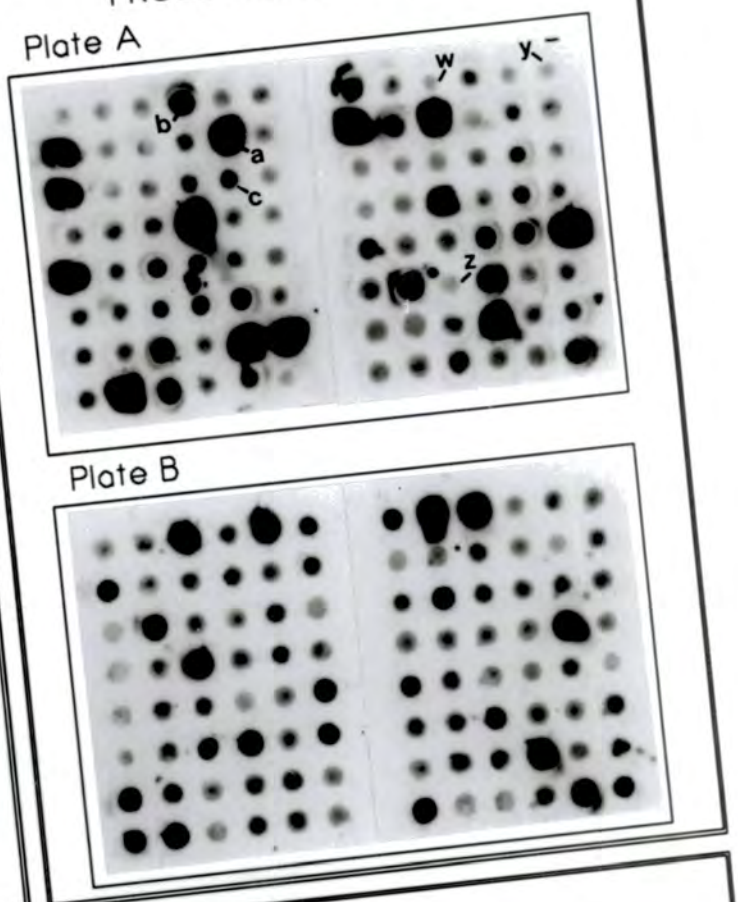
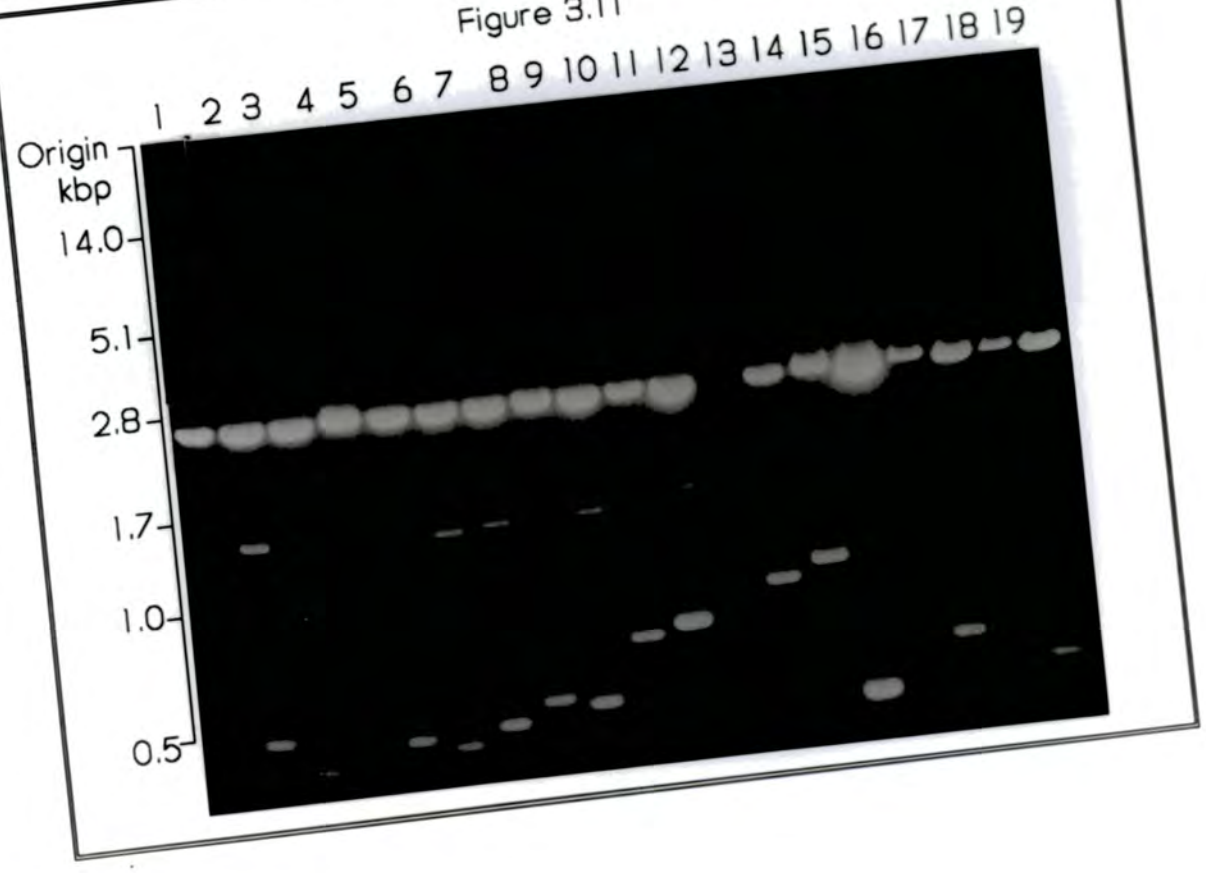


Figure 3.11



the relative abundancy of each clone. Although every effort was made to ensure that scoring was carried out from autoradiographs which had identical background clone hybridization, by varying the time of exposure of the autoradiographs, some still showed varying degrees of background hybridization.

Total cDNA and therefore the cDNA library should generally represent all the mRNA species that are present in the specific tissue from which the total RNA was isolated. Several probes were prepared from different sources and used to probe the cDNA library in an attempt to analyse its quality and specificity. Sets of filters containing the cDNA clones were separately probed with two rape seed-specific clones namely pC1 and pN2 which encode cruciferin and napin, the two major rape seed storage proteins, and also pHA1 which codes for ribosomal DNA. pHA1 contains the whole sequence of one ribosomal gene repetitive unit from pea which cross-hybridises with most, if not all, higher plant rRNA genes. The cDNA library was also probed with total cDNA from leaf tissue of a high glucosinolate DNA variety Jet Neuf and the results are discussed later in Section 3.3.6 together with the results from screening with total cDNA from pod tissue (excluding the seeds) of a high glucosinolate variety Jet Neuf.

When pC1 and pN2 were used as probes for the cDNA library, the percentage of cDNA clones that hybridised to these characterised clones was 6% and 4% respectively. According to DeLisle and Crouch (1989), the levels of cruciferin and napin mRNA during cell expansion were 11% and 8% respectively of the total mRNAs in seeds. The cDNA library analysed was constructed from pod tissues that include the developing seeds where the proportion of the developing seed material in the total pod material was about 40 to 50% by weight. Consequently, the figures of 6% and 4% of the cDNA clones hybridising to pC1 and pN2 respectively are therefore closely correlated with the levels of mRNA specific for cruciferin and napin in rape seed. From this result, it can therefore be suggested that the cDNA library has provided a true representation of the mRNA population in the seed and by inference, also in pod tissue.

The cDNA library was then probed with pHA1 in an attempt to analyse for any contamination of ribosomal RNA in the poly-A^+ RNA preparation during the construction of the library. The result showed that about 5% of the cDNA clones

hybridised with pHA1 indicating the presence of some ribosomal clones; therefore, confirming that the poly-A^+ RNA isolated was, to a small extend, contaminated with ribosomal RNA (Data not shown). According to Scutt (1990), the acceptable level of contaminating ribosomal clones should be less than 1% of the whole cDNA library.

A possible cause for the contamination was that the oligo-dT-cellulose batch method employed in the poly-A^+ isolation step was not completely efficient in separating poly-A^+ RNA from ribosomal RNA contamination. At best, poly-A^+ is always contaminated with rRNA. Thus the repurification step in which the poly-A^+ RNA fraction isolated was reloaded onto and eluted from a freshly regenerated batch of oligo-d(T) cellulose proved inadequate to completely remove the ribosomal RNA. It can therefore be suggested that the batch method is unsuitable for selecting poly-A^+ RNA to be used in the construction of a cDNA library due to the high level of ribosomal RNA contamination remaining.

3.3.5 Differential Screening of the cDNA Library

To identify cDNA clones which might be associated with high glucosinolate content in rape, differential screening of the cDNA library was carried out. cDNA synthesized from total poly-A^+ RNA isolated from pod material from a low glucosinolate variety at identical stage of development to that used for the cDNA library was used as a probe for a duplicate set of nitrocellullose filters containing the whole cDNA library (1536 clones). Figure 3.13 shows two examples of the autoradiographs from the differential screening. The autoradiographs were analyzed in an identical manner as before and the relative abundance of each sequence was scored. The relative abundance of each cDNA clone was scored during the differential screening experiments, and compared, and those clones showing detectable differences in relative abundance and, hence, containing differentially expressed sequences, were identified. As a result of the differential screening the cDNA clones were classified into two categories: (i) those that showed identical levels of abundance in both with high and low glucosinolate probes and, (ii) those that showed a high abundance when probed with high glucosinolate cDNA than with the low glucosinolate cDNA and *vice versa*. Of the 1536 cDNA clones screened, 29 cDNA clones showed different levels of expression between high and low glucosinolate

lines. These cDNA clones only showed differing levels of expression; none showed expression with one probe and no expression with the other probe. These cDNA clones were selected and used for secondary screening.

The different levels of hybridisation described in this screen could also be the result of differing amount of plasmids containing the cDNA inserts arising from unequal bacterial growth in each colony. Therefore, to obtain a more accurate account of the relative abundance of the selected cDNA clones, differential screening of the selected cDNA clones with equal amounts of plasmids was carried out. Thus, an experiment was set up for a secondary differential screening with the high and low probes using equal amounts of plasmids isolated from each of the 29 cDNA clones selected. Plasmids were isolated from each clone and their concentrations were determined by Fluorimetric Method (Diaminobenzoic Acid Assay) as described in Materials and Methods. 2 μg of plasmids from each cDNA clones were dot-blotted onto a nitrocellullose filter according to the procedure described in Materials and Methods. 2 sets of nitrocellullose filters containing the dot-blotted plasmids of the cDNA clones were made and each set was separately probed with high and low glucosinolate cDNAs. One clone which showed an equal level of hybridisation to both high and low probes during the first differential screening and plasmid pUC18 was also included on the dot-blot filter as positive and negative controls respectively. The result of this secondary differential screening is shown in Figure 3.14. 5 cDNA clones appeared to be differentially expressed and therefore were selected for subsequent analysis. Inserts from the 2 of 5 cDNA clones showing the most significant differential expression were isolated and each was used as $^{\alpha}_{\lambda}$probe for southern blots of rape genomic DNAs from high and low glucosinolate rape varieties. The result of the southern blots showed that none of the selected clones were able to reveal unique RFLP patterns which could be linked to the glucosinolate content of a variety. All the varieties showed identical band patterns of one or two bands of the same size when the two cDNA clones were used as probes in all enzyme systems tested namely: *Bam* HI, *Eco* RI, *Eco* RV, *Hind* III, *Kpn* I, *Pst* I, *Xba* I, and *Xho* I. Representative autoradiographs of the band pattern revealed by one of the two cDNA clones with three different enzymes are presented in Figure 3.15 (a, b, and c).

Figure 3.13 — Primary differential screening of cDNA library. Autoradiographs of 2 duplicate filters containing samples of cDNA clones separately probed with pod cDNA from high and low glucosinolate varieties. Clones which were potentially differentially expressed, are indicated by a, b, c, and d. The filters were washed at high stringency (0.1x SSC).

Figure 3.14 — Secondary differential screening preparation of cDNA clones selected from the primary screen. Autoradiographs of 2 duplicate filters containing isolated plasmid preparations from clones selected from the primary screening of the cDNA library. Equal amounts of plasmid DNA's were denatured, dot-blotted onto nylon filters and separately probed with pod cDNA from high and low glucosinolate varieties (identical cDNAs as was used in the primary screening). Potential differentially expressed cDNA clones are indicated by v, w, x, y, z. Positive controls consisting of a cDNA clone highly abundant in both high and low glucosinolate varieties (3 or 4 dots applied) and a negative control consisted of pUC18 (1 dot applied) were included (indicated by p and n respectively). The cDNA clone indicated by y corresponded to clone b in Figure 3.13.

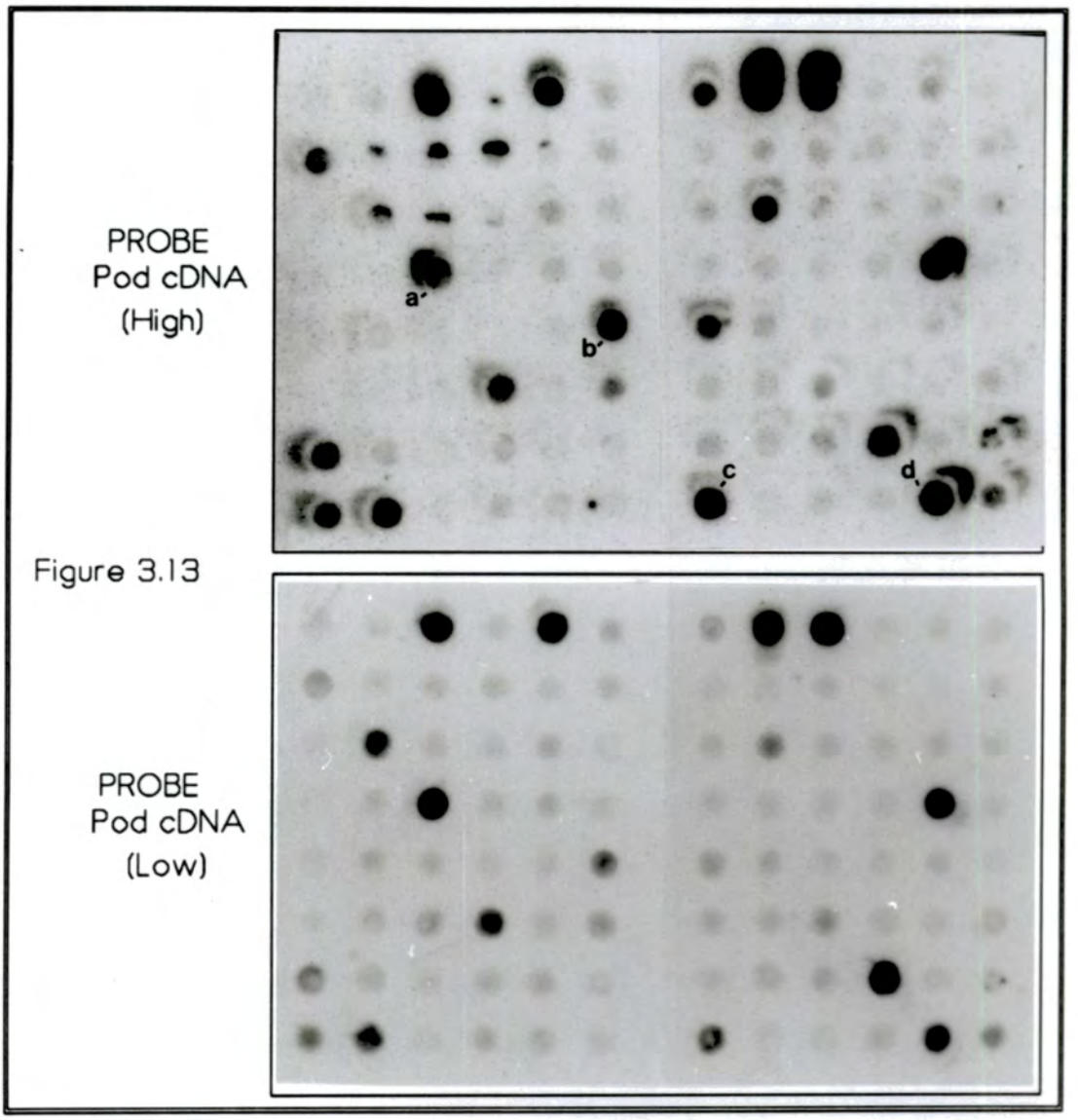


Figure 3.13

Figure 3.14

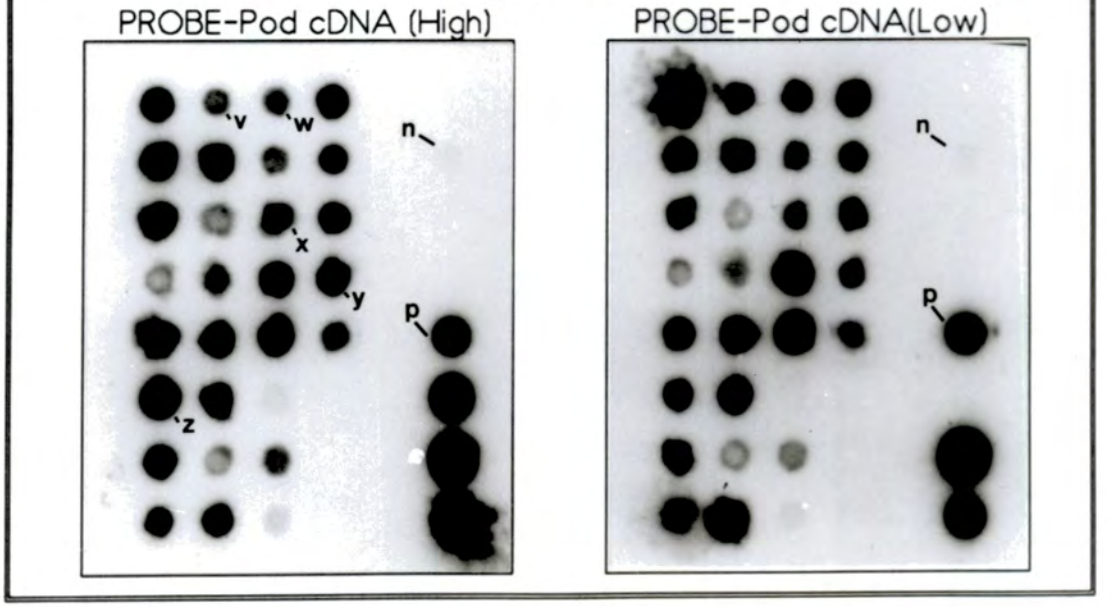
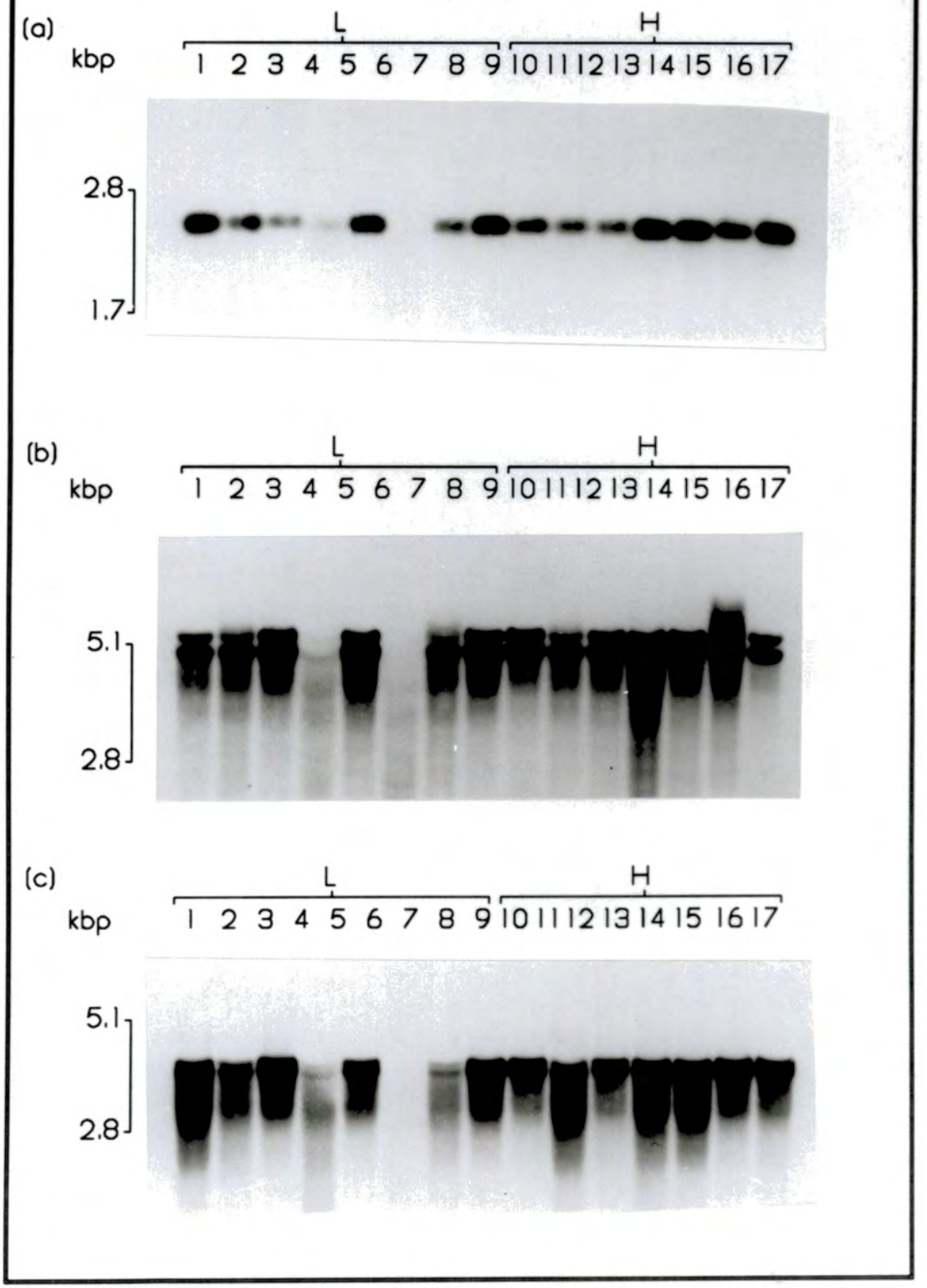


Figure 3.15 — Autoradiographs of rape DNA's digested with *Bam* HI, *Eco* RV, and *Xba* I and probed with a differentially expressed cDNA clone. Rape genomic DNA's from low (Ariana (1), Cobra(2), Drakkar (3), Libravo (4), Lictor (5), Lirawell (6), Topas (7), and Bronowski (8)) and high (Astec (9), Bienvenu (10), Brutor (11), Mikado (12), Pasha (13), Rafal (14) and Willi (15)) glucosinolate varieties were digested with (a) *Bam* HI, (b) *Eco* RV, and, (c) *Xba* I and probed. The filters were washed at high stringency (0.1x SSC) and exposed for 1 hour. Genomic DNAs of varieties Libravo (4), and Lirawell (6) apparently have degraded during the digestion.

Figure 3.15


(a)
kbp
L
H
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
2.8
1.7
(b)
kbp
L
H
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
5.1
2.8
(c)
kbp
L
H
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
5.1
2.8

3.3.6 Screening for Pod-Specific cDNA clones

The cDNA library was analysed by probing with cDNAs synthesised from leaf mRNA and pod mRNA isolated from pod tissues without the developing seeds which henceforth is called pod-only cDNA. Both leaf and pod tissues were from a high glucosinolate variety Jet Neuf. The results revealed that about 76% of the cDNA clones hybridised with leaf cDNA while about 72% hybridised with pod only cDNA. 98% of the cDNA clones that hybridised to pod-only total cDNA were also hybridised to leaf total cDNA. As was observed in the differential screening, the remaining 2% of the cDNA clones only showed differing levels of expression; none showed expression with one probe (pod-only total cDNA) and no expression at all with the other probe (leaf total cDNA). The high percentage (76%) of cDNA clones in the library hybridising to leaf cDNA was undoubtedly due to the photosynthetic nature of the pod and seed tissues and clones hybridising to both were most likely to be identical to those expressed in leaf tissue. Besides their photosynthetic nature, both pod and leaf tissues are known to be actively expressing glucosinolates at the same level (D. Murphy, pers. comm.). Therefore, mRNAs responsible for the expression of the glucosinolates would be the same and further contribute to the high percentage of hybridisation.

Among the cDNA clones which hybridised to the pod only cDNA probe, 15 clones showed a slightly greater hybridisation when compared with the hybridisation to leaf cDNA (Data not shown). Plasmids were isolated from these clones and analysed on agarose gel. They were also evaluated for cross-reactivity where each clone was separately used as probe for dot-blots containing plasmid of all 15 cDNA clones. One of the clones was found to be cross reacting. 8 out of the 14 non-crossreacting clones selected which contained inserts greater than 400 bp were then used separately as probes for sets of northern blots containing total RNA samples isolated from pod tissues of low and high glucosinolate varieties, and seed and leaf tissues from high glucosinolate variety. All 8 cDNA clones hybridised to components in all the RNA samples showing that these cDNA clones were most probably related to photosynthesis although hybridisation to the seed total RNA (Lane 3, Figure 3.16) was comparatively lower. This was probably due to the lower photosynthetic activity of developing seeds as judged by their lighter green colour compared with pod and leaf tissues. The hybridisation was also comparatively

lower in high glucosinolate pod total RNA (Lane 1, Figure 3.16). A representative autoradiograph of the Northern blots is shown in Figure 3.16.

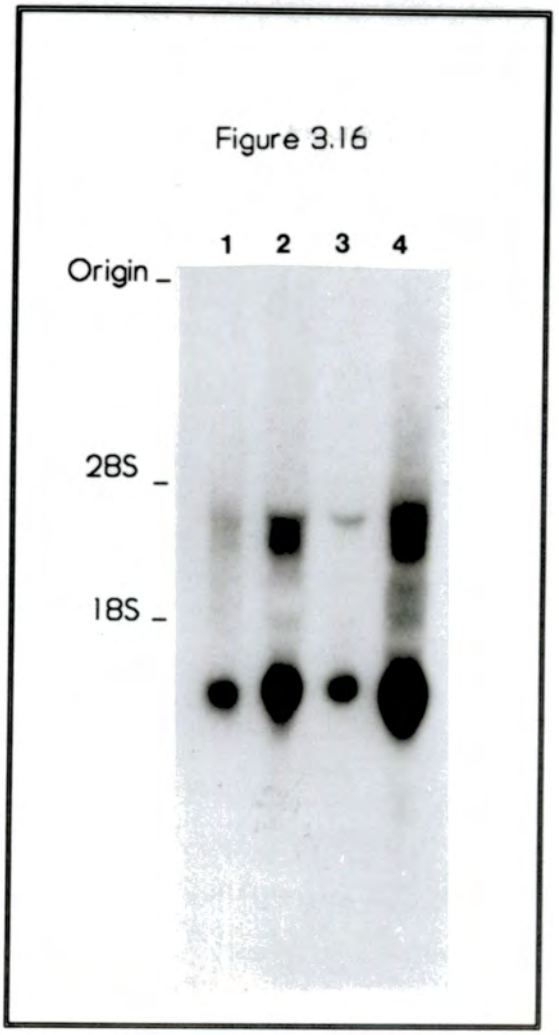
3.3.7 Concluding Remarks

From the analyses conducted on the constructed cDNA library, it was found that the library, although contaminated to a small extent with ribosomal clones, was relatively good and should therefore adequately represent all the mRNAs present in the pod tissue. One of the reason for the apparent failure of the differential screening to detect cDNA clones associated with high glucosinolate content was most probably due to the small number of cDNA clones screened. One possible solution would be to perform differential screening on more than 1536 cDNA clones as previously screened in a cDNA library. It is therefore suggested that, possibly, at least 20,000 cDNA clones should be screened.

The complexity of the pathway of glucosinolate biosynthesis will remain the main obstacle to progress towards a better understanding of their production in *Brasica sp.*. A thorough knowledge of the biosynthetic pathway will open up the possibility of using molecular biological techniques in the complete elimination of glucosinolates in the seeds of *B. napus* while maintaining high level of the compounds in other tissues. This is being worked on at Newcastle University (Agriculture Department) and John Innes Institute.

Figure 3.16 — Autoradiograph of Northern blots of total RNAs from different tissues of rape. Total RNA samples from pod tissues from a high and low glucosinolate varieties (Lanes 1 and 2 respectively) and seed and leaf tissues (Lanes 3 and 4 respectively) from a high glucosinolate variety were analysed on a Northern blot and were probed with one of the apparently pod specific" cDNA. The filter was washed with medium stringency (2x SSC) and exposed overnight.

Figure 3.16



3.4 Hind III REPETITIVE SEQUENCE

Digestion of genomic DNAs from a few varieties of oilseed rape with *Hind* III and followed by electrophoresis in 0.8% (w/v) agarose gel in the presence of EtBr, revealed heavily stained bands of DNA forming a unique "step-ladder" pattern. This was consistently observed in all of the *Hind* III digestions. Figure 3.17 shows a typical stained gel of rape DNAs from different varieties digested with *Hind* III as described. Initial studies have indicated that there may be varietal differences in the banding intensity and size as depicted in Figure 3.17. The banding patterns of rape DNAs could be arbitrarily divided into two groups where varieties in Lanes 1, 3, and 6 (Figure 3.17) belonged to Group 1 and varieties in Lanes 2, 4, and 5 belonged to Group 2.

The heavily stained bands of DNA, the smallest size (about 180 bp) being the most intense, showed all the characteristics of highly abundant repetitive sequences in a genomic DNA digests. Consequently, it was suspected that this might be due to the presence of a family of highly repetitive sequences in the genomic DNA in which the lowermost band constituted the monomer and the upper bands constituted tandem repeats of the monomer to make dimers, trimers, tetramers, etc.. If the larger bands constituted the polymeric forms of the monomer, then they would have sequences identical to the monomer and therefore would hybridise at high stringency when the monomeric sequence was used as a probe for a Southern blot of *Hind* III-digested DNAs. An attempt was therefore made to clone the lowermost band which was suspected to contain the monomer of the repeat sequence and to use this cloned sequence as a probe to investigate the speculations mentioned above.

3.4.1 Cloning of the Monomer in the Repetitive Sequence

Rape genomic DNA from the variety Mikado was digested to completion with *Hind* III and electrophoresed on a 1% (w/v) agarose gel for 12 hours for maximum resolution of the low molecular weight DNA fragments. A low molecular weight DNA marker (pBR322/*Alu* I) was also electrophoresed to estimate the size of the monomer. The smallest band of DNA with an approximate size of 180 bp was excised from the gel, purified and the DNA recovered for subsequent cloning into *Hind* III-digested pUC18. White transformants were selected and the clones were

Figure 3.17 — Analysis of rape genomic DNAs digested with *Hind* III. After digestion with *Hind III*, the DNA samples from several varieties were loaded onto a 0.8% (w/v) agarose gel, electrophoresed in the presence of 5 μg/ml ethidium bromide and photographed. The lowest band was approximately 180 bp (monomer), followed by 360 bp (dimer), 540 bp (trimer), 720 bp (tetramer) and 900 bp (pentamer), etc., with increments of 180 bp. These bands are arrowed in the figure.

Figure 3.18 — Autoradiograph of Southern blot of rape genomic DNAs probed with pRS1. DNA samples from variety Mikado were digested with *Hind* III, *Bam* HI, *Eco* RI, *Xba* I, *Xho* I, *Bgl* II, *Pst* I and *Sal* I (Lanes 3 to 10 respectively). Clone pRS1 digested with *Hind* III was also included in the blot. The 180 bp insert from pRS1 (Lane 1) corresponds exactly to the monomer in the *Hind* III-digested DNA (Lane 2) and hybridised strongly to the polymeric forms indicating sequence homology to pRS1 (Lane 3). Exposure time was 24 hours.

Figure 3.17

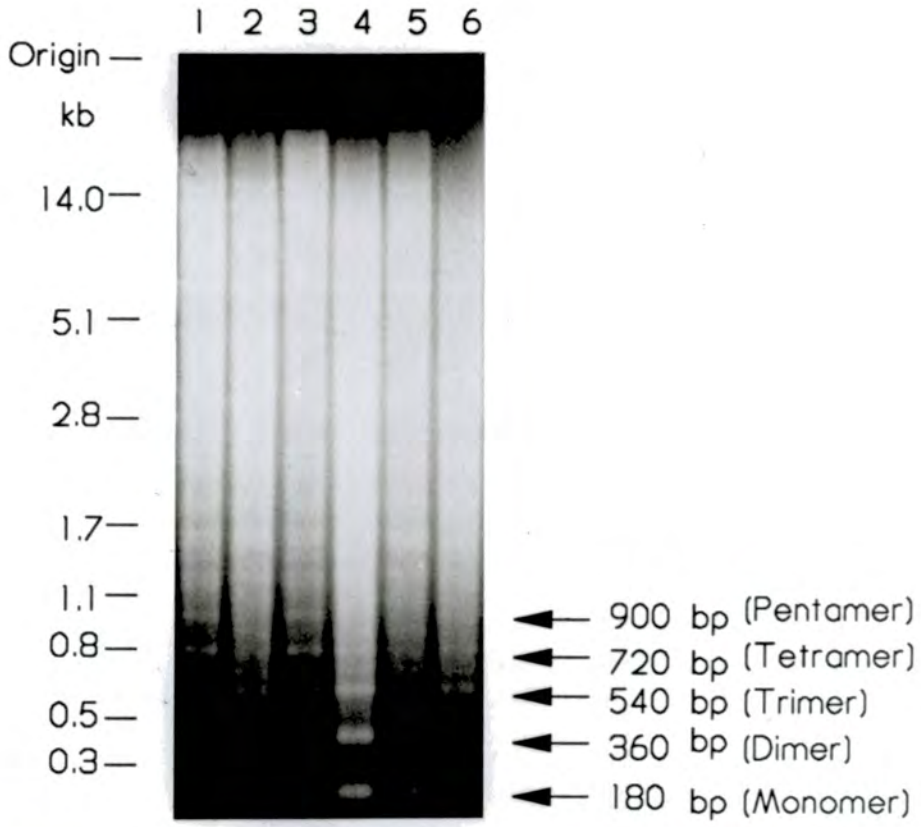
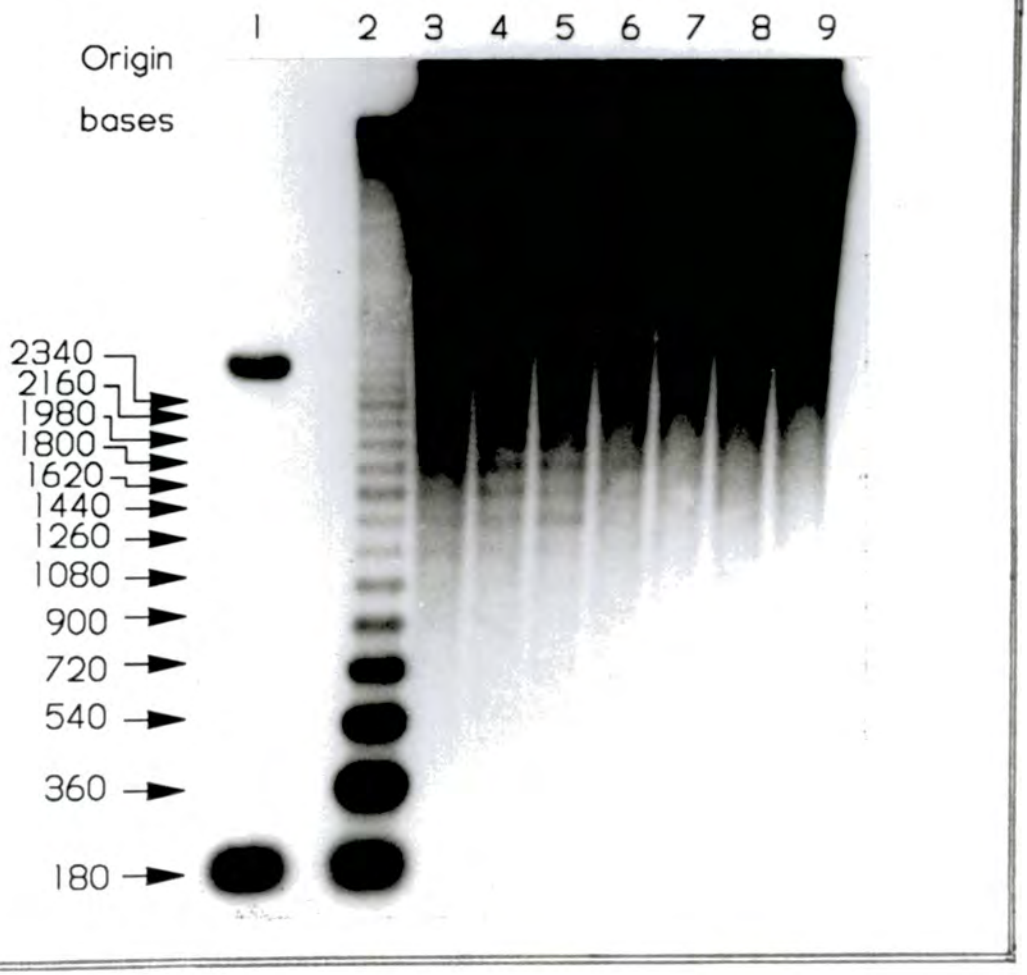


Figure 3.18



digested with *Hind* III and electrophoresed. Several clones with approximately 180 bp inserts were isolated from the transformation. One of these clones, which was later designated pRS1 (repetitive sequence) was chosen at random, digested with *Hind* III, and the insert was used as a probe for hybridising to a Southern blot containing *Hind* III-digested rape genomic DNAs.

The result of the Southern hybridisation using the insert from pRS1 showed that it hybridised strongly to the smallest band and also to each of the bands in the "step-ladder" pattern of the *Hind* III-digested genomic DNA at high stringency (0.1x SSC) (Figure 3.18). The intensity of the bands in the step-ladder pattern obtained by hybridisation and autoradiography appeared to correlate well, to the band intensities observed by EtBr staining of the gel. This further confirmed that the cloned sequence was indeed from the 180 bp band and that the larger bands which formed the "step-ladder" pattern also contained sequences homologous to the cloned sequence. The fact that the hybridisation of the monomeric sequences to other bands in the "step-ladder" was stable at high stringency (0.1x SSC) provided the evidence that the sequence homology was at least 98%. This, however, could only be confirmed by sequence comparison between the monomer and the higher polymeric forms of the repetitive sequences. This aspect is dealt with in later sections of this chapter.

Other enzymes, namely *Bam* HI, *Eco* RI, *Xba* I, *Xho* I, *Bgl* I, *Pst* I, and *Sal* I (Figure 3.18, lanes 4 to 10 respectively) were also included in the Southern blot probed with pRS1. These commonly used enzymes was included in the analysis to investigate the presence of restriction sites for these enzymes within the repetitive sequences and its flanking regions. As depicted in Figure 3.18 (lanes 4 to 10), all the restriction enzymes chosen in the analysis failed to give the characteristic "step-ladder" pattern of *Hind* III digestion. The probe, on the other hand, strongly hybridised mainly to the high molecular weight region of the digested DNA samples suggesting that the tandemly arranged sequences was not cleaved by the selected enzymes and remained integrated with the high molecular weight DNA. This result also suggested that the restriction sites of these enzymes were not present in the monomeric sequence. The presence of a restriction enzyme site in the monomeric sequence would probably give a "step-ladder" pattern comparable to that of the *Hind* III digest or in the case of a conserved site, a low molecular weight band

coincident with the 180 bp monomer. There was however, a slight indication of a faint "step-ladder" pattern when the restriction enzymes *Eco* RI and *Xba* I were used in the analysis (lanes 5 and 6 respectively). A possible explanation for this phenomenon was the presence of *Eco* RI or *Xba* I restriction sites for these enzymes on a very limited number of repeats (i.e. an *Eco* RI or *Xba* I subclass of the RS units). Again, such speculation can only be confirmed by sequence analysis of the repetitive sequence.

It can therefore be concluded that pRS1 contained a 180 bp insert corresponding to the monomeric form of the repetitive sequences. The homology of pRSI with the higher MW forms was confirmed by its hybridisation at high stringency to the step-ladder pattern observed in *Hind* III digestion.

3.4.2 Characterisation of the Clone pRS1

Clone pRS1 was further characterised by subjecting it to plasmid DNA sequencing and its sequence together with the restriction enzyme map is presented in Figure 3.19. When the nucleotides originally from the pUC18 cloning vectors were excluded, the actual size of the insert (monomer) was 176 bp. The monomeric sequence begins with ↓AGCTT (i.e. the second nucleotide in Figure 3.19) and ends with TAAGA↓ (i.e. the 177th nucleotide) (Arrows indicate the cleavage point of the enzyme *Hind* III).

In an attempt to investigate the possibility that the sequence codes a protein, the sequence was translated and analysed for open reading frame as depicted in Figure 3.20. The complete 6-phase translation analysis is presented in Figure 3.20(A) and the schematic representation of this analysis is summarised in Figure 3.20(B). In the forward sequence translation, there is a very short open reading frame of 33 nucleotide long with start codon and a termination codon (first phase (1), Figure 3.20), and also an apparent reading frame spanning almost the entire length of the monomer but do not contain any start/stop codons (third phase (3), Figure 3.20). In the reverse sequence translation, there is an open reading frame of 69 nucleotide long with two start codons but no termination codon within the sequence. The complete (with start and stop codons) open reading frame indicated that it is possible that a protein of 11 amino acids long could be produced but further analysis is necessary to confirm this. Alternatively, the expression of the

repetitive sequence could be studied by probing Northern blots containing total RNAs from various rape tissues with insert from pRS1.

In the preceeding section, it was observed that there was apparently a faint "reversed step-ladder" pattern when *Eco* RI and *Xba* I were used. "Reversed step-ladder" was so named because contrary to that observed in *Hind* III-digested DNA sample, the pattern revealed by these enzymes has band intensities which gradually decreased with decreased in the fragment size. Analysis of the cloned sequence revealed that there were no restriction sites for *Eco* RI and *Xba* I present in this monomeric sequence.

3.4.3 Analysis of Repetitive Sequences in Oilseed Rape Varieties

The observed qualitative and quantitative differences between varieties was further investigated by analysis of DNA samples, to investigate whether there were any genuine varietal differences in the "step-ladder" pattern. Genomic DNAs from several varieties of oilseed rape were digested with *Hind* III under standard condition of 3U/μg DNA enzyme for 3 hours. analysed as before, Southern blotted onto nitrocellulose filter and then probed with pRS1. The result are presented in Figure 3.21.

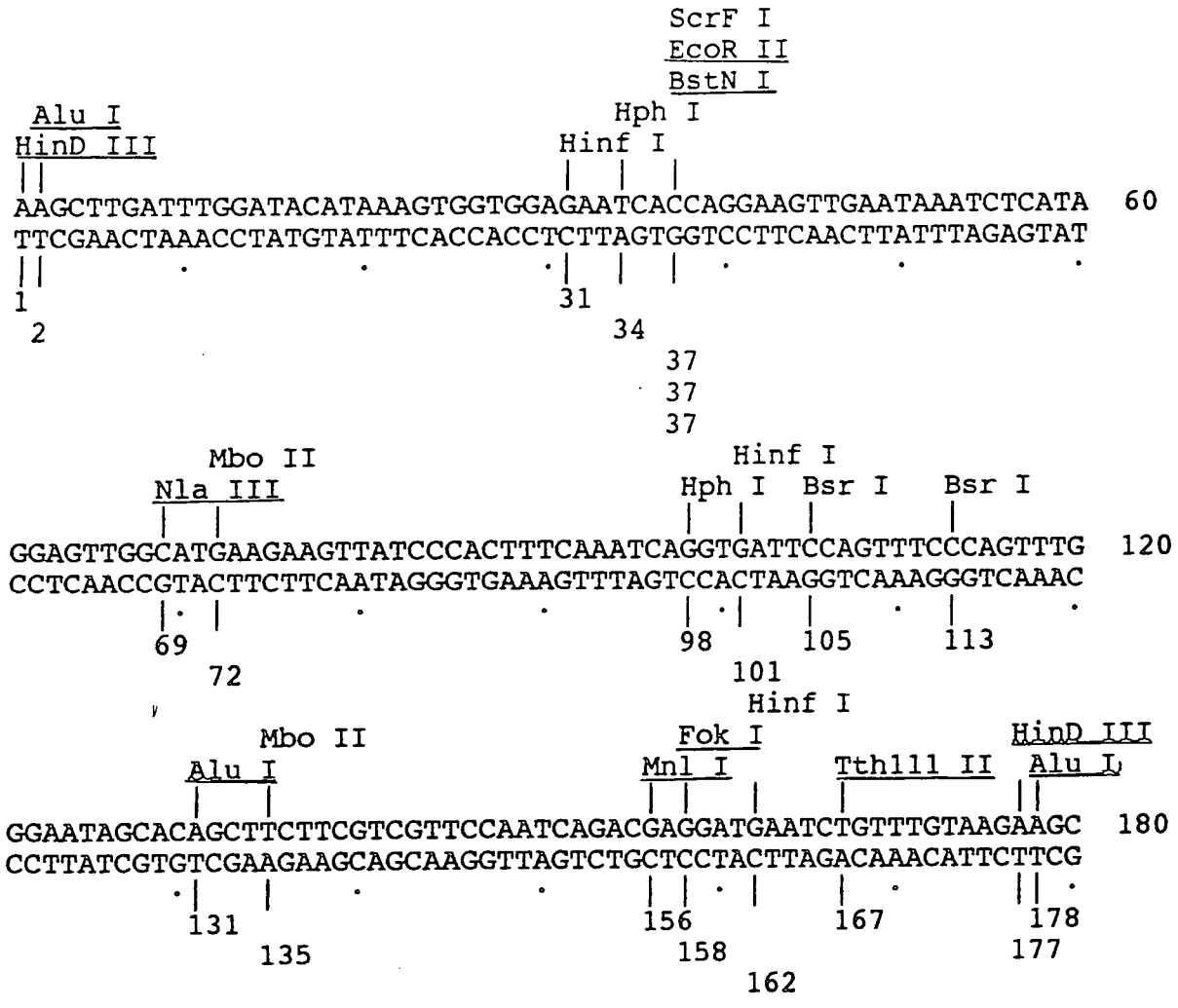
Although the sizes of the DNA bands showing up in the "step-ladder" pattern were apparently identical in all of the varieties tested, there was a marked difference in the intensities of the bands. A closer examination of the autoradiograph (Figure 3.21) revealed that the pattern could be classified into three intensity patterns as exemplified by lanes 2, 3, 4, and 8 (Pattern 1), lanes 1, 5, 6, 7, 9, 10 and 11 (Pattern 2), and lanes 12 and 13 (Pattern 3). In Pattern 1, most of the repetitive sequences were still integrated in the high molecular weight DNA with an apparent "step ladder" in which there was a progressive increase in intensity of the higher polymeric forms. In Pattern 2, almost all of the repetitive sequences has dissociated from the higher molecular weight DNA and formed a "step-ladder" with the middle size polymeric forms showing high intensity and a progressive decrease in intensities of the higher and lower polymeric forms of repetitive sequences. Finally, in Pattern 3, only the lowest molecular weight bands showed up as the strongest

Figure 3.19 — Sequence data and the restriction map of the monomer in pRS1. (A) Clone pRS1 was sequenced using a plasmid DNA sequencing procedure and analysed for restriction sites using a programme called DNA strider (Marck, 1988) in Apple Macintosh SE/30. The actual monomeric sequence begins with ↓AGCTT and ends with TAAGA↓. The full *Hind* III recognition sequence was included so as to facilitate the analysis. (B) Schematic representation of the restriction map with commonly used enzymes.

Figure 3.19

(A)

Positions of Restriction Endonucleases sites (unique sites underlined)



(B)

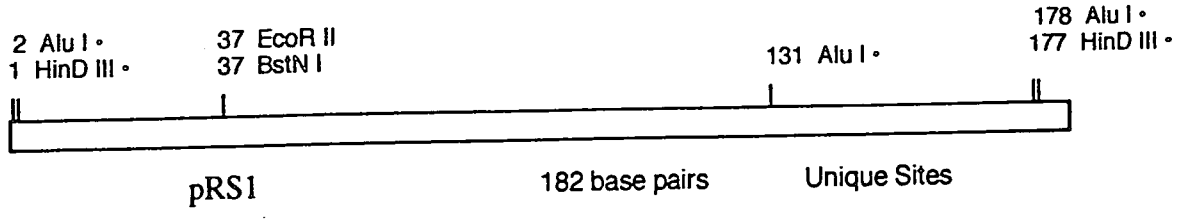


Figure 3.20 — Translation of the pRS1 Monomeric Sequence. The monomeric sequence translated in all six reading frames to determine any potential coding regions and the possible amino acids sequence(s) was determined (A). OPA, AMB, and OCH denote stop codons TGA, TAG, and TAA respectively. Schematic representation of the open reading frame map is in (B). The arrows indicate open reading frames with short lines and tall vertical lines denoting the start and stop codons, respectively. Positive and negative numbers represent the phases of forward and reverse sequence translations. The long bidirectional arrow indicates the open reading frame in the forward sequence translation which does not contain the start/stop codons within the sequence.

(A)

Figure 3.20

DNA Strider™ 1.1 ### Wednesday, September 4, 1991 3:10:44 pm

prs1 -> 6-phase Translation

DNA sequence 182 b.p. AAGCTTGATTTG ... TGTAAGAAGCTT linear

```
1/1                                                    31/11
  ala OPA phe gly tyr ile lys trp trp arg ile thr arg lys leu asn lys ser his arg
 ser leu ile trp ile his lys val val glu asn his gln glu val glu OCH ile ser AMB
lys leu asp leu asp thr OCH ser gly gly glu ser pro gly ser OPA ile asn leu ile
AAG CTT GAT TTG GAT ACA TAA AGT GGT GGA GAA TCA CCA GGA AGT TGA ATA AAT CTC ATA
TTC GAA CTA AAC CTA TGT ATT TCA CCA CCT CTT AGT GGT CCT TCA ACT TAT TTA GAG TAT
leu lys ile gln ile cys leu thr thr ser phe OPA trp ser thr ser tyr ile glu tyr
 ala gln asn pro tyr met phe his his leu ile val leu phe asn phe leu asp OPA leu
  ser ser lys ser val tyr leu pro pro ser asp gly pro leu gln ile phe arg met pro
61/21                                                  91/31
  ser trp his glu glu val ile pro leu ser asn gln val ile pro val ser gln phe gly
 glu leu ala OPA arg ser tyr pro thr phe lys ser gly asp ser ser phe pro val trp
gly val gly met lys lys leu ser his phe gln ile arg OPA phe gln phe pro ser leu
GGA GTT GGC ATG AAG AAG TTA TCC CAC TTT CAA ATC AGG TGA TTC CAG TTT CCC AGT TTG
CCT CAA CCG TAC TTC TTC AAT AGG GTG AAA GTT TAG TCC ACT AAG GTC AAA GGG TCA AAC
ser asn ala his leu leu OCH gly val lys leu asp pro ser glu leu lys gly thr gln
 leu gln cys ser ser thr ile gly ser glu phe OPA thr ile gly thr glu trp asn pro
  thr pro met phe phe asn asp trp lys OPA ile leu his asn trp asn gly leu lys pro
121/41                                                 151/51
  asn ser thr ala ser ser ser phe gln ser asp glu asp glu ser val cys lys lys
 glu AMB his ser phe phe val val pro ile arg arg gly OPA ile cys leu OCH glu
gly ile ala gln leu leu arg arg ser asn gln thr arg met asn leu phe val arg ser
GGA ATA GCA CAG CTT CTT CGT CGT TCC AAT CAG ACG AGG ATG AAT CTG TTT GTA AGA AGC
CCT TAT CGT GTC GAA GAA GCA GCA AGG TTA GTC TGC TCC TAC TTA GAC AAA CAT TCT TCG
ser tyr cys leu lys lys thr thr gly ile leu arg pro his ile gln lys tyr ser ala
 phe leu val ala glu glu asp asn trp asp ser ser ser ser asp thr gln leu phe
  ile ala cys ser arg arg arg glu leu OPA val leu ile phe arg asn thr leu leu
```

(B)

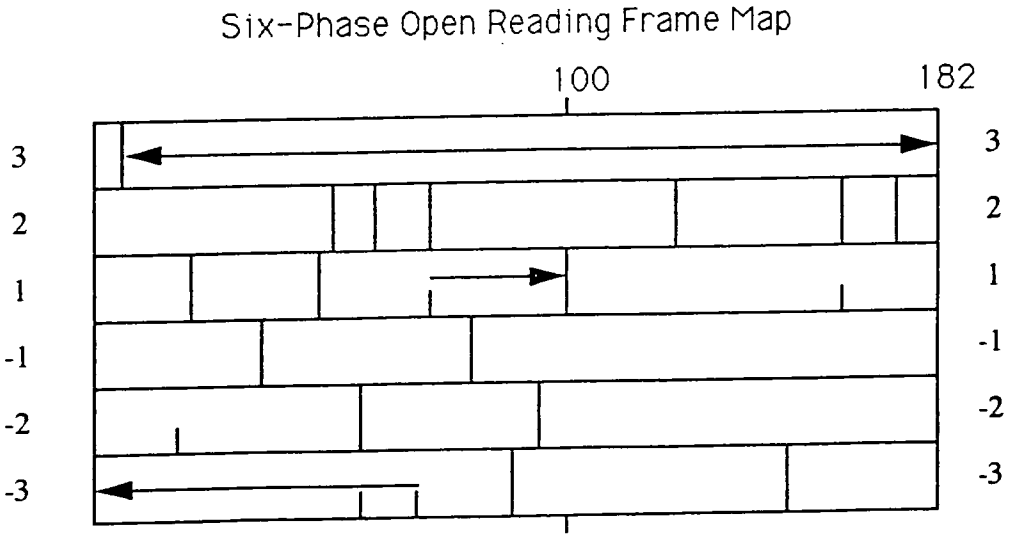
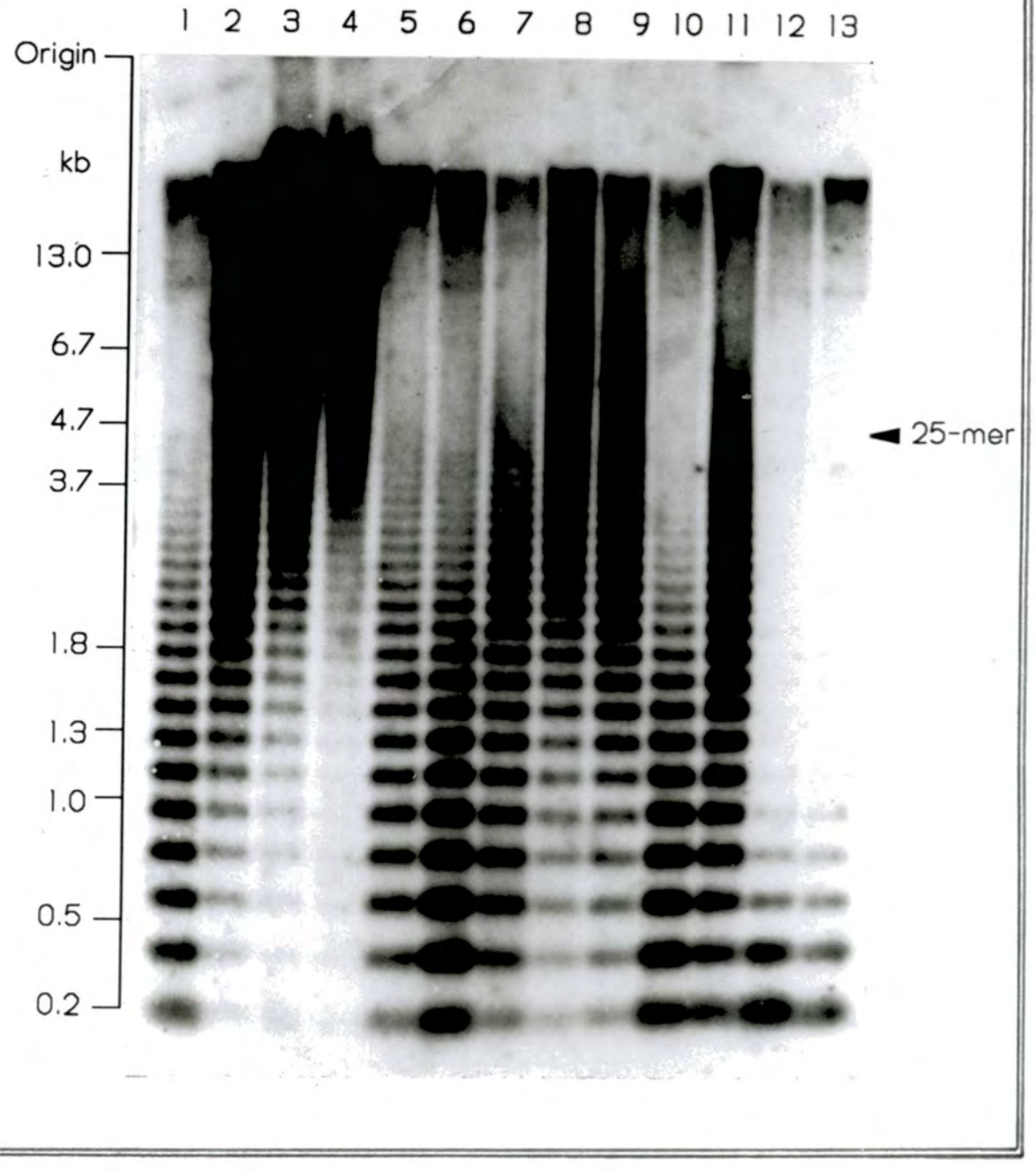


Figure 3.21 — Autoradiograph of *Hind* III-digested oilseed rape DNAs from 13 different varieties probed with pRS1. DNA samples from 13 varieties of oilseed rape namely, Ariana, Bienvenu, Brutor, Cobra, Drakkar, Jet Neuf, Libravo, Lirawell, Mikado, Optima, Pasha, Topas and Willi (Lanes 1 to 13 respectively). Quantitative variations in the "step-ladder" pattern are apparent and were classified into three intensity patterns as exemplified by lanes 2, 3, 4 and 8 (Pattern 1, lanes 1, 5, 6, 7, 9, 10 and 11 (Pattern 2) and lanes 12 and 13 (Pattern 3). The number of visually analysable bands that made the step-ladder was at least 25 with the size of about 4.5 kilobasepairs (kbp) (arrowed). Exposure time was 1 week.

Figure 3.21



hybridising bands.

The bands that made the "step-ladder" were counted to estimate the longest polymeric sequences present in rape genome. It was apparent that at least 25 distinct bands were present in the step-ladder indicating that the longest polymeric sequence is made up of at least 25 monomeric sequences (i.e. about 4.5 kbp) arranged in tandem array in rape genome even on prolonged exposure. This estimation was carried by counting the number of distinct bands that could be visually resolved. There was no doubt that this estimation was very conservative as hybridisation data revealed by those varieties belonging to Pattern 1 showed that pRS1 hybridised to DNA with a molecular weight higher than 4.5 kbp. However, this observed hybridisation might represent hybridisation of pRS1 to repetitive sequences that were still associated with other high molecular weight, non-related genomic DNA sequences and not hybridisation to genuine, intact (tandemly arranged) repetitive DNA.

Further investigations were carried out to see whether these apparent differences in pattern intensity were variety specific. The danger in assessing differences between highly repetitive DNA using the enzyme which reveals the pattern is that of partial digestion products as a result of differences in the purity of the DNA. To eliminate this possibility, two experiments were carried out to investigate the effects of enzyme concentration and prolonged incubation period on the "step-ladder" patterns; conditions always associated with partial digestion.

3.4.4 Effects of Enzyme Concentration and Incubation Period

Five μg of DNA samples from varieties Bienvenu, Jet Neuf, and Mikado were separately digested with different amounts of enzymes (1U, 3U, and 6U) per μg DNA of *Hind* III for 3 hours and electrophoresed in a 0.75% (w/v) agarose gel. These varieties were chosen as representative varieties showing "step-ladder" pattern belonging to Pattern 1 and 2 which were suspected to be resulted from partial digestion. The gel was then southern-blotted and the filter probed with pRS1. The autoradiograph is shown in Figure 3.22.

All DNA samples eventually showed the same final pattern of bands (Pattern 3) when digested with 6U/μg DNA of enzyme suggesting that this pattern (Pattern

Figure 3.22 — Effect of enzyme concentrations on the "step-ladder" pattern in *Hind* III-digested rape genomic DNAs. Three genomic DNAs from rape varieties Bienvenu (Lanes 1, 5, and 9), Jet Neuf (Lanes 2, 6, and 10), and Mikado (Lanes 3, 7, and 11) were separately digested with 1U/μg DNA (Lanes 1, 2, and 3), 3U/μg DNA (Lanes 5, 6, and 7), and 6U/μg DNA (Lanes 9, 10, and 11) of *Hind* III. A progression of changes of band pattern intensities ranging from that of Pattern 1 (Lane 2, 4 and 6) to that of Pattern 3 (Lanes 7, 8 and 9) was observed. Exposure time was 6 hours.

Figure 3.23 — Effect of prolonged incubation periods on the "step-ladder" band pattern of rape DNA digested with excess *Hind* III. 5 μg of rape genomic DNAs from two varieties, Jet Neuf (lane 1, 3, and 5) and Mikado (lanes 2, 4, and 6) were digested with 6U/μg DNA *Hind* III for 6 hours (Lanes 1, and 2), 12 hours (Lanes 3, and 4), and 18 hours (Lanes 5, and 6). For the 18 hours incubation time, a fresh aliquot of enzyme was added at 12 hours incubation. Exposure time was 6 hours.

Figure 3.22

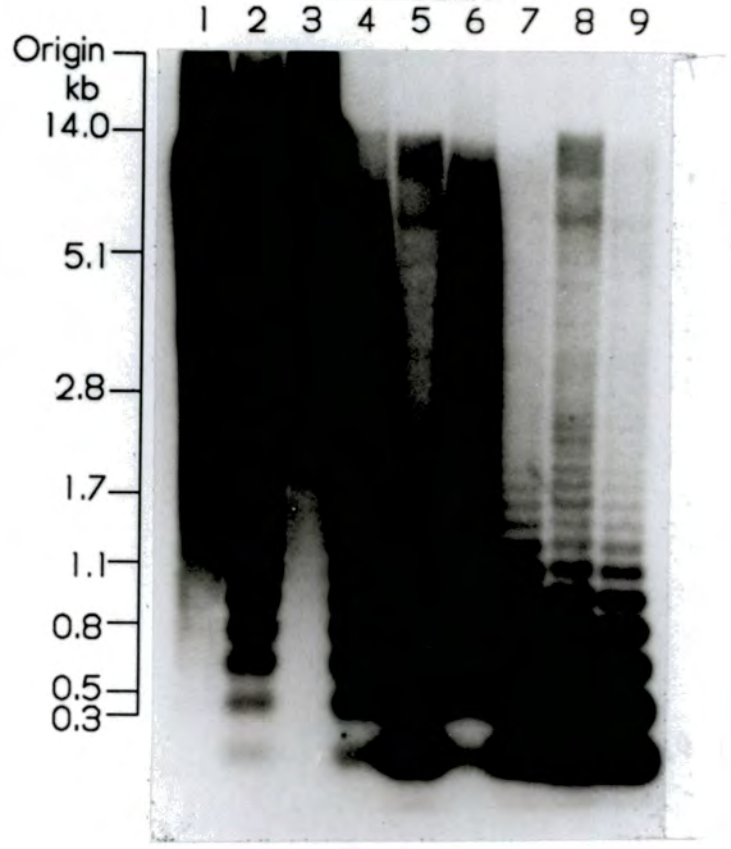
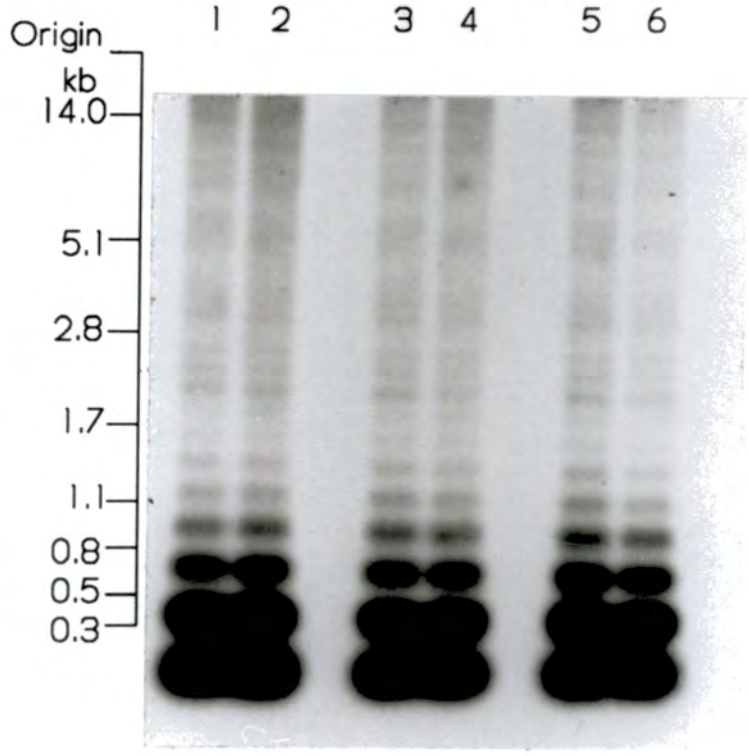


Figure 3.23



3) is stable and therefore represents the pattern of the complete digest of rape DNA with *Hind* III. The original differences in patterns of band intensity was therefore caused by incomplete digestion of the genomic DNA samples tested due to inadequate enzyme for the required digestion or genomic DNAs of different quality.

A further set of experiments was set up to investigate whether a prolonged incubation period had any effect on the distribution of band intensities. In this case, 5 μg of rape DNA from varieties Jet Neuf and Mikado were separately digested with *Hind* III at 6U/μg DNA for periods of 6, 12, and 18 hours. After 12 hours incubation, a second aliquot of 6U/μg DNA *Hind* III restriction enzyme was added. The restriction conditions were adjusted accordingly by adding appropriate volumes of 10x restriction buffer and water and the digestion allowed to continue for another 6 hours (for the 18-hour incubation time). The samples were then electrophoresed on a 0.75% (w/v) gel, southern-blotted, and probed with pRS1. The autoradiograph is shown in Figure 3.23. The band patterns from both DNA samples were identical in all 3 incubation times. A stable band pattern (Pattern 3) was obtained after 6 hours incubation time and remained so up to the 18 hours incubation time. This further suggested that the incubation time of 6 hours with 6U/μg DNA *Hind* III was more than enough for complete digestion. In fact, as already shown in previous experiment, an incubation period of 3 hours was sufficient for a complete digestion of all the susceptible *Hind* III sites in the DNA. Excessive *Hind* III enzyme in the reaction mixture had no further effect on the step-ladder pattern.

From these experiments involving enzyme concentration and prolonged incubation period it was concluded that the "step-ladder" pattern observed (Pattern 3) was stable even after digestion with excess enzyme and prolonged incubation period and was not due to partial digestion. Conditions established to achieve Pattern 3 were used subsequently in similar studies involving repetitive sequences.

3.4.5 Methylation Study of Rape DNA

Another parameter which may affect the endonuclease activity of restriction enzymes is the methylation pattern of the genomic DNA. Thus, having eliminated partial digestion as the underlying basis for the "step-ladder" pattern, one other

explanation could be the methylation state of *Brassica napus* genomic DNA and in particular, the regions of the repetitive sequences affecting the digestibility of the DNA.

Experiments were designed to investigate the methylation pattern of rape genomic DNA particularly in the region of the repetitive sequences. 5 μg of DNA from rape variety Mikado were separately digested with *Apy* I, *BstN* I, *Dpn* I, *EcoR* II, *Hpa* II, *Mbo* I, *Msp* I, *Sma* I, and *Xma I.* The recognition sequences of these restriction endonucleases with information on their sensitivity to different DNA methylation states are summarised in Figure 3.24. These enzymes were chosen based on the fact that their endonuclease activities are affected by being sensitive, dependent, or resistant to DNA methylation as suggested by Kessler*et al*, (1985). These enzymes were chosen in pairs of isoschizomers i.e. recognising the identical recognition sequence but where one enzyme is sensitive to DNA methylation, its isoschizomer is not. In order to ensure that the conditions for each enzyme were suitable to achieve complete digestion of all susceptible sites, 6U/μg of each enzyme were used and the digestion was carried out for 16 hours (overnight). This represented a large excess of restriction capacity. All the enzymes used in this experiment were newly purchased and were used immediately. Figure 3.25 shows the ethidium bromide-stained gel of DNAs digested with these enzymes.

The enzyme *Bst* NI, an enzyme not affected by the 5-methyl cytosine (5-MC) methylation in its recognition sites showed excellent restriction of rape DNA (Lane 2). *Eco* RII (Lane 3), which is sensitive to 5-MC methylation at the same position showed a slightly lower degree of restriction compared to *Bst* NI (Lane 2). The isoschizomer of *Bst* NI, *Apy* I, an enzyme which actually requires 5-MC methylation at the same position for its endonulease activity, showed good restriction comparable to *Eco* RII but it is lower than that of the methylation resistant *Bst* NI. In another pair of isoschizomers, *Msp* I and *Hpa* II, a similar phenomenon occured whereby *Hpa* II (Lane 5) which is sensitive to 5-MC methylation of 'external' and the 'internal' cytosine, showed a lower degree of restriction endonuclease activity compared with *Msp* I (lane 4) which is resistant to 5-MC methylation of the 'internal' cytosine. A similar observation was also found with another pair of isoschizomers, *Sma* I and *Xma* I. *Sma* I, which is sensitive to 5-MC of the 'internal' cytosine showed a lower degree of restriction when compared with *Xma*

Enzyme	Site	Base Seq. Within RS1 Sequence
Apy I	$C^{+}C^{m}$ ↓(A/T)GG	Yes
Bst NI	$C^{\circ}C^{\circ}$ ↓(A/T)GG	Yes
Eco RII	↓$C^{\circ}C^{+}$(A/T)GG	Yes
Hpa II	C^{+} ↓C^{+}GG	No
Msp I	C^{+} ↓C°GG	No
Sma I	CCC^{+} ↓GGG	No
Xma I	C↓$C^{+}C^{\circ}$GGG	No
Dpn I	GA^{m} ↓TC	No
Mbo I	↓$GA^{+}TC^{\circ}$	No
Hind III	$TTC^{+}GA$↓A^{+}	Yes

Figure 3.24 — Recognition sequence of restriction endonucleases affected by methylation pattern of DNA. The cleavage site of restriction endonucleases is represented by the symbol (↓). The symbols (A^{+}) or (C^{+}) indicate the inhibition of a restriction endonuclease by a N^{6}-methyl adenine or 5-methyl cytosine residues (5-MC) respectively within the recognition sequence. The symbols (A^{m}) or (C^{m}) indicate that N^{6}-methyl adenine or 5-methyl cytosine residue within the recognition sequence are a prerequisite for the enzymatic activity of the restriction endonuclease. The symbols (A°) or (C°) show that the digestion of the DNA with the restriction endonuclease is not influenced by a N^{6}-methyl adenine or 5-methyl cytosine residue within the recognition sequence. The information was obtained from Kessler *et al.* (1985). The presence or absence of the recognition sequences within the RS1 sequence is indicated in Column 3.

Figure 3.25 — Effects of methylation on digestion of rape DNA. DNA from rape variety Mikado was separately restricted with restriction enzymes *Apy* I, *Bst* NI, *Eco* RII, *Msp* I, *Hpa* II, *Sma* I, *Xma* I, *Dpn* I, *Mbo* I, and *Hind* III (Lanes 1 to 9 and 11 respectively) and electrophoresed on a 0.8% agarose gel in the presence of 5 μg/ml ethidium bromide. Unrestricted DNA was in Lane 10. A lower degree of restriction was observed for the C-methylation sensitive enzymes *Eco* RII, *Hpa* II and *Sma* I (Lanes 3,5 and 6 respectively) compared to their respective C-methylation resistant isoschizomers, *Bst* NI, *Msp* I and *Xma* I (Lanes 2, 4 and 7 respectively). The C-methylation dependent enzyme, *Apy* I (Lane 1), also showed a lower degree of restriction when compared with *Bst* NI. The A-methylation dependent enzyme, *Dpn* I (Lane 8) failed to restrict rape DNA while its isoschizomer, *Mbo* I (Lane 9) showed complete restriction. "Step-ladder" patterns were apparent in the digestions with *Apy* I, *Bst* NI, and *Eco* RII.

Figure 3.26 — Autoradiograph of Southern blot of rape DNAs restricted as in Figure 3.25, probed with pRS1. Restriction digests were as described in Figure 3.25 and the blot was probed with pRS1. Restriction enzymes *Apy* I (Lane 1), *Eco* RII (Lane 3), and to a lesser extent, *Bst* NI (Lane 2) showed a "step-ladder" pattern similar to *Hind* III while *Msp* I, *Xma* I and *Mbo* I showed patterns similar to that of partial *Hind* III digestions. The step-ladder patterns observed in the stained gel (above) was revealed by probe pRS1 showing that the bands contained homologous sequences. Exposure was 5 days.

Figure 3.25

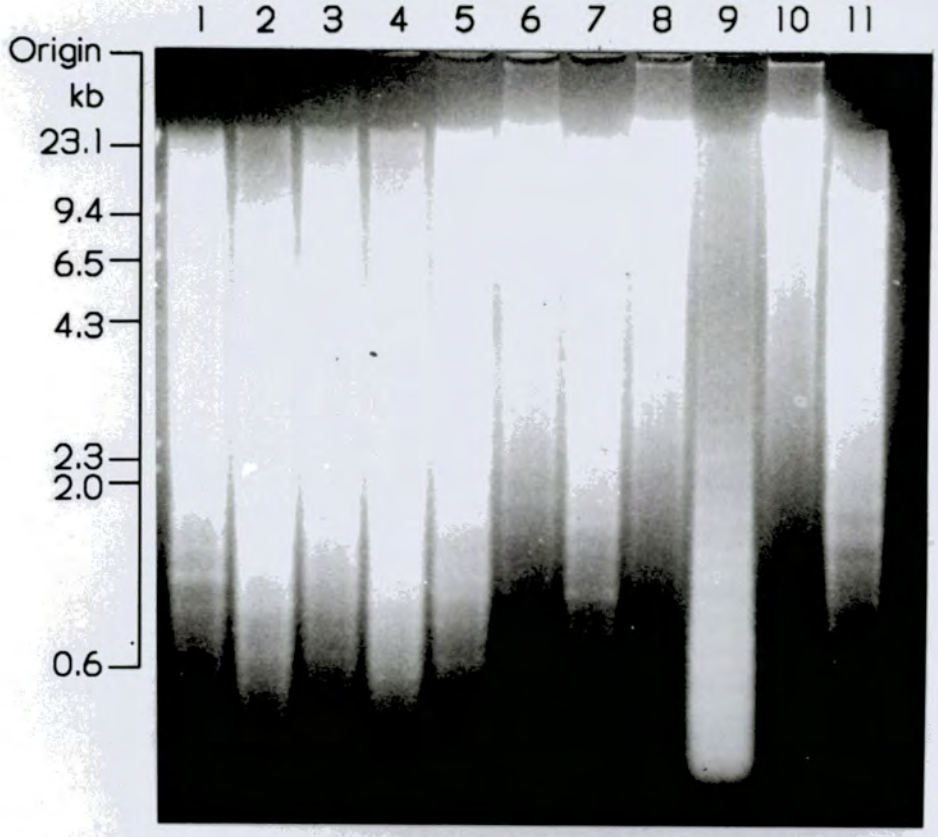
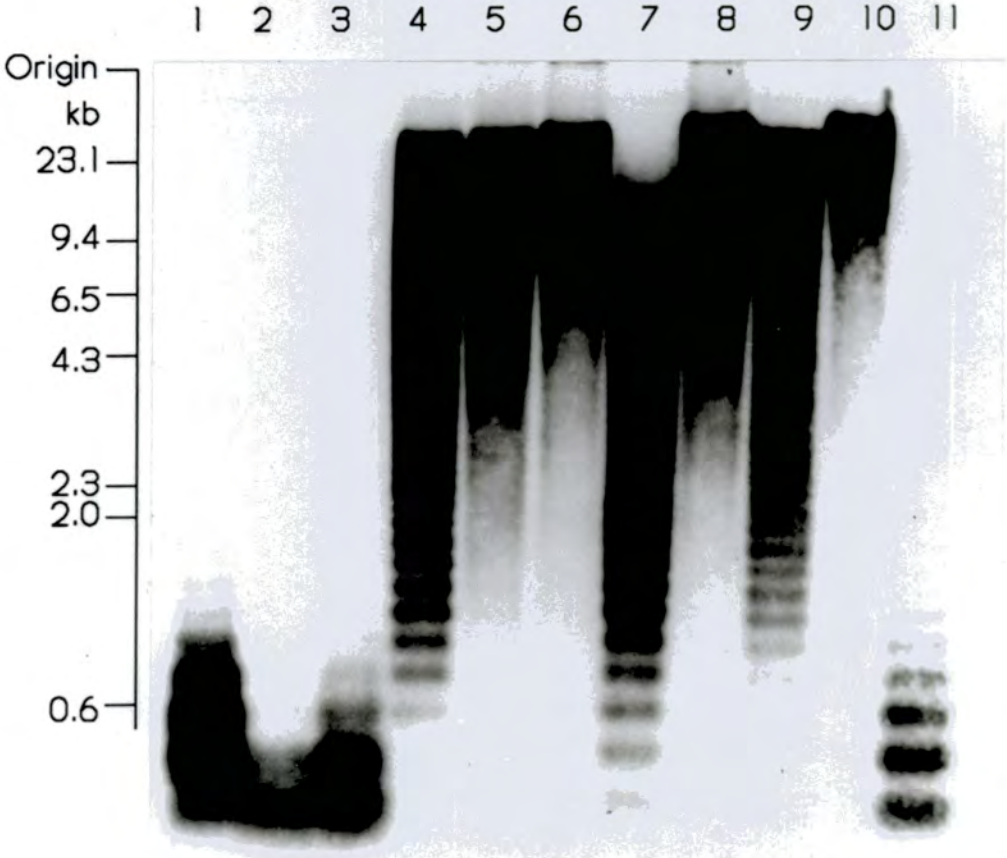


Figure 3.26



I, which is resistant to 5-MC methylation at the same position (Lanes 6 and 7 respectively). N^6-methyl adenine methylation (A-methylation) which is reported not to be present in plant DNA (Vanyushin *et al.* (1960) was also studied. Restriction enzyme *Mbo* I which is sensitive to A-methylation was able to restrict rape DNA (Lane 9) successfully. On the other hand, *Dpn* I, an enzyme which requires A-methylation for its endonuclease activity, failed to restrict rape DNA (Lane 8). Unrestricted rape DNA (lane 10) was included for comparison. It was apparent that there was some degree of degradation of rape DNA digested with *Dpn* I when compared with the unrestricted rape DNA (Lanes 8 and 10 respectively) though this was not thought to be due to *Dpn* I restriction. Therefore, consistent with the findings of Vanyushin *et al.* (1960), there is no evidence for N^6-methyl adenine methylation in rape DNA.

In plants, the principal methylated nucleic acid base is 5-MC and accounts for up to 32% of all cytosine residues (Vanyushin et al., 1960). The two methylated sequences occurring in plants as well as animals are CpG and CpNpG (where N is any nucleotide). The latter methylation, CpNpG was first suggested by Gruenbaum *et al.* (1981) and accounted for most of the difference in methylation level between plants and animals. As shown by the methylation study carried out earlier, rape DNA evidently was also methylated at the 5-methyl cytosine residues of both CpG (internal cytosine), as shown by the two pairs of isoschizomers, *Hpa* II (sensitive) and *Msp* I (insensitive), and *Sma* I (sensitive) and *Xma* I (insensitive), as well as CpNpG (external cytosine) as exemplified by *Eco* RII (sensitive), *Bst* NI (insensitive) and *Apy* I (dependent).

In order to survey the methylation pattern in the regions of tandemly repetitive sequences of rape DNA, the gel in Figure 3.25 was Southern-blotted and probed with pRS1. The autoradiograph is shown in Figure 3.26. Identical step-ladder pattern with 180 bp monomer was observed in rape DNA digestions using *Apy* I, *Eco* RII and to a certain extent, *Bst* NI (Lanes 1, 3 and 2 respectively). The recognition site for these enzymes, which in this case is identical, is present in the monomeric sequence. For *Bst* NI, a major band of 180 bp and a minor band of 360 bp, with possibly a very faint third band of 540 bp, was observed. The effect of methylation in the region of highly repetitive DNA in the genome was clearly demonstrated by these enzymes where *Bst* NI (methylation-resistant)

managed to digest the step ladder pattern to a much greater extent compared with *Apy* I (methylation-dependent) and *Eco* RII (methylation-sensitive), where the step-ladder was more apparent. The apparently less prominent "step-ladder" pattern observed in *Bst* NI digest (Lane 2, Figure 3.26) was probably due to underexposure of the autoradiograph. However, the step-ladder observed in this case cannot be exclusively attributed to the methylation pattern of the DNA. The presence of modified recognition sites for these enzymes in rape DNA is another basis for the production of similar step-ladder patterns.

Other enzymes included in the study, namely *Msp* I, *Xma* I and *Mbo* I (Lanes 4, 7 and 9 respectively, Figure 3.26), showed "step-ladder" patterns in the autoradiograph similar to that of a partial *Hind* III digestion (Pattern 1 of Figure 3.21) although these enzyme showed a complete digestion as shown in Figure 3.25 (Lanes 4,7 and 9). Sequence analysis of the monomeric sequence revealed no restriction sites for these enzymes. It was speculated that as a result of point mutation(s), the sites for these enzymes was modified and hence not detected in this particular cloned monomeric sequence. So, the sequence was subjected to the analysis of degenerate recognition sites for these enzymes. It was found that in both *Msp* I and *Mbo* I, single-base mutations within the recognition site was enough to generate functional restriction sites for these enzymes while in the case of *Xma* I, at least double-base mutations was needed to produce a site. This phenomenon of single or double-base mutations which resulted in the modification of certain restriction sites was later confirmed when sequences from the polymeric forms of repetitive sequences (trimer and tetramer) were analysed (Section 3.4.5). It was found that single and double-base mutations have resulted in a modified *Hind* III recognition sites.

Closer analysis of the "step-ladder" pattern revealed by digestion of different enzymes showed that the size of polymers were all identical indicating that they were of the same identity.

An interesting observation worth noting is on the digestion with the frequent-cutting enzyme *Mbo* I, where the repetitive sequences remained relatively intact as shown by hybridisation to pRS1 (Figure 3.26, Lane 9) although a major proportion of the DNA was digested to sizes less than 1 kbp (Figure 3.25, Lane 9).

The endonuclease activity of *Hind* III can also be affected by the methylation pattern of the DNA. As shown in Figure 3.24, methylation of the C-residue (5-MC) in the *Hind* III recognition sequence will inhibit its endonuclease activity (Kessler *et al.* (1985). It was therefore possible that the stable higher polymeric forms of the repetitive sequences could be a consequence of the methylation state of rape DNA in the region of *Hind* III recognition sequences, thus rendering them resistant to *Hind* III digestion. This, however, did not exclude the possibility of internal *Hind* III sites present in the polymeric repetitive sequences modified due to factors such as point mutation and which were therefore resistant to *Hind* III digestion. The cloning of the higher polymeric forms of the repetitive sequence followed by sequence analysis was therefore deemed highly desirable in order to confirm this observation.

3.4.6 Repeat Sequence Comparison with Other *Cruciferae*

Genomic DNA samples from related plant species belonging to the same Family (*Cruciferae*) as oilseed rape, were obtained and used in an experiment designed to survey the presence of repetitive sequences homologous to the rape repetitive sequence isolated. The DNA samples were restricted to completion with *Hind* III, electrophoresed, southern-blotted, and probed with pRS1 at high stringency (0.1x SSC). Figures 3.27 and 3.28 show the EtBr-stained gel and the corresponding autoradiograph of the Southern blot.

The results showed that other *Brassica* species, namely *Brassica campestris, Brassica oleraceae, Brassica carinata*, and *Brassica juncea* gave the same characteristic "step-ladder" band pattern showing the presence of homologous sets of sequences in these species. The "step-ladder" band pattern observed, as well as the intensity of the bands were apparently identical to that of *Brassica napus*. *Brassica napus* is an amphidiploid hybrid of *Brassica oleraceae* and *Brassica campestris* and therefore might have been expected to contain such related sequences. Other *Cruciferae*, namely *Arabidopsis thaliana, Hesperis matronalis, and Iberis amara* (Lanes 6, 7, and 8; Figures 3.27 and 3.28) apparently do not have sequences homologous to pRS1. However, a "step-ladder" pattern similar to the *Brassica sp.* pattern was observed in the stained gel when *Iberis amara*

Figure 3.27 — Analysis of genomic DNA from other species belonging to the same Family *Cruciferaceae* as oilseed rape for homologous repetitive sequences. *Hind* III-digested genomic DNAs from several species belonging to the *Cruciferaceae* were electrophoresed in an 0.8% (w/v) agarose gel in the presence of ethidium bromide. The DNA's analysed included *Brassica napus*, *Brassica campestris*, *Brassica oleraceae*, *Brassica carinata*, *Brassica juncea*, *Arabidopsis thaliana*, *Hesperis matrionalis* and *Iberis amara* (lanes 1 to 8 respectively). All *Brassica sp.* showed the characteristic "step-ladder" band pattern. Genomic DNA from *Iberis amara* (Lane 8) also showed a similar "step-ladder" pattern.

Figure 3.28 — Autoradiograph of the Southern blot of *Hind* III-digested genomic DNAs from species belonging to the Family *Cruciferaceae* probed with pRS1. The step-ladder pattern characteristic of *Hind III*-digested rape DNA (Lane 1) is reproduced in the other *Brassica* species tested, namely *B. napus*, *B. campestris*, *B. oleraceae*, *B. carinata* and *B. juncea* (Lanes 1, 2, 3, 4 and 5 respectively) indicating the presence of homologous sequences in their DNAs while no hybridisation was observed to the non-*Brassica sp.*, namely *Arabidopsis thaliana*, *Hesperis matrionalis* and *Iberis amara* (Lanes 6, 7, 8 respectively). The filter was washed to high stringency of 0.1x SSC.

Figure 3.27

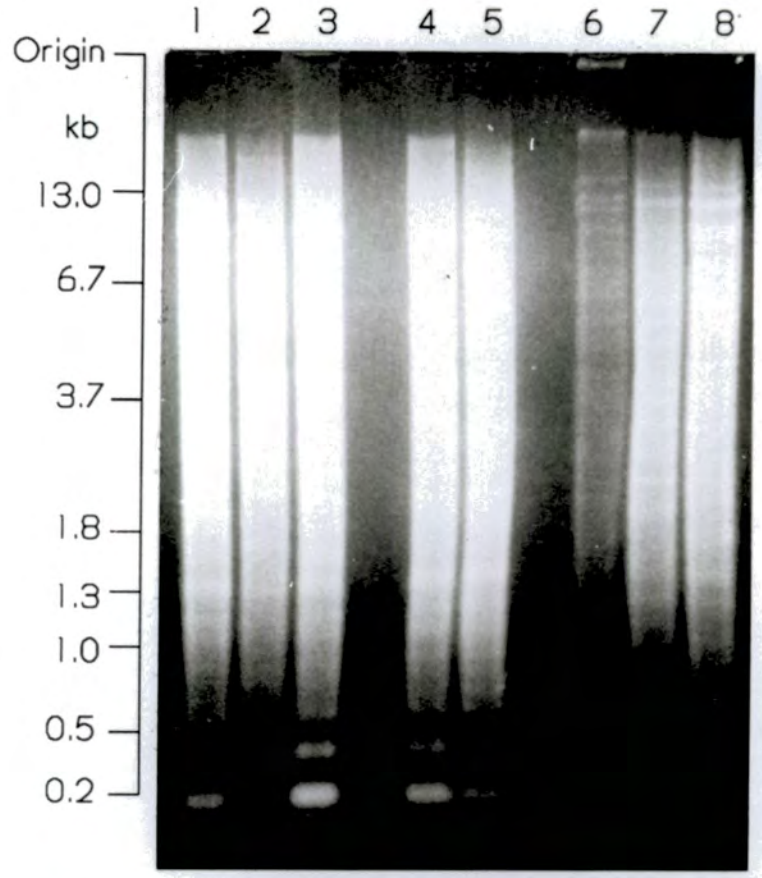
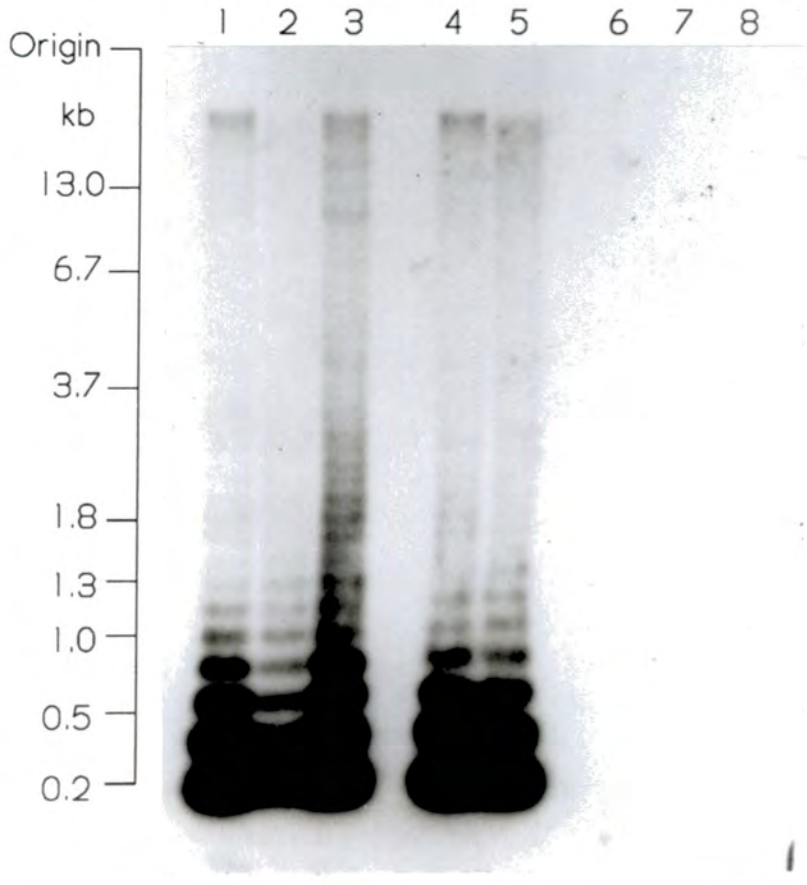


Figure 3.28



was digested with *Hind* III, indicating a similar though non-homologous *Hind* III repetitive sequence.

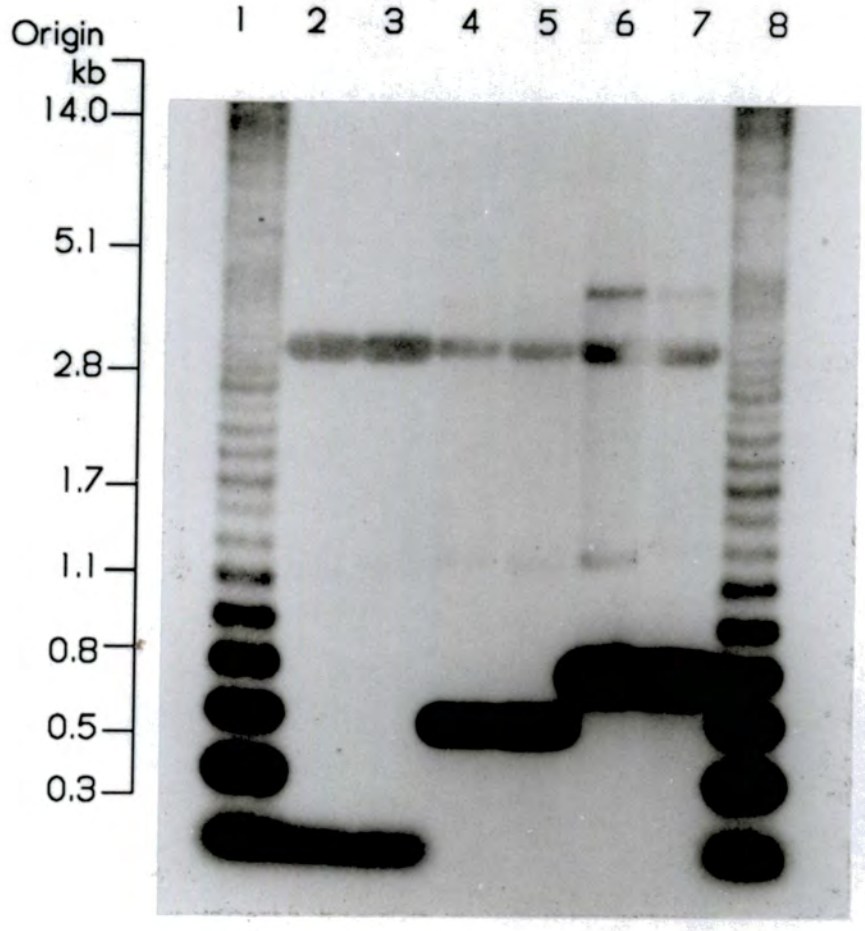
3.4.7 Cloning of the Higher Polymeric Forms

In previous experiments to test the effects of enzyme concentration and prolonged incubation period, the *Hind* III step-ladder pattern was stable even after 18 hours incubation time with fresh aliquots of restriction enzyme. This suggested that the higher polymeric forms of the repetitive sequence were resistant to *Hind* III digestion. An attempt was therefore made to clone and isolate representative higher polymeric forms resistant to *Hind* III digestion to investigate the underlying mechanism of resistance. Genomic DNA from rape variety Mikado was digested to completion according to the conditions established previously and separated on a 0.8% (w/v) agarose gel. The sections of the gel containing the fragments corresponding to the dimer, trimer, and tetramer were excised and each of the DNAs were separately recovered by electroelution. After phenol-chloroform extraction and ethanol precipitation the DNA samples were then subjected to a second *Hind* III digestion and the above process of isolation and purification repeated. After the second *Hind* III digestion, the three polymeric forms still retained their resistance to *Hind* III digestion and showed no evidence of the presence of the monomeric forms. These polymeric DNA fragments were individually cloned into *Hind* III-digested pUC18. The resulting recombinants were then screened by colony-hybridisation using pRS1 as probe. Positive clones were selected, plasmid DNA isolated and analysed by digestion with *Hind* III, electrophoresed, southern-blotted, and probed with pRS1 for positive confirmation of the clones. Two of the positive clones, designated pRS3 (trimer) and pRS4 (tetramer) were confirmed to contain the required inserts resistant to *Hind* III digestion. All positive clones analysed from the dimer cloning (pRS2) were subsequently shown to be susceptible to *Hind* III digestion giving bands corresponding to the monomer (data not shown).

Figure 3.29 shows the autoradiograph of the clones pRS3 and pRS4 (with and probed with pRS1 pRS1 as comparison) after digestion with *Hind* III. Clearly, the autoradiograph shows that the inserts from pRS3 and pRS4 correspond exactly with the trimer and the tetramer of the original *Hind* III-digested rape DNA respectively. The

Figure 3.29 — Autoradiograph of duplicate samples of polymeric clones pRS3 and pRS4 digested with *Hind* III . Duplicate samples of clones pRS3 (Lanes 4 and 5) and pRS4 (Lanes 6 and 7). They were digested with *Hind* III and were shown to give inserts corresponding to the original trimer and tetramer bands respectively of rape genomic DNA digested with *Hind* III (Lanes 1 and 8). pRS1 digested with *Hind* III was also analysed together for comparison (Lanes 2 and 3).

Figure 3.29



resistance of the inserts of pRS3 and pRS4 to prolonged and multiple *Hind* III digestions suggested that the internal *Hind* III restriction sites had been lost. The susceptibility of pRS2 (dimer) sequence to *Hind* III indicates the recovery of susceptible site from a previously resistant *Hind* III site.

The repetitive sequence clones, pRS1, pRS3, and pRS4 were subcloned into M13 and sequenced to examine the sequences and in particular, the regions of the *Hind* III internal restriction sites.

3.4.8 Sequence Analysis of Repetitive Sequence Clones

Complete sequence information was obtained for pRS1, pRS3 and pRS4 and is summarised in Figure 3.30. For the purpose of simplicity in presentation and ease of comparison, sequences from pRS3 and pRS4 have been presented as multiples of 3 and 4 monomeric sequences respectively, compared to the single monomeric sequence from pRS1, presented at the top.

From the sequence information analysed, it can be concluded that clones pRS3 and pRS4 contain repeat sequences highly homologous to the monomeric sequence of pRS1 arranged in tandem array of 3 (trimer) and 4 (tetramer) respectively (Figure 3.30). In order to illustrate the homologies between inserts of these clones precisely, the dot-matrix method of sequence comparison (Courtesy of Dr. R.R.D. Croy) was carried out and is presented in Figure 3.31. Homologous sequences between the monomer and itself, the monomer and the trimer and the monomer and the tetramer are indicated by straight line plot(s) (Figures 3.31(A), 3.31(B), and 3.31 (C), respectively). Thus, the 3 and 4 straight-line plots shown from left to right in Figures 3.31(B) and 3.31(C), clearly indicated that the monomeric sequence was repeated 3 times for the trimer and 4 times for the tetramer. The short straight-line plots observed beside the long straight-line plot indicated the presence of internal repeats and this will be discussed later in Section 3.4.11.

Closer examination was also carried out on both polymeric sequences in the region where the three and four repeat units were joined. It was found that all the "internal" *Hind* III sites of the trimer (two internal sites) and the tetramer (three internal sites) were lost. The 6 bp *Hind* III recognition sites were changed as a result of point mutations in at least one of the six base positions. These mutations

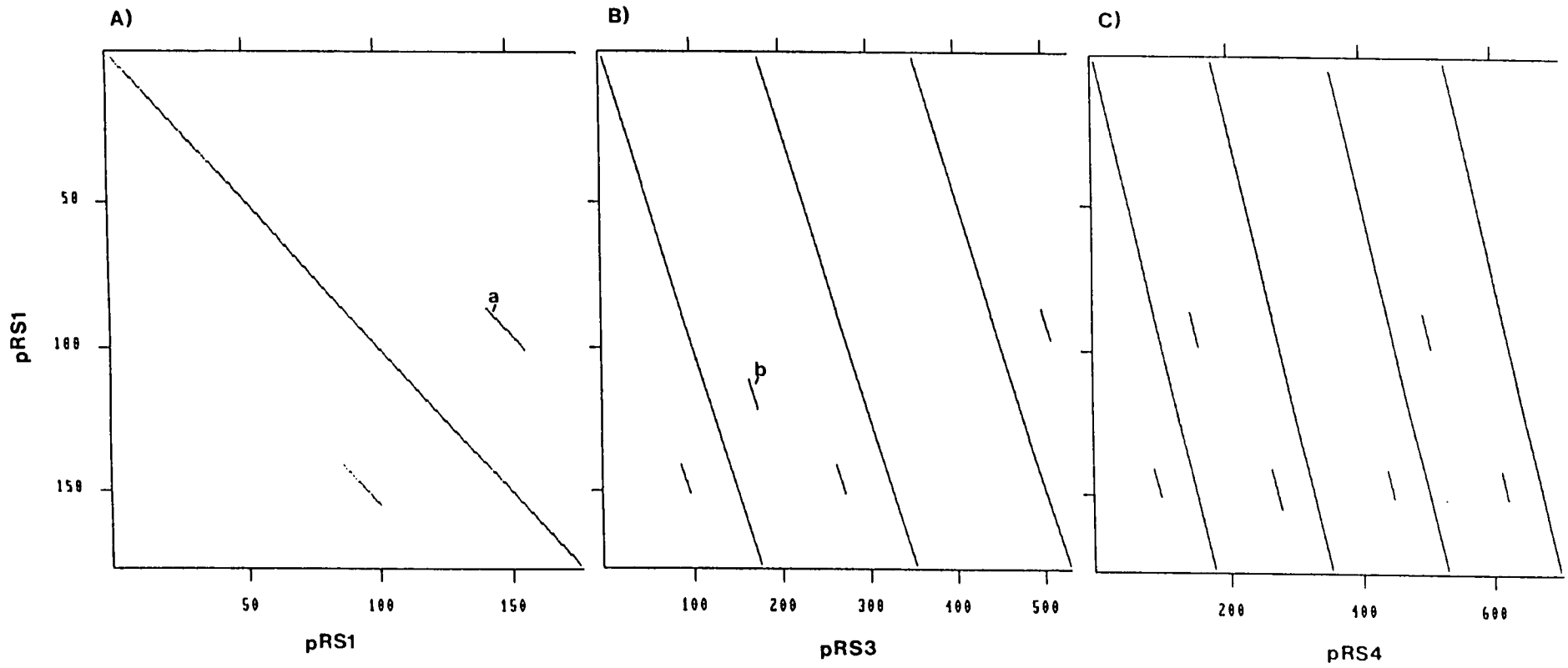
Figure 3.30 — Sequence comparisons of the repeat units in clones pRS1, pRS3 and pRS4 and the consensus repeat sequence. The eight copies of the monomeric sequence from clones pRS1, pRS3 and pRS4 were aligned and their consensus sequences determined by scoring the most frequent base residue at each of the 176 bp positions. Positions in the monomeric sequence which are different from the consensus sequence were indicated by (∧).

Figure 3.30

```
1   AGCTTGATTTGGATACATAAAGTGGTGGAGAATCACCAGGAAGTTGAATAAATCTCATAG PRS1

1   AGCTTGATTTGGATACCTAAAGTGGTGGAGAATCACCAGGAAGTTGAATAAATCTCATAG PRS3(1)
                  ^

1   ATCTTGATTTGGATACATAAAGTGGTGGAGAATCACCAGGAAGTTCAATAAATCTCATAG PRS3(2)
     ^                                            ^

1   AGATTTATTTGGATACATAAAGTGGTGGAGAATCACCAGGACGTTGGATAAATCTCATAG PRS3(3)
      ^  ^                                    ^    ^

1   AGCTTGATTTGGATACATAAAGTGGTGCAGAATCACCAGGAAGTTGAATAAATCTCTTAG PRS4(1)
                                 ^                          ^

1   ACCTTGATTTGGATAGATAAAGTGGTGGAGAATCACCAGGAAGTTGAATATATCTCATAG PRS4(2)
     ^               ^                                  ^

1   TCCTTGATTTGGATACATAAAGTGGTGCAGAATCACCAGGAAGTTGAGTAAATCTCATAG PRS4(3)
    ^^                               ^                  ^

1   ATCTTGATTTGGATACATACAGTGGTGGAGAATCACCAGGAAGTTGAATAAATCTCATAG PRS4(4)
     ^                    ^

1   AGCTTGATTTGGATACATAAAGTGGTGGAGAATCACCAGGAAGTTGAATAAATCTCATAG CONS.


61  GAGTTGGCATGAAGAAGTTATCCCACTTTCAAATCAGGTGATTCCAGTTTCCCAGTTTGG PRS1
             ^

61  GAGTTGGCAAGAAGAAGTTATCCCACTTTCAAATCAGGTGATTCCAGTTTCCCAGTTTAG PRS3(1)
                                                           ^

61  GAGTTGGCAAGAAGAAGTTATCCCACTTTCAAATCAGGTGATTCCAGTTTCCCAGTTTGG PRS3(2)

61  CAGTTGGCAAGAAGAAGTTATCCCACTATGAAATCAGGTGATTCCAGTTTCCCAGTTTGG PRS3(3)
    ^                                ^ ^

61  GAGTTGGCAAGAAGAAGTTATCCCACTTTCAAATCAGGTGATTCCAGTTTCCCAGTTTGG PRS4(1)

61  CAGTTGGCATGATGAAGTTATCCCACTTTCAAATCAGGCGATTCCAGTTTCCCAGTTTGG PRS4(2)
    ^         ^  ^                        ^

61  GAGTTGGCATGATGAAGTTATCCCACATTCAAATCAGGTGATTCCAGATTCCCAGTTTGG PRS4(3)
              ^  ^                 ^                   ^

61  GACTAGGCAAGAAGAAGTTATCCCACTTTCAAATCAGGTGATTCCAGTTTCCCAGTTTGG PRS4(4)
     ^ ^

61  GAGTTGGCAAGAAGAAGTTATCCCACTTTCAAATCAGGTGATTCCAGTTTCCCAGTTTGG CONS.


121 GAATAGCACAGCTTCTTCGTCGTTCCAATCAGACGAGGATGAATCTGTTTGTAAGA      PRS1
                                      ^  ^                  ^

121 GAATAAGACAGCTTCTTCGTCGTTCCAATCAAACCAGGATGACTCAGTTTGTGAGA      PRS3(1)
         ^^                                    ^  ^

121 GAATAGCACAGCTTCTTCGTCGTTCCAATCAAACCAGGATGAATCTCTTTGTGAGA      PRS3(2)
                                                  ^

121 GGATAGCACAGCTTCCTCGTCGTTCCATTCAAACCAGGATGAATCTGTTTGTAAGA      PRS3(3)
     ^                ^                ^                  ^

121 GAGTAGCACACCTTCTTCGTCGTTCCAATCAGACCAGGATGAATCTGTTTGTGAGA      PRS4(1)
      ^         ^                       ^

121 GAATAGCACAGCTTCTTCGCCGTTCCAATCAAACCAGGATGAATCTGTTTGTGAGA      PRS4(2)
                        ^

121 GAATAGCTCAGCTTGTTCGTCGTTCCAATCAGACCAGGATGAATCTGTTTGTGAGA      PRS4(3)
             ^          ^                ^

121 GAATAGGACAGCTTCTTTGTCGTTCCAATCAAACCAGGATGATTCTGTTTGTAAGA      PRS4(4)
            ^           ^                          ^          ^

121 GAATAGCACAGCTTCTTCGTCGTTCCAATCAAACCAGGATGAATCTGTTTGTGAGA      CONS.
```

Figure 3.31. Homology comparison of inserts from clones pRS1, pRS3, and pRS4 using a dot-matrix method. The monomeric sequence from pRS1 was plotted separately against complete sequences of (A) itself (monomer), (B) pRS3, and (C) pRS4 inserts. The monomeric sequence is plotted from top to bottom, and the sequence to be compared is plotted from left to right of each graph. Region of homologies are indicated by straight-line plot(s). The presence of internal repeats is indicated by the short straight-line plots as exemplified by a and b.

Figure 3.31


A)
B)
C)
pRS1
50
100
150
a
b
pRS1
pRS3
100
200
300
400
500
pRS4
200
400
600

rendered the sites resistant to *Hind* III digestion. Two nucleotide changes occurred in the second junction of pRS4 (Positions 2 and 3). Sequence information on the modified *Hind* III sites at the junctions between the repetitive units is presented in Figure 3.32.

From the available data, it can be concluded that as a result of point mutations, nucleotide changes occurred most frequently (4 times) at Position 3 (replacing 'G') of the 6 bp recognition site while a mutation occurred at both Position 2 (replacing 'A') and Position 4 (replacing 'C'). At Position 3, G was changed to either C or T (purine to pyrimidine transversion), at Position 2, A was changed to T (purine to pyrimidine transversion) and at Position 4, C was converted to A (pyrimidine to purine transversion. So, the predominant type of point mutation resulted in the loss of *Hind* III sites was transversion from purine residue to pyrimidine residue or *vice versa.* Positions 5 and 6 were not changed in any site sequenced.

3.4.9 Comparison of the Monomeric and Consensus Sequences

As shown in Figure 3.33, the eight copies of the monomeric sequences from the 3 clones were aligned for comparison. It was found that the homology between any two of the eight monomeric sequences was at least 98% showing that these repetitive sequences present in rape DNA are highly conserved. However, no two monomeric sequences amongst the eight analysed were identical indicating a low degree of heterogeneity. The consensus sequence for the monomer was deduced from the eight repeated monomeric sequences and is shown in Figure 3.30. The exact length of the monomeric unit is 176 bp. The deviation from the consensus sequence of all the eight monomeric sequences was less than 3% and the nucleotide differences were scattered randomly along the length of the monomeric sequence indicating the absence of any mutational 'hot-spot'.

Closer examination of the consensus sequence revealed that each of the 176 bp monomers could be subdivided into three 58/59 bp subsequences. The 58/59 bp subunits can be aligned with each other by aligning the short sequences ACCAGG, ATCAGG and ACCAGG (Boxed) as shown in Figure 3.33. The presence of these short sequences within the monomeric units which can be aligned, is interesting as they provide evidence that the monomer can be further subdivided into three

1 2 3 4 5 6	(Nucleotide position)
A A G C T T GATTT	(*Hind* III recognition site)
GTGAG A A **T** C T T GATTT	(1st junction of pRS3)
GTGAG A A G **A** T T TATTT	(2nd junction of pRS3)
GTGAG A A **C** C T T GATTT	(1st junction of pRS4)
GTGAG A **T** **C** C T T GATTT	(2nd junction of pRS4)
GTGAG A A **T** C T T GATTT	(3rd junction of pRS4)

Figure 3.32 — Changes in *Hind* III recognition sequence at the junctions between two monomeric sequences in clones pRS3 and pRS4. Nucleotide positions where changes have take place are in **bold** for each modified site. The two external *Hind* III sites used for cloning were not altered.

58/59 bp subsequences, though these sequences are highly divergent in nature. The sequence homology between these subsequences with one another only ranged from 33% to 48%. Kato *et al.* (1984) reported a 59 bp repeat in broad bean, while Grellet *et al.* (1986) reported a 177 bp repeat (multiples of 59 bp) in radish which belonged to the same *Cruciferaceae* family as *B. napus.* Grellet *et al.* (1986) further contended that the monomer could be further divided into three shorter related blocks with about 60% homology. He further concluded that these divergent blocks of DNA were patchworks of short sequences as indicated by the few direct perfect repeats and several imperfect ones suggesting that many interactions between these blocks have taken place before they were amplified as a single block i.e. as the whole monomer (Grellet *et al.*, 1986).

3.4.10 Sites of Possible DNA methylation

According to Adams *et al.* (1983) and Deumling (1981), those sequences containing the most methylation represent for the most part the repeated or heterochromatic regions of the genome. Consequently, the consensus monomeric sequence was analysed for the presence of CpG and CpNpG (5-MC methylation sites). The preceeding section described the presence of the sequence CCAGG at 2 positions in the consensus monomeric sequence (one position in the monomeric sequence of pRS1). The methylation pattern of rape DNA at these sites may affect the enzymatic activity of *Apy* I (methylation-dependent) and *Eco* RII (methylation-sensitive) compared with *Bst* NI (methylation insensitive). Although the restriction sites were still intact, 5-MC methylation in some of these sites will affect the enzymatic activity of *Apy* I and *Eco* RII and could therefore give rise to the "step-ladder" band pattern. The lower degree of enzymatic activity as shown by the comparatively less prominent "step-ladder" for *Bst* NI digestion (Lane 2, Figure 3.26) further supports the speculation of a possible effect of the methylation. Enzyme *Bst* NI is resistant to 5-MC methylation and therefore can cleave all the susceptible sites in rape DNA. As discussed previously, the CCAGG sites could be modified and therefore resistant to the digestion, in this case, by *Bst* NI. Sequence information obtained from clone pRS1 indicated that one of the CCAGG sites was modified by point mutation suggesting these sites are also prone to mutation as has been discussed earlier with *Hind* III recognition sites. Further experiments are therefore necessary which involve the cloning of the polymeric

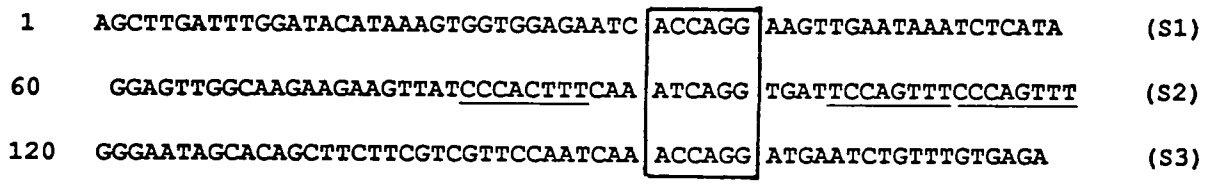


Figure 3.33 — Three subunits of the *Brassica napus* consensus monomeric sequence. The three subunits present in the consensus monomeric sequence aligned at the short sequences ACCAGG and ATCAGG (boxed). When so aligned each subunit contains 58 or 59 bp. Three 8-nucleotide direct repeats in one of the subunits are underlined.

form of *Bst* NI-digested rape DNA and analysing the sequence especially at the junction(s) between two monomers, to prove this.

The difficulty in cloning dimers might also be attributed to the methylation state of the DNA. It was possible that the dimeric DNA fragments generated by *Hind* III digestion were inhibited by methylation during the isolation of the fragments and then the site were recovered when cloned in methylase-bacterial strain.

Besides the methylation sites just described which were associated with the isoschizomers examined, the presence of other CpG and CpNpG motifs was also examined in the consensus sequence. As shown in Table 3.3, there are 12 methylation sites in the monomeric sequence; six of each type. Hepburn *et al.* (1987) contended that the presence of large numbers of these CpG and CpNpG "islands" is always associated with high mutation rates. Therefore it can be speculated that the presence of these islands in the repetitive sequence of rape may have been the cause of most of the mutation observed in the sequence analysis and the production of the polymeric forms.

3.4.11 Further Characterisation of the Consensus Sequence

Flavell (1982) proposed that the evolution of repetitive sequences began with the amplification of short, moderately heterogenous pieces of DNA sequences, sometimes only a few bases long. This was followed by rearrangements and finally, the resulting diverged arrays, were reamplified and this cyclic event was repeated until eventually fixed in the species (homogenisation). Earlier studies on the characterisation of repetitive sequences have mentioned the presence of short, unique sequences in the form of direct and inverted repeat, palindromic sequences, etc. in the sequence analysed (Reddy *et al.*, 1989; Simoens *et al.*, 1988; Benslimane *et al.*, 1986; Grellet *et al.*, 1986). The 176 bp repetitive sequence was therefore further analysed for shorter unique sequences and the result was summarised in Table 3.3.

Besides the presence of simple direct and inverted repeats of 3 to 5 bp in length in the consensus monomeric sequence, the sequence also contain other unique sequences. The three direct repeats of the sequence (C/T)CCA(C/G)TTT

Table 3.3 — Summary of Unique Sequences found in the Consensus Sequence of the Monomer.

Sequence Type	Nucleotide Sequence	Position[a]
Direct Repeat	(C/T)CCA(C/G)TTT	82, 104, 110
"	GAAGTT	40, 75
"	GAATA	46, 122
"	AGCTT	1, 131
"	AATCA	30, 49
Inverted Repeat	GATAC	12
"	CATAG	56
Inv. Complement	TGATTTG	5
"	CAAATCA	91
Palindromic	GAAGAAG	71
"	TGTTTGT	166
"	GCTTCTTCG	133
Methylation	CpG	2, 67, 127, 132, 39, 42
"	CpNpG	37, 96, 106, 114, 130, 155

[a] denotes the position of the nucleotide at the start of the sequence.

82 to 118 were unique in that they occurred within only one of the 58/59 bp subunits. It was possible that these 8-nucleotide tandem repeat units represent one of the building blocks of the 58/59 bp subunits which in turn became the building blocks of the 176 bp monomeric sequence. However, perhaps due to the highly divergent nature of the 58/59 bp subunits this 8-nucleotide repeat unit was not present in the other two subunits. Another unique sequence was the heptanucleotide TGATTTG (Position 5 to 11) where its complementary inverted repeat CAAATCA was found between Position 91 to 97. Other unique sequences include the two palindromic sequence GAAGAAG (Position 71 to 77) and TGTTTGT (Position 166 to 172).

3.4.12 Sequence Homology With Other *Cruciferaceae*

After the complete sequencing of the monomer, trimer and the tetramer of the repetitive sequences, the monomeric sequence was analysed and compared with published sequences in Genetic Sequence Data Bank "GenBank", NIH. It was found that the sequence was highly homologous to several satellite DNAs from a number of plant species including those belonging to the *Cruciferaceae* family. The definition of satellite DNA has been discussed earlier in the Introduction. A comparative study between these published sequences and the *Brassica napus* consensus monomeric sequence was carried out using MicroGenie Software running on an IBM-AT compatible 386 computer. Homology of 97-98% was observed between the consensus monomeric sequences of closely related *Brassica* species including the diploid *B. oleraceae* (Benslimane *et al.*, 1986) and *B. campestris* (Lakshmikumaran and Ranade, 1990), the two ancestors of *B. napus*, and also the amphidiploid *B. juncea* (Reddy *et al.*, 1989). High homology was also found between the rape sequence and sequences from *Sinapis alba* (white mustard) (75%) (Capecius, 1983) and *Raphanus sativus* (radish) (73%) (Grellet *et al.*, 1986). For *Arabidopsis thaliana*'s satellite DNA, a non-Brassica *Cruciferaceae*, the homology was only 40% (Simoens *et al.*, 1988).

3.4.13 The Copy Numbers of Rape Monomeric Sequences

An experiment was carried out to estimate the total number of copies of the monomeric sequences in the rape genome using the method described by Grellet

et al. (1986) for radish. Gene copy equivalents of the monomeric sequence were calculated using the formula described in the Materials and Methods and the calculation indicated that 1.79 pg of plasmid pRS1 contained the equivalent of 1 gene copy per μg of rape genomic DNA. Fluorimetric estimation (DABA assay) as well as the normal spectrophotometric estimation were used to accurately determine the concentration of DNAs to be used in these experiments. Two identical experiments were carried out to determine the copy number. The slot-blot filter was probed with excised insert from pRS1 and the autoradiograph is shown in Figure 3.34.

After exposure, the sections of the filter containing hybridised ^{32}P labelled DNA from each dilution of plasmid pRS1, rape genomic DNA and control pUC18 were excised and each was subjected to scintillation counting. The counts in counts per minute (cpm) from the rape genomic DNA dilutions and from the pRS1 dilutions after substracting counts from the corresponding pUC18 control dilutions, were plotted separately against the total amount (in μg) of rape genomic DNA and gene copy equivalents of pRS1 respectively. Figures 3.35 and 3.36 depict the graphs obtained by plotting cpm versus the concentration of the genomic DNA and also corrected cpm versus the gene copy equivalent of pRS1 respectively. These figures were intended to illustrate the linear correlation in both of the above cases and only the relevant points were plotted.

A linear correlation between the cpm and the amount of rape genomic DNA was obtained at dilutions corresponding to 0.25 μg and below while that of the cpm and gene copy equivalents (pRS1) was observed at dilutions corresponding to 0.25 million copies and below; In both cases, the cpm at higher concentrations were less than expected indicating that some interference in the hybridisation was taking place . Similar phenomenon was also reported by Grellet et al., (1986). The estimation of the number of copies of the monomeric sequence was then calculated by first determining the corresponding cpm for an amount of genomic DNA on the first graph and secondly used that cpm reading to determine the corresponding gene copy equivalent on the second graph. The cpm readings were taken within the region of linear correlation of the two variables. For example, 0.17 μg of genomic DNA gave a cpm reading which corresponded to about 0.05 million copies.

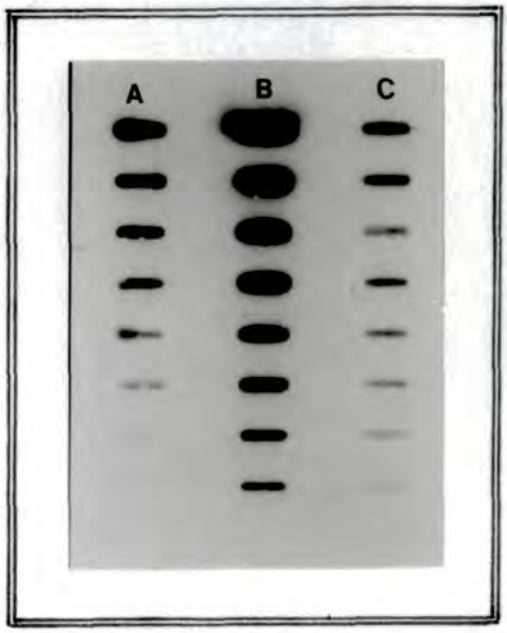


Figure 3.34 — Autoradiograph of slot-blot to estimate copy numbers of the monomeric sequences in rape genomic DNA. Slot-blots of a serial 2-fold dilutions of rape genomic DNA, pRS1, and pUC18 (Rows A, B, and C). Serial two-fold dilutions of the amount of plasmid pRS1 corresponding to 2 x 10^6, 10^6, 5 x 10^5, down to 1.5 x 10^4 gene copies per haploid genome and a serial two-fold dilutions of rape genomic DNA at 1 μg, 0.5 μg, down to 7.8 ng were slot-blotted onto nylon filter using a Minifold II Slot-Blot Apparatus (Schleicher and Schuell, Dassel, W. Germany) according to the procedure for dot-blotting described in Sambrook *et al.*, (1989). Plasmids pUC18 (background control), serially diluted to the same DNA concentration as pRS1 were also included in the experiment for comparison. Non-linear correlation of the bound probes versus the amount of DNA was apparent at higher dilutions.

Figure 3.35 — Graph of the bound ^{32}P-DNA(cpm) versus the concentration of Rape Genomic DNA. From the standard curve, a linear correlation was observed at dilutions corresponding to 0.25 μg genomic DNA and below. The values for calculations of gene copy equivalents were read within this region of linear correlation.

Figure 3.36 — Graph of the bound ^{32}P-DNA (cpm) versus the gene copy equivalents of clone pRS1. From this graph, linear correlation was observed at dilutions corresponding to 0.25 million copies and below. As before, the values for gene equivalent calculations were read within this region of linear correlation.

Figure 3.35

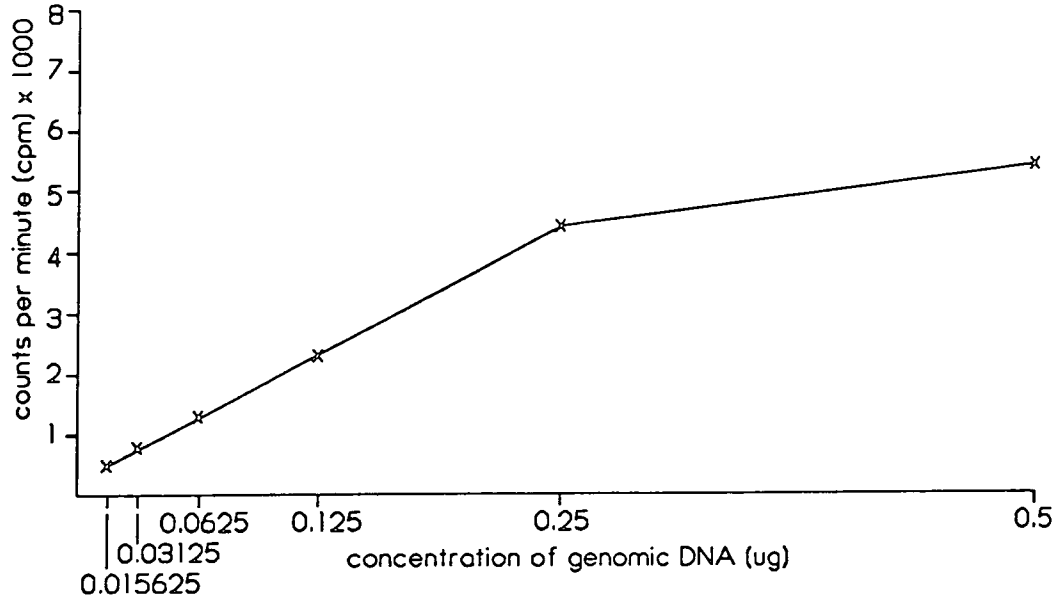
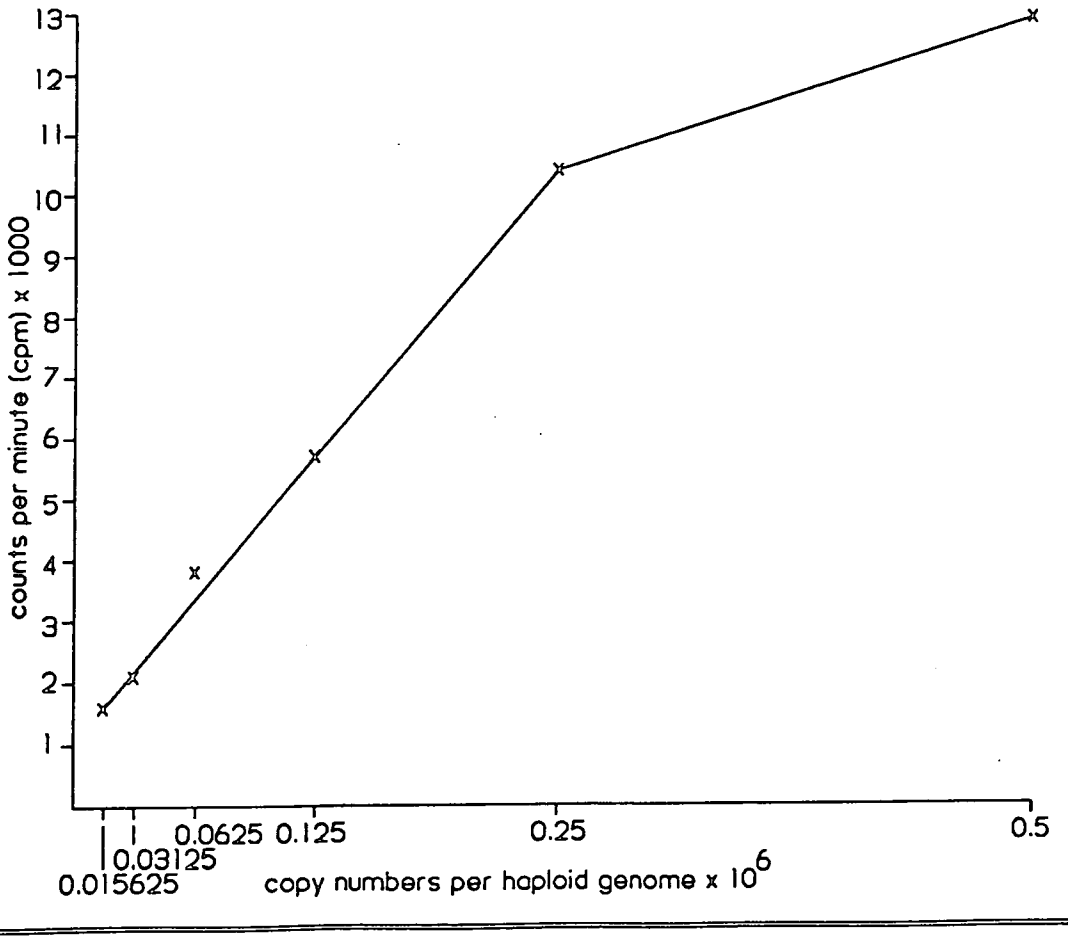


Figure 3.36



Therefore, the gene copy equivalence per haploid genome for 1 μg of genomic DNA, was calculated to be 0.29 million copies.

The result from two independent experiments described above showed that there are approximately 0.3 million copies of the repetitive sequences (monomer equivalents) per haploid genome. In radish, the number of copies of the monomeric sequences in a diploid genome was estimated at approximately 0.6 million (Grellet *et al.*, 1986) which is equivalent to 0.3 million copies per haploid genome.

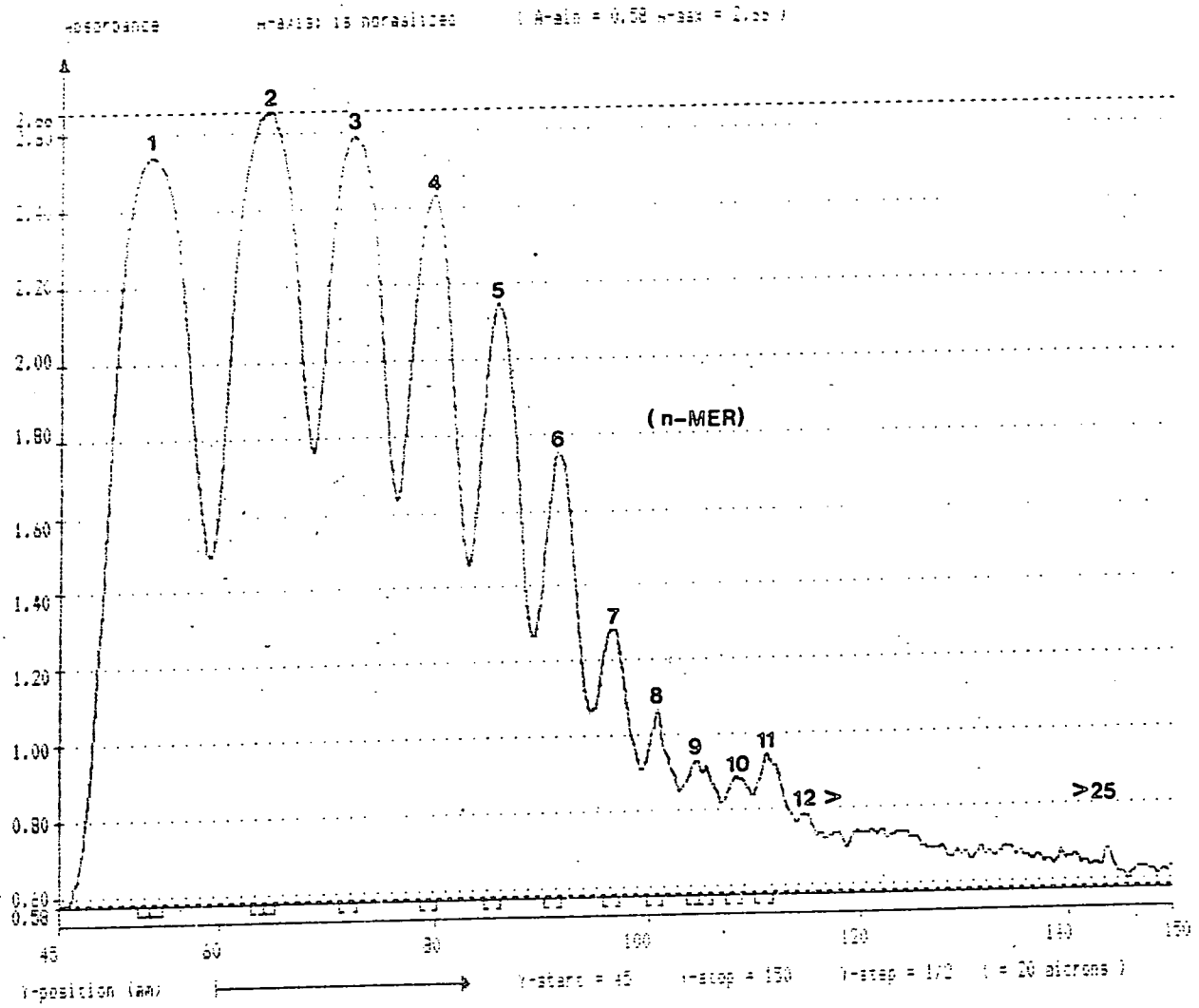
3.4.14 Estimation of the Copy Numbers of the Polymeric Forms

Copy numbers of the individual polymeric forms of the repetitive sequencs in the rape genome were estimated by determining the proportion of each polymeric form and calculating the copy numbers based on the estimates of the total number of copies of the monomeric sequences in rape genome. Autoradiographs of the "step-ladder" patterns from three DNA samples, representative of a complete digestion, were scanned with a laser densitometer (LKB Ultroscan, LKB Biotechnologies) and the area under each peak was estimated. Figure 3.37 depicts one of the densitometric traces obtained.

From the densitometric trace, it was found that the densitometer was able to resolve only eleven peaks and the area under each of the peaks was calculated as relative percentage of these peaks. After scanning the first eleven peaks, the densitometer was unable to resolve any more bands since they were very close to each other and they were of low intensity and there was significant interference from the background. Earlier observation of at least 25 visually identifiable bands was carried out on the basis of determining the longest polymeric sequences occurring naturally in rape genome by examining the "step-ladder" from an incomplete *Hind* III digestion. Thus, the bands visualised earlier of up to at least 25-mer, might represent, in a large proportion, restrictable polymeric sequences containing susceptible internal *Hind* III sites. However, this did not exclude the possibility that genuine *Hind* III-resistant 25-mer repetitive sequence were present in a very low copy numbers which were visually non-detectable in a complete digestion. It was apparent however that the bands beyond the 11-mer were extremely minor and together amounted to no more a few percent of the total.

Figure 3.37 — Densitometric trace of the "step-ladder" pattern of repetitive sequences in the rape genome. The "step-ladder" pattern representative of a complete *Hind* III digestion, was scanned with a laser densitometer (LKB Ultroscan, LKB Biotechnologies). The peaks starting from left are the monomer, dimer, trimer, etc., of the "step-ladder". Areas under the peak and the relative percentages of the peak were automatically calculated and the information obtained was used in the estimation of the copy number of each polymeric sequence.

Figure 3.37


Absorbance
A
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
1
2
3
4
5
6
7
8
9
10
11
12 >
>25
(n-MER)
45
60
80
100
120
140
150
Y-position (mm)
Y-start = 45
Y-stop = 150
Y-step = 1/2

It was noted in the densitometric trace that the peak of the monomer was relatively lower than that of the dimer. From the stained gel (Figure 3.17), it was observed that the monomeric sequences were the major restriction product in the step-ladder. It is possible that this apparent discrepancy was due to the inability of smaller DNA fragments, like the 180 bp monomers, to efficiently bind to the nitrocellulose and therefore were lost during hybridisation. Another possible explanation was the spread of band where the lowest size of fragments spreads more especially in Southern blotting. Therefore, the peak represent an underestimation of the actual copy number of the monomeric sequences.

The areas under the peak calculated as relative percentages were used to estimate the total copy numbers of the monomeric sequences in each peak. The total copy numbers of each polymeric sequence were then calculated by dividing the total monomeric copy numbers by the number of corresponding monomeric sequences present in the polymeric sequences. The result of these calculations is summarised in Table 3.4 and Figures 3.38 and 3.39. Figure 3.38 depicts the total number of repeat units in each of the polymeric forms plotted against the corresponding polymeric form while in Figure 3.39 depicts the estimated copy number of each polymeric form per haploid genome plotted against the corresponding polymeric form. From the plot in Figure 3.39, it was found that the monomeric sequences had the highest copy number per haploid genome although it was underestimated. This was followed by the subsequent polymeric sequences in decreasing copy numbers. The copy numbers per haploid genome of the higher polymeric form of the repetitive sequences decreased rapidly until the heptamer and decreased slowly in subsequent polymeric forms.

The results presented on the estimation of copy numbers of the polymeric sequence present in rape genome should be considered as an overestimation as the calculated percentages did not include those of other higher polymeric sequences (above 11-mer) which could not be resolved due to interferences from the background hybridisation and close proximity to each other, and also the underestimation of the copy numbers of the monomeric sequence due to inefficient binding during Southern blotting as mentioned earlier in this section.

Table 0.1 — The Copy Numbers of Polymeric Sequences Per Haploid Genome of Rape

Polymeric Sequence[a]	Copy Numbers of Monomers in each repetitive form	Copy Numbers of Polymers (per hap. genome)
1	1.17×10^5	2.34×10^4
2	1.22×10^5	6.10×10^4
3	8.31×10^4	2.77×10^4
4	7.54×10^4	1.87×10^4
5	6.33×10^4	1.26×10^4
6	4.59×10^4	7.65×10^3
7	2.81×10^4	4.11×10^3
8	1.84×10^4	2.34×10^3
9	2.01×10^4	2.23×10^3
10	1.20×10^4	1.20×10^3
11	1.42×10^4	1.29×10^3

[a] The numbers denote monomers (1), dimers (2), trimers (3), etc..

Figure 3.38 — Total Copy Numbers of the Monomeric Sequences for each Polymeric Forms of Repetitive Sequences. The total copy numbers of the monomeric sequences present in each polymeric form was estimated based on the areas under the peak of the densitograph. The total amount of genomic DNAs used was 5 μg.

Figure 3.39 — The Copy Numbers of each Polymeric Sequences Per Haploid Genome. The copy numbers of each polymeric sequence per haploid genome were plotted against their corresponding polymeric sequence.

Figure 3.38

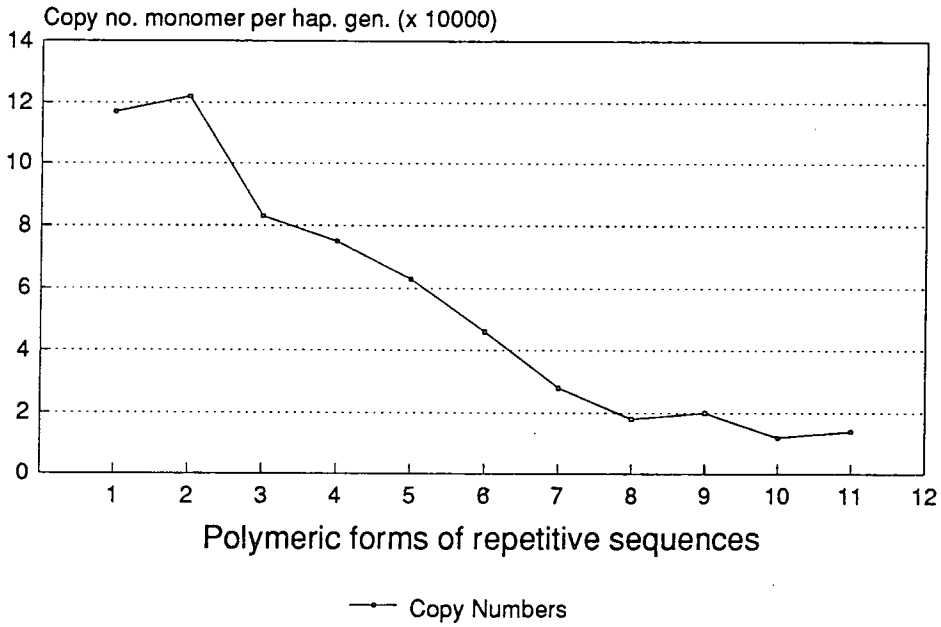
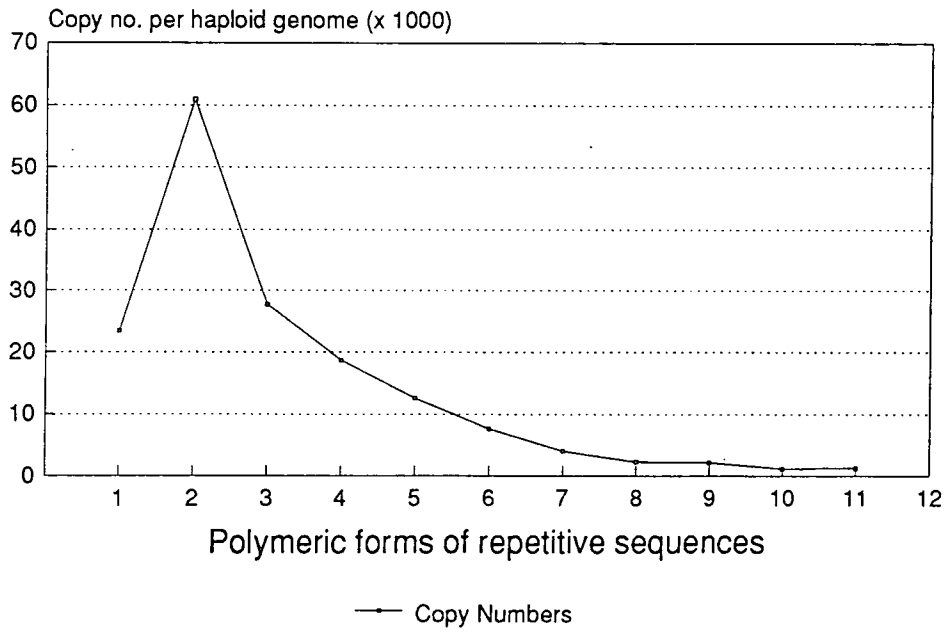


Figure 3.39



3.5 RANDOM AMPLIFIED POLYMORPHIC DNA (RAPD)

3.5.1 Rationale

DNA polymorphism assay is an important tool in genome analysis whereby some information about the genetic make-up of an individual can be determined. DNA markers generated from DNA polymorphism assays such as RFLP are proving extremely useful for trait introgression in plant and animal breeding programmes, gene cloning, DNA fingerprinting, etc., and especially if they are genetically linked to a trait of interest. RFLP analysis and its many applications and the attempts to identify cloned probes which could give RFLP patterns associated with glucosinolate content have been discussed in previous sections. Recently, a new method of DNA polymorphism assay based on the polymerase chain reaction (PCR) amplification of random DNA segments using single oligonucleotide primers of arbitrary nucleotide sequence, was simultaneously published by William *et al.* (1990), and Welsh and McClelland (1990). The method was called Random Amplified Polymorhic DNA (RAPD) by Williams *et al.* (1990) or Arbitrarily Primed PCR (AP-PCR) by Welsh and Mc Clelland (1990). The primers used in the reaction can detect DNA polymorhisms which in turn can function as useful genetic markers for the construction of genetic maps. As contended by Williams *et al.* (1990), genetic mapping using RAPD markers has, besides the rapidity of PCR technology, several additional advantages over other methods: (i) a universal set of primers for genomic analysis can be used for a wide variety of species. (ii) does not require the cloning, preparation of DNA probes and filters for hybridisation, and (iii) rapid and simple information transfer in collaborative research programmes where communicated primer sequences instead of material DNA clones are exchanged.

Due to RAPD's many advantages and especially so in its ability to even distinguish bacterial strains of the same species as suggested by Welsh and McClelland (1990) through genome fingerprinting, a study was initiated to explore the possibility of using the method to detect DNA polymorphisms in *B.napus* varieties. The possibility of RAPD's ability to detect polymorphisms which could be associated with glucosinolate content was also explored by using DNA samples from

rape varieties with known glucosinolate content. Initially, six phylogenetically related but distinct species of *Brassica* were analysed to determine the feasibility of using the RAPD technology to measure genetic relatedness among closely related species.

3.5.2 Sources of Primers and Template DNA

The arbitrary primers used in this study were synthesized as described in the Materials and Methods according to the sequence information described by Williams *et al.* (1990). Five 10-nucleotide long (10-mer) oligonucleotide primers were selected based on their ability to detect polymorphisms in soybean varieties as suggested by Williams *et al.* (1990). The primers were designated as AP4c, AP5a, AP12h, AP5h and AP6 according to Williams *et al.*, (1990) . In addition, six other oligonucleotide primers, 10-mers GE1, GE2 and GE3 and 9-mers GE4, GE5 and GE6 were a generously gift by Dr. G. Edwards (Shell Research Limited, Sittingbourne Research Centre, Sittingbourne, Kent) and were also used in this study. The sequences of the individual primers are shown in Table 3.5 . Firstly, template DNA from six phylogenetically related species – *B. napus*, *B. carinata*, *B. juncea*, *B. campestris*, *B. oleraceae*, and *B. nigra* were analysed to test the feasibility of the method. PCR amplification were performed as described in Materials and Methods. Secondly, template DNA samples from 9 low glucosinolate varieties – Ariana, Cobra, Drakkar, Libravo, Lictor, Lirawell, Optima, Topas and Bronowski, and 8 high glucosinolate varieties – Astec, Bienvenu, Brutor, Jet Neuf, Mikado, Pasha, Rafal and Willi were also analysed.

3.5.3 RAPD and its use as genetic marker

Like RFLP, RAPD polymorphism assay can offer information on the genetic variability of an individual and also serve as genetic markers by studying the segregation data of appropriate crosses. Due to its recent discovery, RAPD analysis has not been applied in many crop plant species apart from rice, soybean, and maize as described by Williams *et al.* (1990) and Welsh and McClelland (1990, 1991). So far, RAPD analysis has not been applied to *Brassica sp.*. In an attempt to explore the feasibility of using RAPD analysis to detect DNA polymorphism in *Brassica sp.*, genomic DNAs from closely related *Brassica sp.* were used as

Table 3.5 — The primer lengths and sequences used in RAPD analysis

Primer	No. of Bases	Sequence (5' — 3')
AP4c	10	TCTCGATGCA
AP5a	10	CTGTTGCTAC
AP12h	10	CGGCCCCTGT
AP5h	10	CACATGCTTC
AP6	10	GCAAGTAGCT
GE1	10	GGATCTCGAC
GE2	10	GCGTTCCATG
GE3	10	CAAGCCAGGA
GE4	9	GGATCTCGA
GE5	9	CAGTTGCGA
GE6	9	ATCGGAAGG

templates for the RAPD amplification reactions. Moreover, this study would also be useful in exploring the possibility of using RAPD in the study of phylogenetic relationship of plant species such as *Brassica sp.* as a measure of DNA relatedness among those species. The feasibility of using RAPD analysis to study varietal differences among varieties of oilseed rape was also explored.

Genomic DNA from six phylogenetically related *Brassica sp.* with known genotypes were used in RAPD analysis. The genotype of each *Brassica sp.* and their phylogenetic relationship has been described in the Introduction (Section 1.3.1). The template genomic DNAs were from *Brassica campestris* (aa), *Brassica nigra* (bb) and *Brassica oleraceae* (cc), all being primary species, while *Brassica juncea* (aabb), *Brassica carinata* (bbcc) and *Brassica napus* (aacc) are the amphidiploids resulting from hybridisations between the primary species.

All the primers used in the experiment were able to give polymorphic amplification products of varying degree of polymorphism. Some primers were able to give highly polymorphic band patterns while some gave relatively conserved polymorphic band patterns. The results of 4 of the primers used in the amplification reactions are presented in Figures 3.40, 3.41, 3.42, and 3.43 corresponding to primers GE1, AP5h, AP12h and AP5a respectively. In Figure 3.40, genomic DNA from *B. campestris* (1), *B. oleraceae* (2), *B. nigra* (3), *B. juncea* (4), *B. carinata* (5) and *B. napus* (6) was used in the amplification reactions using primer GE1. Another genomic DNA sample of *B.napus* but at a lower concentration of 5 ng instead of of 25 ng and control reaction without template DNA were also included in the reaction (Lanes 7 and 8 respectively). The inclusion of lower concentration of template DNA of *B. napus* in an amplification reaction was to check the effect of template DNA concentration on the generation of amplification products. Control reaction was included to confirm that the observed bands were amplified genomic DNA, and not primer artefacts (Williams *et al.* (1990) or non-specific amplifications. As shown in Figure 3.40, the polymorphic band patterns of each species though phylogenetically related were exceedingly complex with primer GE1, although some of the bands were apparently conserved between the species. There were at least 6 polymorphic positions in the amplification products using primer GE1. As an example, bands in *B. napus* (Lane 6) were compared with those of its two ancestors namely *B. campestris* (Lane 1) and *B.*

oleraceae (Lane 2). Polymorphic bands are indicated by numbered arrows where the numbers indicate the positions from the top of the gel.

From the analysis of the polymorphic bands of these three species, it can be suggested that at Positions 3, 5 and 6, the bands in *B. napus* (6) came from its two ancestors while at Positions 1 and 2 the bands came from *B. campestris* (1) alone. However, the band at Position 4 in *B. campestris* were not present in *B. napus* (6). Although similar bands were found at Positions 5 and 8 for *B. juncea* (4) and *B. campestris* and also at Position 7 for *B. nigra* (3) and *B. carinata* (5), the phylogenetic relationship of these species with their respective ancestors could not be determined conclusively as the polymorphic patterns of these amphidiploids were too complex. Further analysis of the band pattern revealed species-specific amplification products, examples of which were indicated by a, b, c, d and e for *B. campestris*, *B. oleraceae*, *B. juncea*, *B. carinata* and *B. napus* (Lanes 1 to 6) respectively. Also observed was the presence of genotype-specific bands at Position 2 where there is a strong evidence that this band belong to the aa genotype as it is present only in species containing that genotype i.e. *B. campestris* (1) and *B. napus* (6). The effect of template concentration on the amplification reaction was also studied whereby in Lane 7, due to the reduction of the template concentration, the intensity of the bands decreased considerably and thus difficult to be visually analysed. The effect of the concentration of template DNA on the generation of amplification products will be discussed in Section 3.5.5.

Another primer used in this analysis was AP5h and the pattern of bands produced is shown in Figure 3.41 . The polymorphic band patterns of each species were not as complex as with that of primer GE1 (Figure 3.40). Similar analysis was made by comparing the polymorphic band patterns of *B. napus* (Lanes 6 and 7) and that of *B. campestris* and *B. oleraceae* (Lanes 1 and 2 respectively). It was found that, with the exception of bands at Positions 2 and 5 which belong to *B. oleraceae*, all other bands in *B. napus* were present in both of its two ancestors although the band at Position 7 on *B. oleracea* was less intense. In fact, the band pattern of *B. campestris* (1) was identical to that of *B. napus* (6). Another comparison was made between *B. juncea* (4) and its ancestors *B. campestris* (1) and *B. nigra* (3). It was found that the band pattern of *B. juncea* was similar to that of *B. campestris* where 3 bands (Positions 3, 4 and 8) were present in both. When

Figure 3.40 — RAPD analysis of six phylogenetically related *Brassica sp.* using primer GE1. Template genomic DNA samples from six related *Brassica sp.* namely *B. campestris* (1), *B. oleracea* (2), *B. nigra* (3), *B. juncea* (4), *B. carinata* (5), and *B. napus* (6) (Lanes indicated in brackets). Two concentration of *B. napus* genomic DNA(25 ng, Lane 6)and (5 ng, Lane 7) were used as template in the RAPD analysis. Lane 8 contained the control reaction without added template DNA.

Figure 3.41 — RAPD analysis of six phylogenetically related *Brassica sp.* using primer AP5h. Template genomic DNA samples and lane assignment are exactly as described in above figure. The polymorphic bands patterns of *B. juncea* (Lane 4), *B. napus*, and *B. campestris* were highly conserved with *B. napus* and *B. campestris* showing identical band patterns. Species-specific bands were present in *B. nigra* (Lane 3, eg. Positions 1 and 7) and *B. oleracea* (Lane 2, Position 5) as indicated by arrows.

Figure 3.40

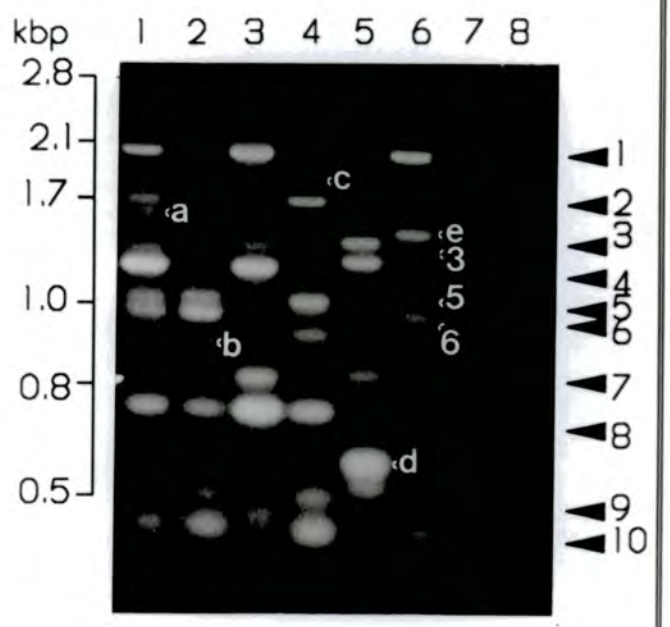
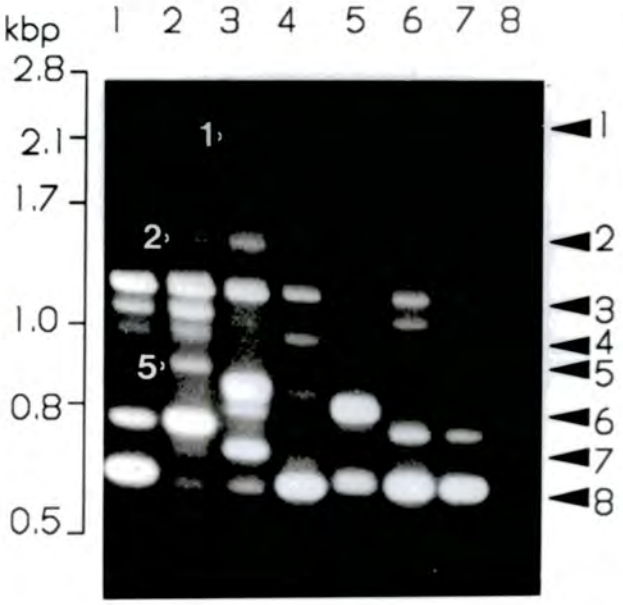


Figure 3.41



the band pattern of *B. juncea* (4) was compared to that of *B. nigra* (3), though several bands were present in both, at least 4 other bands that were present in *B. nigra* were not found in *B. juncea* (Positions 1, 2, 6 and 7). For the amphidiploid *B. carinata* (5), only two bands (Positions 6 and 8) were amplified by primer AP5h where both were also present in *B. nigra* (3), one of its ancestor. One species-specific band at Position 7 was present in *B. nigra*. In Lane 7 where the template DNA concentration was reduced to 5 ng instead of 25 ng, the intensity of the amplification products have markedly decreased and made the analysis difficult, thus, confirming the previous observation when GE1 was used.

Two other primers used in the amplification reaction that gave polymorphic amplification products were AP12 h and AP5a where the results are presented in Figures 3.42 and 3.43. By using primer AP12 h in the reaction it was observed in the case of *B. napus* (7) that the band pattern was similar to that of *B. campestris* (2). It was also observed that the pattern were, in general, highly conserved among the primary species (2, 3 and 4) but highly divergent among amphidiploid species (5, 6 and 7). However, in both cases, species-specific band patterns were observed in some species. Highly conserved pattern among an amphidiploid and its ancestors was observed when AP5a was used as depicted in Figure 3.43 where *B. campestris*, *B. oleraceae* and *B. napus* were represented in Lanes 1, 2 and 6 respectively. For other amphidiploids and their ancestors, some bands were found at identical positions of the polymorphic band patterns while some were unique to a particular species. Due to the hypervariability of the amplification products generated especially when using AP12h as primer (Figure 3.42), it was indeed possible to distinguish between the *Brassica sp.* analysed. As will be discussed later, primer AP12h belongs to a class of RAPD primers that can generate hypervariable amplification products that make it unsuitable in the analysis of genetic relatedness between species unlike primer AP5a and AP5h. The effect of increasing the template DNA concentration was also studied. In Figure 3.43, Lane 7, the concentration of the template DNA was doubled to 50 ng and as shown, the intensity of the bands were increased and therefore facilitate the analysis of the polymorphic bands. Therefore, it can be suggested that by increasing the concentration of the template DNA to an optimum level, it is possible to generate more amplification products of the existing bands and making them more prominent; thus facilitating the analysis and increased the confidence of scoring polymorphic bands. In Lane

Figure 3.42 — RAPD analysis of six phylogenetically related *Brassica sp.* using primer AP12h. Lane 1 contains the control reaction without template DNA. Lanes 2 to 7 is the amplification patterns of *B. campestris*, *B. oleraceae*, *B. nigra*, *B. juncea*, *B. carinata* and *B. napus*. The band patterns, though remarkably conserved among the primary *Brassica sp.* (Lanes 2, 3 and 4), are also species-specific. The amphidiploids share identical bands at some positions (arrowed) with their respective ancestors but the overall patterns were extremely divergent.

Figure 3.43 — RAPD analysis of six phylogenetically related *Brassica sp.* using primer AP5a. The genomic template DNA were from *Brassica sp.* of *B. campestris*, *B. oleracea*, and *B. nigra* (Lanes 1, 2 and 3) and *B. juncea*, *B. napus*, *B. napus* (with 50 ng instead of normal 25 ng of template DNA), and *B. carinata* (Lanes 5, 6, 7 and 8). Lane 4 is the control reaction. The band patterns were identical for *B. napus* (Lanes 6 and 7) and *B. campestris* (Lane 1). In Lane 7, the doubling of the concentration of the template DNA increases the intensity of the bands.

Figure 3.42

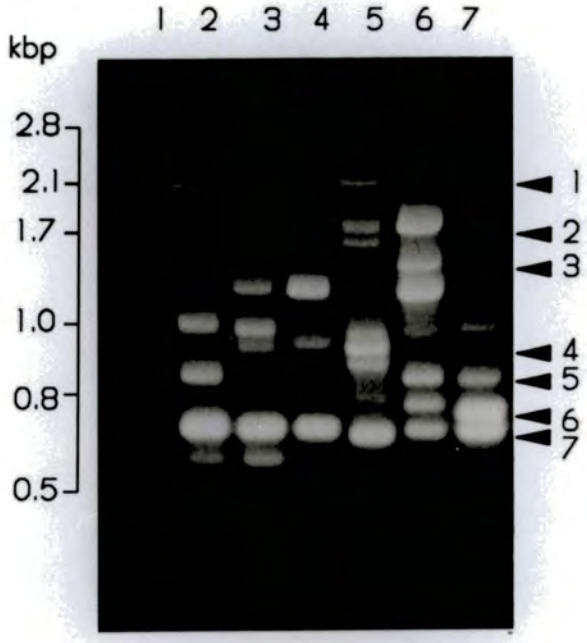
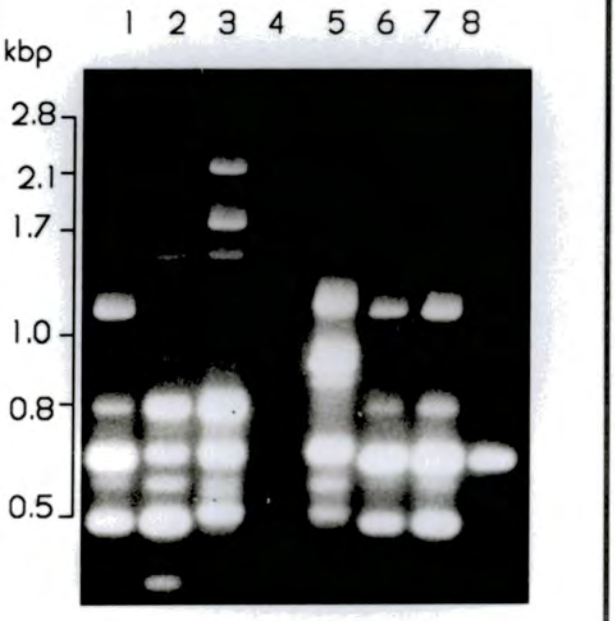


Figure 3.43



8 (*B. juncea*), with the exception of the major bands, the amplification products were too faint, probably as a result of low template DNA concentration due to pipetting error.

After analysing six phylogenetically related *Brassica sp.* in RAPD DNA polymorphism assay, it was observed that with some primers (eg. AP5h (Figure 3.41) and AP5a (Figure 3.43) common patterns can be generated so as to identify between closely related species while some primers (eg. GE1 (Figure 3.40) and AP12h (Figure 3.42) gave complex patterns that complicate the determination of the species' genetic relatedness. It was also found that the six *Brassica sp.* studied cannot be used to demonstrate the expected additive RAPD pattern as a result of a cross or hybrid as the amphidiploid *Brassica sp.* used in the analysis is not generated from a true hybrid of their parental (ancestral) species.

3.5.4 RAPD Analysis of *B. napus* Varieties

After successfully demonstrating the usefulness of RAPD analysis in revealing DNA polymorphism among several closely related *Brassica sp.*, a study was initiated to explore the feasibility of using this method to detect DNA polymorphism among *B. napus* varieties. From the experiments conducted using 11 arbitrary primers and 17 varieties of oilseed rape, it was found that with the exception of only two 9-mer oligonucleotide primers, all the primers were able to generate amplification products, some of which showed polymorphic bands. Some of the results representative of the RAPD analysis in oilseed rape varieties are presented in the following sections.

3.5.4.1 RAPD analysis using primer AP4c

Figure 3.44 shows the result of one experiment in which primer AP4c was used to amplify segments of genomic DNA from 17 varieties of oilseed rape. Lanes 1 to 9 contained template DNA from low glucosinolate varieties while Lanes 10 to 17 contained template DNA from high glucosinolate varieties. Lane 18 contained no template DNA and acted as a control reaction. As shown in Figure 3.44, extensive polymorphisms were observed. The amplification products generated ranged from about 150 bp to about 2.3 kb in size. The extension time of 2 minutes was therefore were capable of generating fragments up to 2.3 kbp. The majority

of the bands generated were of comparatively low intensities and were therefore classified as minor bands while the 3 prominent bands (Positions 6, 8 and 9) were classified as major bands. A closer analysis of the amplication products (bands), revealed extensive intervarietal polymorphisms where each polymorphic band could be scored by its presence or absence in a particular variety. Polymorphic bands are indicated by numbered arrows in Figure 3.44 where the numbers indicate the positions from the top of the gel. There are at least 7 scoreable polymorphic bands in the amplification products (Position 1, 2, 3, 4, 5, 7, and 8). The polymorphic band at Position 1 was unique to the varieties Lictor (5), Astec (10), Jet Neuf (13), and Rafal (16) and therefore could be used as a distinguishing RAPD marker for these four varieties. No common feature or genetic relatedness was present between these varieties. Although the majority of polymorphic bands occurred in minor bands (Positions 1, 2, 3, 4, 5, and 7) polymorphic bands were also observed in major bands. In the major band at Position 8, there were 4 varieties (Cobra (2), Drakkar (3), Libravo (4) and Astec (10)) that showed a different band pattern compared to the others. In these varieties, a less discrete band which appeared to be coming from two closely positioned bands was observed instead of only one as that present in other varieties at the same position. An excellent example of polymorphic band was shown by bands at Positions 2 and 3. The varieties can be classified into 3 band patterns namely (1) those with a single band at Position 2 (Lanes 1, 2, 5, 10, 11, 13, 14 and 15), (2) those with a single band at Position 3 (Lanes 7, 8 and 9), and (3) those with 2 bands (one at each Position 2 and 3), (Lanes 3, 4, 12, 15 and 16). It might therefore be suggested that these band patterns may be associated with the zygosity of the varieties where those varieties with one band are homozygous (parental type) while those with 2 bands are heterozygous (F_1 individuals). According to Williams *et al.* (1990), all RAPD markers are dominant where DNA segments of the same length are amplified from one individual but not from another, while codominant RAPD markers were detected only rarely and was observed as different-sized DNA segments amplified from the same locus. In a codominant RAPD marker, it is possible to distinguish between homozygous and heterozygous individuals whereby homozygotes contain only one amplified segment each of a different size and heterozygotes contain 2 amplified segments with each representing each of the DNA segments in homozygous individuals. On the other hand, it is impossible to distinguish whether a DNA segment is amplified from

one locus that is homozygous (i.e. contains 2 copies of the locus) or heterozygous (i.e. contains only 1 copy of the locus) in a dominant RAPD marker. In order to distinguish the zygosity of an individual in a dominant marker, Williams *et al.* (1990) suggested that tightly linked RAPD markers (i.e. to the dominant marker), each diagnostic for a different genotype of the parents, could be used in pairs to access the genotype of the individual. However, the degree of confidence with which heterozygotes could be identified would depend on how tightly linked the paired markers are (Williams *et al.* 1990). In Lane 6 and to a lesser extend Lane 2, the minor amplified products were very faint and therefore very difficult to analyse. As observed in several other reactions using different primers, only major bands were amplified.

In the absence of template DNA, that is, in the control reaction, amplification products were also observed (Lane 18, Figure 3.44). Consistent with the findings of Williams *et al.* (1990), these amplified products were not visible in the reactions containing the template genomic DNA. Although numerous polymorphic bands were observed in using AP4c as a primer for RAPD, no unique pattern of the amplified products associated with glucosinolate content was detected.

3.5.4.2 RAPD analysis using primer GE3

Another primer that was able to generate polymorphic amplification products is GE3 and the photograph of the stained gel is presented in Figure 3.45. In Figure 3.45, although minor bands (as indicated at Position 1, 2 and 3) appear to be missing in some varieties, the scoring of polymorphic bands at these positions was difficult due to the variation in the band intensities and thus rendered them unsuitable to be scored as polymorphic bands with a high degree of confidence. It was noted however that in some varieties, for examples varieties Ariana (1), Lirawell (6) and Brutor (12), all three bands appear to be missing. A more reliable scoring could be taken at Positions 4 and 5 where varieties in Lanes 6, 7, 8 and 10 did not contain the band at Position 4 while varieties in Lanes 2 and 3 did not contain the band at Position 5. As has been shown, primer GE3 was able to generate polymorphism in the major bands but not in the minor bands due to the difficulty in scoring the minor bands because of their variable intensities.

Figure 3.44 — Analysis RAPD amplification products generated by primer AP4c. Genomic DNA samples from seventeen varieties of *B. napus* were used as templates in the amplification reactions using primer AP4c. Low glucosinolate varieties ("L") used were Ariana (1), Cobra (2), Drakkar (3), Libravo (4), Lictor (5), Lirawell (6), Optima (7), Topas (8) and Bronowski (9) (Lanes indicated in brackets), and high glucosinolate varieties ("L") Astec (10), Bienvenu (11), Brutor (12), Jet Neuf (13), Mikado (14) , Pasha (15), Rafal (16) and Willi (17) (Lanes indicated in brackets). Control reaction i.e. no template DNA was in Lane 18. Amplification products were analysed in a 1.4%(w/v) agarose gel and detected by staining with EtBr. Polymorphic band patterns were observed in at least 7 positions as indicated by numbered arrows. Variety-specific bands were exemplified at Position 1 by varieties in Lanes 5, 10, and 13. The best polymorphic band patterns were found at Positions 2 and 3.

Figure 3.45 — Analysis of RAPD amplification products generated by primer GE3. Genomic DNA samples used and their lane assignments on the gel (variety Rafal was not done) are as described in Figure 3.44. L and H denote low and high glucosinolate varieties respectively. Control reaction is in Lane 17. Although irregular band intensities interfere with the analysis of minor bands in the amplification products, the polymorphism in the major band positions are easily scored as indicated by the numbered arrows.

Figure 3.44

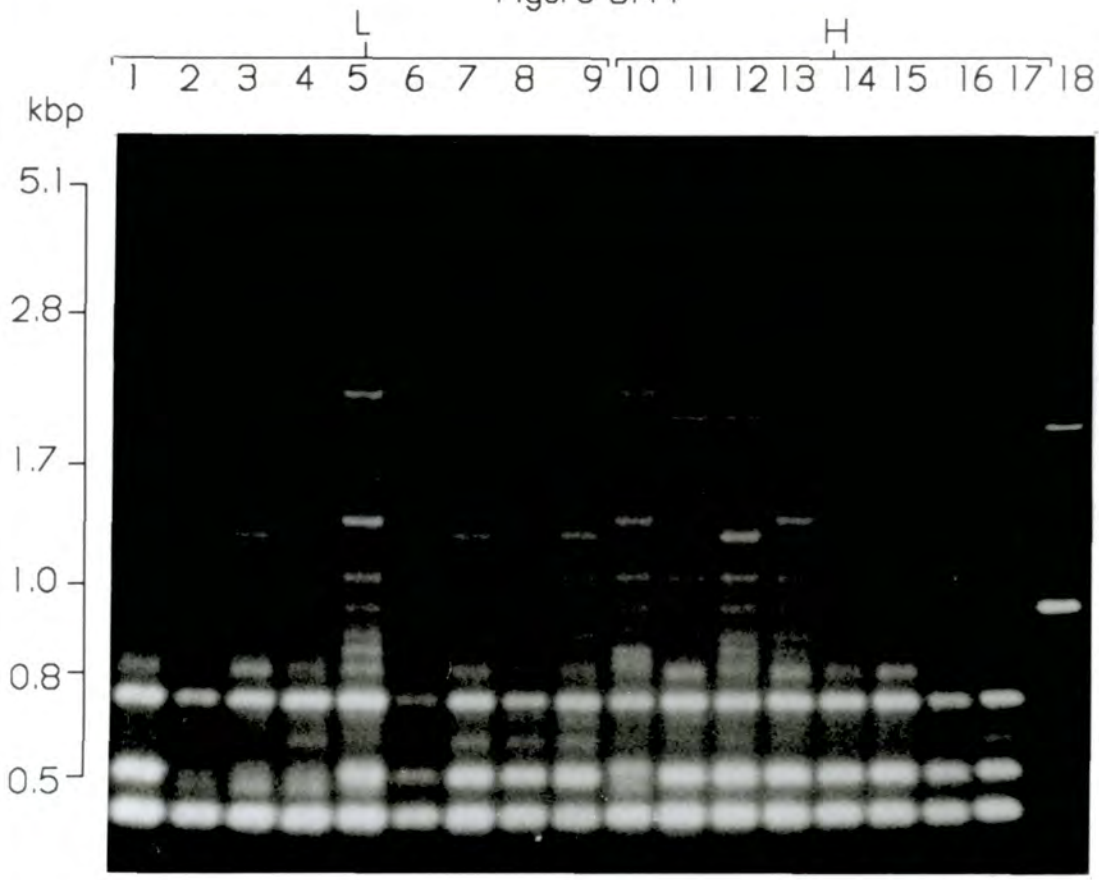
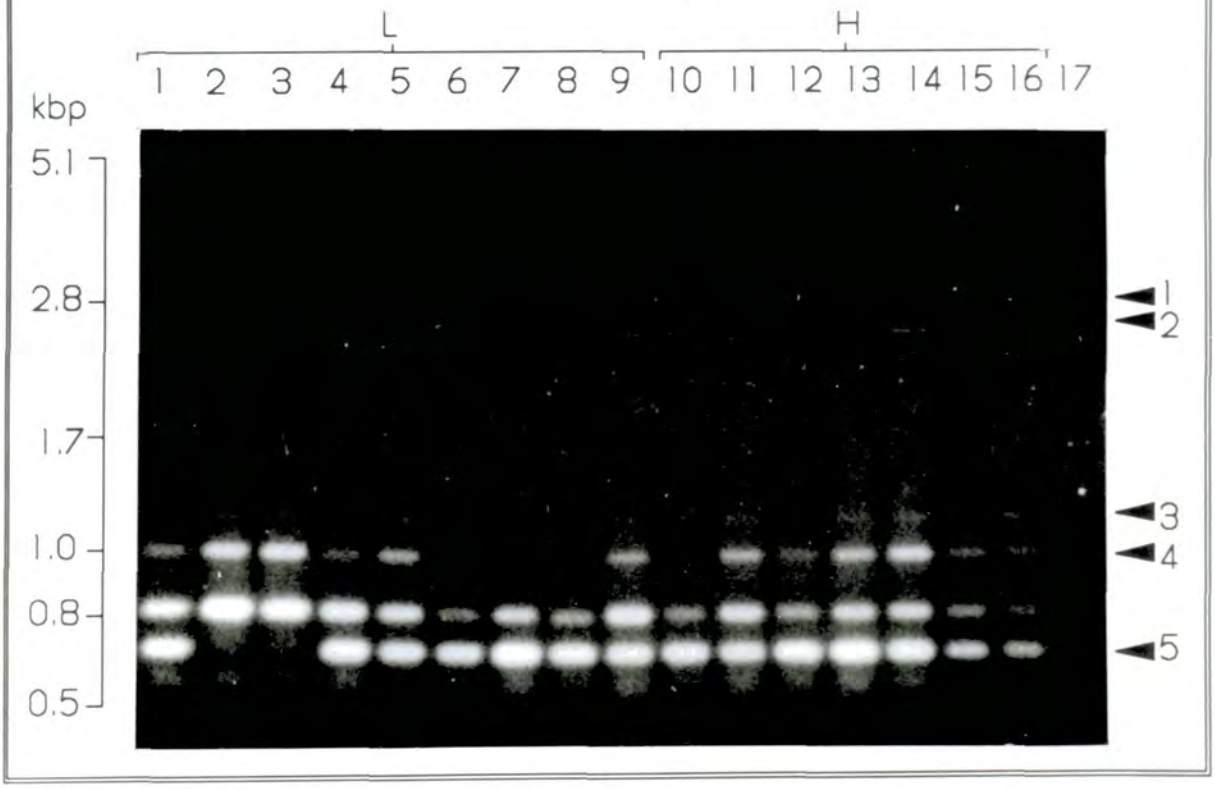


Figure 3.45



3.5.4.3 RAPD analysis using primer AP5a

A similar situation to primer GE3 in relation to the variable band intensities at various positions were observed when primer AP5a was used in the reaction as shown in Figure 3.46. Bands at Positions 1 and 2 showed varying intensities and therefore could not be scored as polymorphic bands. At Positions 3 and 4, although the phenomenon of variable intensities similar to that observed at Position 1 and 2 was present, there was a strong evidence that the band at these positions (Position 3 and 4) was missing in some varieties. No band was observed at Position 3 in varieties Drakkar (3), Topas (8), and Willi (17) and at Position 4 in varieties Cobra (2), Libravo (4), and Jet Neuf (13). The bands at Positions 1 and 2, with the exception of the varieties Ariana (1) and Brutor (12) in which there were amplification reaction failures, were present in all the varieties including variety Cobra (2) where the bands were extremely faint. The effect of variable intensities on the scoring of polymorphic bands will be discussed later in the concluding remarks. It is interesting to note that in the two varieties shown in Lanes 3 and 8, a band was present at Position 4 but not at Position 3 while the variety, for example, in Lane 16 appeared to contain a band at both positions and the variety in Lane 17 did not appear to have any bands at all. These type of RAPD pattern were useful for varietal identifications. The amplification reactions for variety Ariana (1) and Jet Neuf (13) appeared to have failed. In variety Ariana (1), only one major was observed at low intensity (i.e. low amount of amplified DNA segments) and no minor band. It can be suggested that the failure to detect other bands in the amplification reaction could be as a result of insufficient amount of template DNA in the reaction mixture. Low or insufficient template DNA concentration might contribute to the low amount of amplification products, which were thus visually undetectable in the stained gel; as was proven in earlier experiments on the effect of template DNA concentration on the generation of amplification products. Contrastingly, the failure of the amplification reaction in variety Jet Neuf (13) to give discrete bands could be attributed to excess template DNA concentration as suggested by Williams *et al.*, (1990) who stated that a smear instead of discrete bands could be resulted from the presence of high concentration of template DNA in the amplification reaction.

The phenomenon of variable intensities observed in the amplification products generated by primer AP5a has contributed to the inability of reliable scoring of polymorphic positions in some rape varieties. The possible underlying reason for the variable band intensities (qualitatitative versus quantitative scoring strategies) will be discussed later in the concluding remarks.

3.5.4.4 RAPD analysis using primer AP12h

Another primer which showed extensive polymorphism in RAPD is AP12h (Figure 3.47). At least 6 scoreable polymorphic bands were observed in the amplification products as shown in Figure 3.47. Among the varieties which are scoreable at Positions 2 and 3, the varieties in Lanes 14 and 15, contained only a band at Position 2, the varieties in Lanes 5, 7 and 13 contained a band at Position 3 while those in Lanes 10, 11 and 16 contain both bands at these positions. Other positions which showed polymorphic bands were Positions 4, 5, 7 and 8. The most interesting polymorphism observed was at Position 4 where the band at this position indicated a possible relationship to varieties with low glucosinolate content. Lanes 2 to 10 contained RAPD products from genomic DNA of low glucosinolate varieties where the band showed most prominently in Lanes 7, 8, 9 and 10 and to a lesser extend, Lanes 3, 4 and 5. The major band at this position was observed to be almost exclusively related to low glucosinolate rape varieties. Even at its lowest overall intensity (where other bands were observed to be of lower intensities), the band at Position 4 could still be seen (Lanes 3, 4 and 5). A very faint band was observed at the same position in Lanes 15, 16 and 17 which represent the high glucosinolate varieties. The band at Position 4 in these varieties were of comparatively very low intensity when compared to that observed in the low glucosinolate varieties. This band is sufficiently interesting to warrant further analysis in order to confirm this observation and to prove relatedness to glucosinolate content of varieties. This primer was also shown to amplify variety-specific band patterns. As examples, a unique band at Position 1 belonged to variety Optima (8) while another unique band at Position 6 belonged to the variety Jet Neuf.

3.5.4.5 RAPD analysis using primers AP6 and GE2

Not all the primers used gave highly polymorphic amplification products.

Figure 3.46 — Analysis of RAPD amplification products generated by primer AP5a. Template DNAs from low glucosinolate varieties (L) – Ariana (1), Cobra (2), Drakkar (3), Libravo (4), Lictor (5), Lirawell (6), Optima (7), Topas (8), Bronowski (9), and high glucosinolate varieties (H) – Astec (10), Bienvenu (11), Brutor (12), Jet Neuf (13), Mikado (14), Pasha (15), Rafal (16) , and Willi (17) (Lanes numbers indicated in brackets) were used in RAPD analysis using primer AP5a. Control reaction is in Lane 18. Difficulties in analysing the polymorphic band patterns of amplification products was apparent due to variable band intensities as shown at Positions 1 and 2 (indicated by numbered arrows). Scoreable polymorphic bands are observed at Positions 3, 4 and 5. Amplification reactions appeared to have failed in variety Ariana (1) and Jet Neuf (13).

Figure 3.47 — Analysis of RAPD amplification products generated by primer AP12h. Lane 1 is the control reaction. Low glucosinolate varieties (L) are shown in Lanes 2 to 10 while high glucosinolate varieties (L) are shown in Lanes 11 to 18. The varieties are exactly as described in Figure 3.46 Control reaction is in Lane 1. Polymorphic bands at Position 4 appeared to be specific to low glucosinolate varieties where this band is present in Lanes 3, 4, 5, 7 to 10. Unique bands in Lane 14 (Position 6) and Lane 8 (Position 1) are unique to varieti es Jet Neuf and Optima respectively. Amplification reactions for varieties Ariana (2) and Lirawell (6) appeared to have failed in generating other bands as observed in other varieties.

Figure 3.46

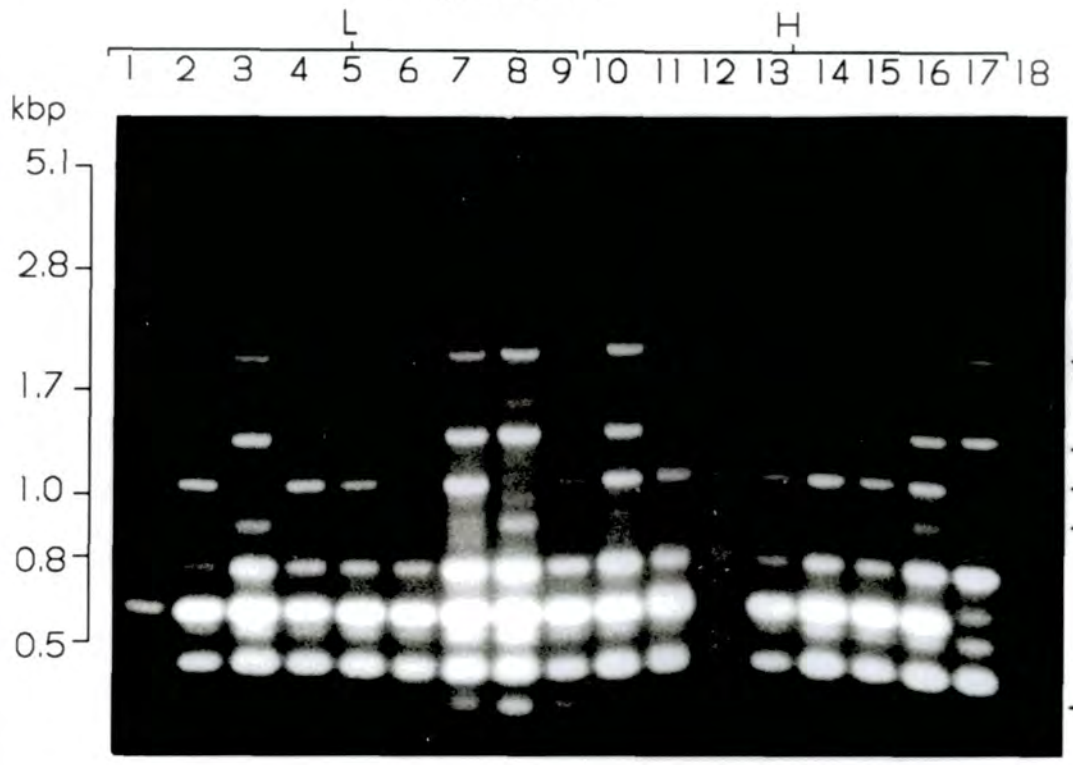
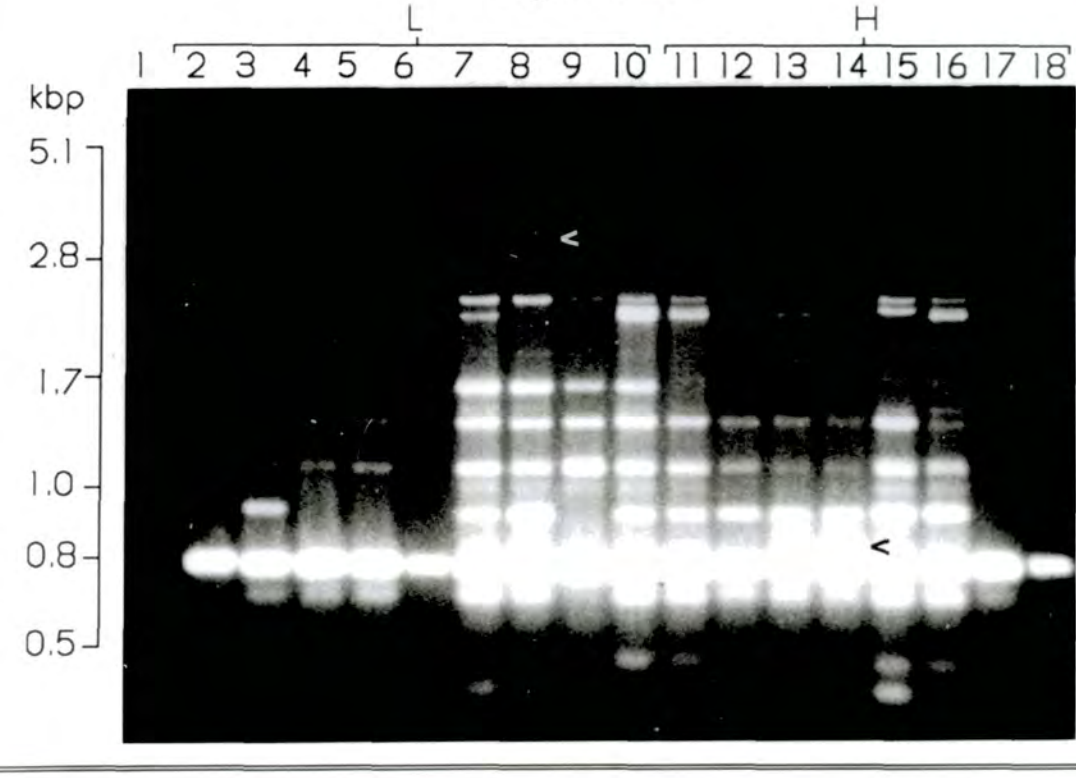


Figure 3.47



Primer AP6 and GE2 gave band patterns which were generally conserved amongst the rape varieties tested as shown in Figures 3.48 and 3.49 respectively. With the exception of 2 varieties, all the varieties gave identical band patterns with primer AP6. The two exceptions, Astec (8) and Bronowski (17) seemed to have lost their band at Position 3. Closer analysis of the amplification products generated by primer AP6 revealed that the two differing varieties have, instead of the band at Position 3, two new bands at Position 1 and 2 (Lane 17) and one new band at Position 4 (Lane 8). It was interesting to note that in variety Astec (8), a new band of lower MW than that lost at Position 3 appeared at Position 4 and in variety Bronowski (17), two new bands appeared at Positions 1 and 2 which were of higher MW than that lost at Position 3. These separate observations could possibly be explained by deletion/insertion of DNA segments (translocation) between two priming sites for the band at Position 3. In variety Astec (8), DNA segments between two priming sites for the band at Position 3, might have been translocated elsewhere in the genome. Consequently, the two priming sites were closer to each other and thus a lower MW band was amplified. In variety Bronowski (17), however, DNA segments each of different lengths might have been inserted between the two priming sites and consequently gave rise to two higher MW amplification products.

A similar phenomenon of conserved band patterns was also observed, to a lesser extent, when primer GE2 was used in the study of RAPD. As shown in Figure 3.49, all varieties gave a similar band pattern; probably with the exception of band at Position 1 which appeared to be polymorphic. The band at Position 1 was absent in varieties Ariana (1), Cobra (2), Libravo (3), Bienvenu (11), Mikado (14) and Willi (17). These varieties were scored at Position 1 after taking into consideration the overall intensities of other bands especially at Positions 2 and 3 which must be of high intensities. As with some of the results of other primers used earlier, the polymorphic position in this case (at Position 1) was not able to be scored with high degree of confidence due to irregular band intensities. Similar observation of highly variable band intensities was also present on bands at Positions 2 and 3. At a first glance, the bands at Position 2 and 3 could be considered as polymorphic; but upon closer analysis, the bands were present in all the varieties with successful amplification reactions, though their intensities were

Figure 3.48 — Analysis of RAPD amplification products generated by primer AP6. The varieties used as template genomic DNA samples were high glucosinolate varieties (H): Bienvenu (1), Brutor (2), Jet Neuf (3), Mikado (4), Pasha (5), Rafal (6), Willi (7) and Astec (8), and low gulcosinolate varieties (L): Ariana (9), Cobra (10), Drakkar (11), Libravo (12), Lictor (13), Lirawell (14), Optima (15), Topas (16), and Bronowski (17) (Lanes indicated in brackets). Control reaction (no template DNA) was in Lane 18. With the exception of the 2 varieties Astec (8) and Bronowski (17), all other varieties produced amplification products with identical band patterns. Amplification reaction of variety Mikado (4) appeared to have failed.

Figure 3.49 — Analysis of RAPD amplification products generated by primer GE2. The varieties used as genomic template DNAs were low glucosinolate varieties (L): Ariana (1), Cobra (2), Drakkar (3), Libravo (4), Lictor (5), Lirawell (6), Optima (7), Topas (8), Bronowski (9); and high glucosinolate varieties (H): Astec (10), Bienvenu (11), Brutor (12), Jet Neuf (13), Mikado (14), Pasha (15), Rafal (16), and Willi (17). Polymorphic band observed at Position 1. At Position 2 and 3, although showing highly variable band intensities were present in all the varieties with successful amplification reactions. The band pattern was relatively conserved in all the varieties. Amplification reaction appeared to have failed completely for varieties Drakkar (3) and Pasha (15). Only one band appeared in variety Rafal (16) indicating also amplification failure.

Figure 3.48

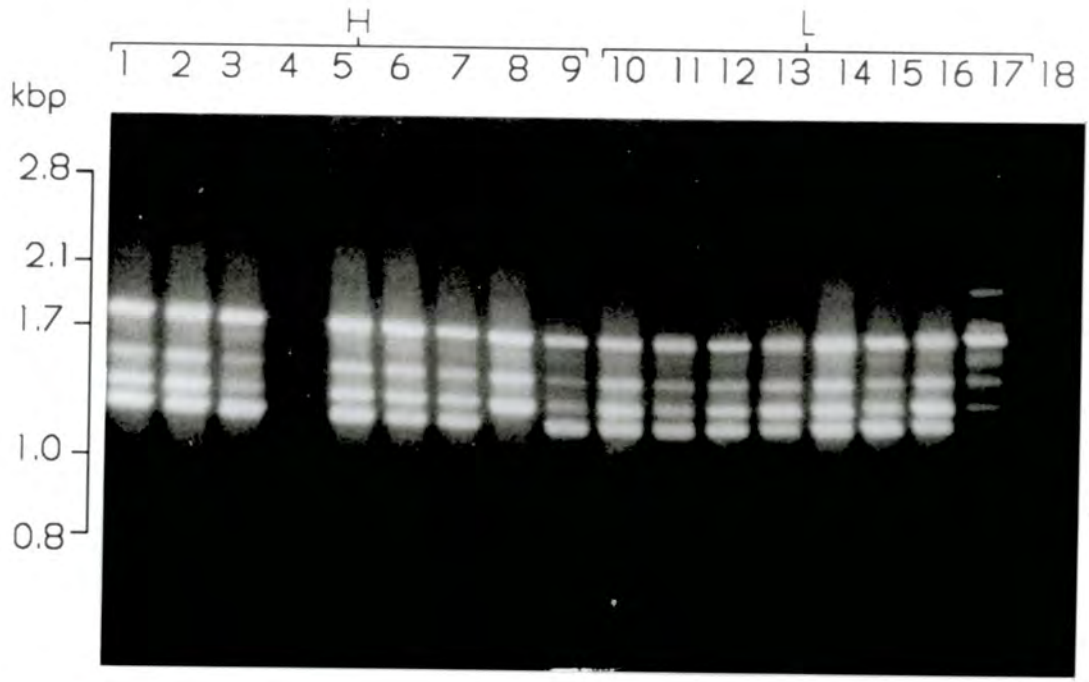
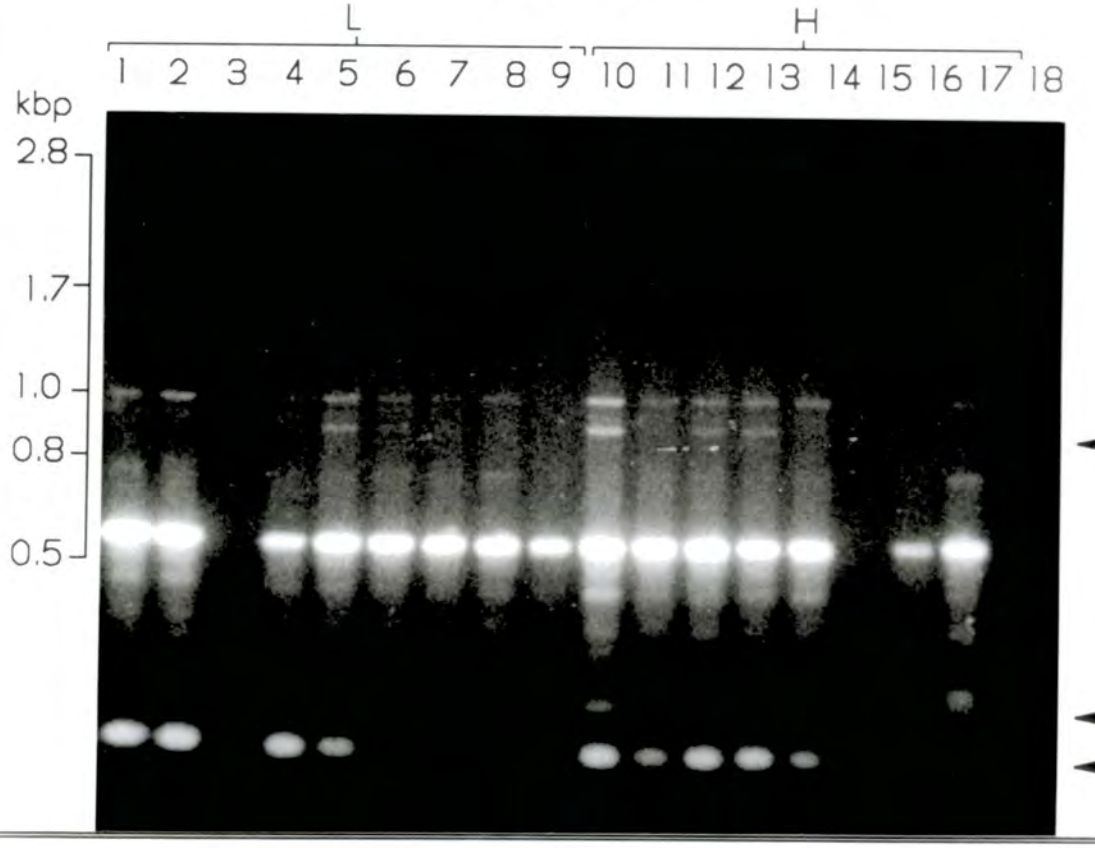


Figure 3.49



extremely variable. The scoring of RAPD markers based on qualitative analysis will be discussed in the concluding remarks.

3.5.5 Concluding Remarks

Random Amplified Polymorphic DNA (RAPD) as a mean of detecting intra- and interspecies DNA polymorphisms has been explored. From the study, it has been shown that, with suitable primers, DNA polymorphism could be detected when 6 phylogenetically related *Brassica sp.* were used where it was possible to determine the genetic relatedness between some of the amphidiploids (eg. *B. napus*) and their respective ancestors *B. campestris* and *B. oleraceae.* It was also found that RAPD analysis was able to reveal DNA polymorphisms among the varieties of oilseed rape. While some primers produce highly conserved banding patterns, other primers could produce polymorphic and even hypervariable banding patterns as has been shown when analysing six *Brassica sp.*.

Although a very limited number of primers were analysed in the study, it has been shown that the nucleotide sequence of each primer played a major role in determining the success of RAPD analysis. Beckmann (1988) suggested, and later confirmed by Williams *et al.* (1990), that primers should contain at least 9 or 10 nucleotide in length, between 50 to 80% G+C in composition, and no palindromic sequence. The primers used in this study met these criteria as their sequences were those described by Williams *et al.* (1990). In this study, it was found that only one of the three 9-nucleotide long (9-mer) primers was able to produce amplification products under the chosen condition of 36°C of annealing temperature suggested by Williams *et al.* (1990). Therefore, the possibility that the 9-mer primers could not support the amplification of DNA segments due to inefficient binding of the primers to the segment of the DNA at that annealing temperature could not be excluded.

Besides the choice of primers, the assay condition of the amplification reactions could also affect RAPD analysis. The concentration of template genomic DNAs and the amount of *Taq* polymerase used in the amplification reaction could greatly affect the performance of RAPD analysis. In this study, the concentration of template DNA was 25 ng in 25 μl reaction volume, the amount of *Taq* polymerase was 0.5U/assay, exactly as described by Williams *et al.* (1990). In earlier

experiments, 1U/assay in 50μl volume was used. Reducing the assay volume from 50 to 25 μl and the amount of *Taq* polymerase from 1U to 0.5 U/assay did not affect the efficiency of generating amplification products. The use of lesser amount of *Taq* polymerase therefore reduces the cost of the assay.

The effect of template DNA concentration on the generation of amplification products was also studied. In some of the amplification reactions carried out, minor amplification products were comparatively very faint and made the analysis an arduous task. An example of this phenomenon was shown in Figure 3.49, where the minor bands in Lanes 6, 7, 8, 9 and 16 were comparatively fainter than those in Lanes 1, 2, 10, 11 and 12. It is very difficult to explain the cause of this phenomenon. It is possible that the primers were somewhat sterically hindered and thus partially prevented them from annealing to the template DNA. Consequently, less amplification products of each DNA band were produced. It could also be due to the lower template DNA present in the reaction mixture. The result of an experiment carried out earlier showed that by lowering the concentration of the template DNA from the normal 25 ng to 5 ng (Data in Section 3.5.3), the amount of the amplification products decreased significantly and therefore were difficult to analyse visually. On the other hand, when the template concentration was increased to 50 ng (i.e. doubled), the amount of the amplification products proportionately increased (Data in Section 3.5.3). It can therefore be concluded that the template concentration can significantly affect the generation of amplification products in RAPD assay. The optimum concentration of template DNAs should be firstly established for every species before analysing using RAPD assay so as to facilitate the analysis of the polymorphic banding patterns. Williams *et al.* (1990) however, argued that high template DNA concentration in an amplification reaction will produce a "smear" instead of discrete bands of the amplification products. Furthermore, Welsh and McClelland (1990) contended that the amplification products produced at optimum template concentration were all represented at higher template concentrations but not at lower template concentrations where the PCR was more sporadic. They further added that template concentration of 10 pg is the lower limit of producing a reliable RAPD for the genome size of approximately 3 x 10^6 bp. For *Brassica napus*, with genome size of approximately 1.6 x 10^9 bp (I.M. Evans, pers. comm.), it was found that the lower limit of the template DNA concentration was 15 ng (Data not shown).

Besides its many advantages over other methods of DNA polymorphism assay (eg. RFLP), the main drawback of PCR-based analysis is the contamination of extraneous DNA. This is of paramount importance especially in RAPD analysis where DNA cross-contamination can be disastrous to the final analysis of the polymorphic banding patterns. Several steps has been taken in this study to prevent DNA cross-contamination. One of them is the preparation of a master mix where the components in the reaction were added in appropriate volumes to a single tube and aliquoted out to several tubes as desired. The reaction mixtures were then overlaid with equal volume of sterile mineral oil. Genomic DNAs were sampled from working stocks and singly added to each tube by carefully pipetting the DNA solution under the mineral oil overlay. Fresh pipette tips were used at every pipetting step. The preparation of a mastermix minimised the pipetting steps involved in subsequent manipulation and hence minimised the contamination. Steps were also taken to ensure that only one tube was open at any one time and should be immediately closed after each pipetting step. Finally, a control reaction where no template DNA was added to the reaction mixture was always included in every amplification run for monitoring cross-contamination and non-specific amplification.

Another disadvantage is the prevalence of irregular band intensities in some of the amplification reactions. Irregular band intensities complicate the scoring of polymorphic bands and by extrapolation, RAPD markers. This problem, however, can probably be solved by determining the optimum template DNA concentration that can give uniform band intensities. Moreover, extremely accurate pipetting is necesary as a low concentration of template DNA is used and therefore the margin of error is very small. This was observed in later RAPD analyses of the varietal differences where although the optimum template DNA concentration was used, as established in earlier experiments involving *Brassica* species, some primers still showed some degree of variable intensities. A higher degree of confidence in scoring polymorphic band of irregular intensities can be achieved by comparing the overall intensities of other bands in the same amplification reaction.

As has been shown in this study, intra- and interspecific genetic variability assessment can be attributed to RAPD. Due to its simplicity, rapidity, and ability to complement the already-established RFLP technology, various applications

of RAPD can therefore be suggested which normally attributed only to RFLP: (1) identification of crop cultivars, (2) assessment of the "purity of inbred lines, (3) selection of the recurrent parental genotype in backcross breeding programmes, (4) characterisation of fusion hybrids, eg. by somatic hybridisation, and (5) evaluation of the extent of somaclonal variation at the molecular level.

CHAPTER IV

CONSPECTUS AND SUGGESTED FURTHER RESEARCH

RFLP and RAPD technology can play a major role in helping plant breeders with the selection of desirable characters in oilseed rape breeding programmes. As has been shown in this study, it is possible to fully utilise these methods in realising DNA polymorphisms among commercial rape varieties using suitable enzyme-probe combinations in the case of RFLP and particularly with random oligonucleotide primers in RAPD.

One of the aims of this research programme was to survey the prevalence of RFLPs in commercial oilseed rape varieties. In general, the study has provided evidence for the occurrence of RFLPs associated with extensin and self-incompatibility related genes. It has also provided evidence on its applicability in analysing segregating populations resulting from sexual crossing experiments, though it was not possible to confirm Mendelian inheritance due to an insufficient number of available DNA samples from segregating F_2 individuals. This was a limitation due to insufficient time to carry out the labour-intensive DNA extractions of leaf materials from more F_2 individuals. Moreover, upon analysis of the RFLP patterns generated, no linkage to the glucosinolate content of the varieties was observed.

In the second part of this research programme, differential screening of a pod/seed cDNA library failed to identify any glucosinolate-related sequences that could be used as RFLP probes. As has been mentioned earlier, one way of increasing the chances of isolating the desired clones would be to differentially screen a much larger cDNA library of about 20, 000 clones as suggested by Sambrook (1989). Alternatively, differentially-expressed genes could be enriched by substractive hybridisation (substractive cDNA library). Principally, mRNA isolated from pod tissue of a high glucosinolate variety is used as a template to produce

a radiolabelled tracer cDNA. The tracer cDNA is hybridised with a sequence excess of driver mRNA isolated from low glucosinolate pod tissue. By passing the mixture through a hydroxyapatite column, where double-stranded nucleic acids (cDNA:mRNA) bind more tightly than single-stranded cDNA, the unhybridised tracer cDNA will contain an enriched population of sequences expressed in high glucosinolate pod tissue but not in low glucosinolate. Double-stranded cDNA synthesised from this enriched population of single-stranded cDNA can then be cloned and screened for possible glucosinolate-related sequences. A more sophisticated variant of this strategy has been devised which involve the immobilisation of the low glucosinolate cDNA on magnetic beads (via biotin labelling and reaction with streptavidin beads) for the purpose of "fishing" out identical sequences from a high glucosinolate cDNA, leaving cDNA sequences unique to the high glucosinolate variety free in solution for use as a probe or for direct cloning.

One of the disadvantages of the low glucosinolate varieties of oilseed rape available commercially today is the reduction in the level of glucosinolates present in the vegetative organs of rape plants which renders them susceptible to fungal attacks and other diseases and also cause harm to grazing animals which feed on the leaves and other vegetative organs by producing excessive wind in the rumen causing them to damage (D. Murphy, pers. comm.). Consequently, the production of rape varieties with differential expression of glucosinolates i.e. low in the seeds while maintaining a high level in other vegetative organs may be essential. One way of achieving this is would be by introducing tissue-specific antisense gene sequences via plant transformation, targetted at an mRNA responsible for controlling the biosynthesis of glucosinolates and thus inhibiting the synthesis in, for example, seeds. Therefore, a thorough knowledge of the pathway of glucosinolate biosynthesis is essential before this can be achieved.

The third objective of this research programme was the study of the sequence and organisation of a class of highly repetitive sequences in *B. napus*. This part of the research programme was highly successful. The cloning of one of each member of the monomeric, trimeric, and tetrameric forms of the repetitive sequences and the complete analysis of their sequences has provided an insight into its organisation and possible evolution in the rape genome. Further analysis

should be directed towards its function in the genome. Localisation of the repetitive sequences, for example during mitosis, using the cloned sequence for *in situ* hybridisations may provide an insight as to its function (if any) in cell division and differentiation.

Lastly, this research programme has provided good evidence for the usefulness of RAPD technology in studying DNA polymorphisms among *Brassica* species and also among varieties of oilseed rape. Even with only a limited number of primers, RAPD analysis has shown its potential by revealing DNA polymorphisms in oilseed rape varieties using a simple, rapid and reproducible method. RAPD analysis in oilseed rape should be further developed by optimising the reaction conditions, and also possibly employing computer automation in its procedure. Welsh and McClelland (1990) used ^{32}P-labelled dCTP in such amplification reactions, followed by analysis on a polyacrylamide gel and autoradiography. The use of radiolabelled dCTP in the reaction and separation on a polyacrylamide gel should facilitate and improve the final analysis of the polymorphic bands. The use of fluorescently labelled bases or primers allows the full automation of the analyses of RAPD amplification products on instruments such as the Applied Bioscience's ABI 373A DNA sequencer.

When one primer (AP12h) was used in the RAPD analysis, a single band was amplified from DNA samples of low glucosinolate varieties which might be related to low glucosinolate content. Although inconclusive, since a faint band of identical size was also amplified in some of the DNA samples from the high glucosinolate varieties, the possible link between this band and the low glucosinolate varieties is a possibility and deserves further exploration. Two approaches are therefore suggested in the study. Firstly, the amplified band can be isolated, reamplified and cloned in a suitable vector (pUC18-*Sma* I) or TA cloning vector (Invitrogen). The insert can then be used as probe to analyse genomic southern blots containing DNAs from high and low glucosinolate varieties and also from well characterised breeding materials from crosses between high and low varieties. Secondly, the insert can be used as a probe for Northern blots from different tissues of high and low glucosinolate varieties to confirm its specific expression. Depending on the result of the Northern blots, the insert can then be used as a probe to isolate cDNA

clones from cDNA library which might be glucosinolate-related.

The potential for this technique is further illustrated by the availability of 6 sets of 100 primers each (i.e. a total of 600) from S.D. Tanksley, Dept. of Biochemistry, Cornell University, Ithaca, NY., USA. Researchers can therefore obtain these available primers, use them in the analyses, identify those that are suitable and synthesize a larger amount of these particular primers for further analyses; thus there is no need for the initial synthesis of random, untested primers in individual laboratories.

CHAPTER V

REFERENCES

1. Adams, R.L.P, R.H. Burdon (1985). Molecular biology of methylation. Springer Verlag, New York.

2. Akbar, M.A. (1989). Resynthesis of *Brassica napus* aiming for improved earliness and carried out by different approaches. *Hereditas* 111(3):239-246.

3. Appels, R., E.S. Dennis, D.R. Smyth, and W.J. Peacock (1981). Two repeated DNA sequences from the heterochromatic region of rye (*Secale cereale*) chromosomes. *Chromosoma (Berl.)* 84:265-277.

4. Apuya, N.B., B.L. Frazier, P. Keim, E. Jill Roth, and K.G. Lark (1988). Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. *Theor Appl Genet* 75:889-901.

5. Austin, R.B., R.B. Flavell, I.E. Henson and H.J.B. Lowe (1986). Oilseed rape. In: *Molecular Biology of Crop Improvement–a case study of wheat, oilseed rape and faba beans.* Austin, R.B. and B. Roger, eds.. Cambridge University Press, Cambridge.

6. Barker, D., M Schafer, and R. White (1984). Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell* 36:131-138.

7. Barnes, S.R., A.M. James, and G. Jamieson (1985). The organisation, nucleotide sequence, and chromosomal distribution of a satellite DNA from *Allium cepa. Chromosoma (Berl.)* 92:185-192.

8. Baum, B.R. (1978). Taxonomy of the tribe *Triticeae* (*Poaceae*) using various numerical taxonomic techniques. III. Synoptic key to genera and synopses. *Can J Bot* 56:374-385.

9. Beckman, J.S., and M. Soller (1990). Towards a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. *Bio/Technology* 8:930-932.

10. Beckman, J.S., and M. Soller (1983). Restriction fragment length polymorphism in genetic improvement: methodologies, mapping and costs. *Theor Appl Genet* 67:35-43.

11. Bennett, M.D. (1972). Nuclear DNA amounts and minimum generation time in herbaceous plants. *Proc Roy Soc ser. B* 191:109-135.

12. Benslimane, A.A., M. Dron, C. Hartmann and A. Rode (1986). Small tandemly repeated DNA sequences of higher plants likely to originate from a t-RNA gene ancestor (1986). *Nucleid Acids Res* 14(20):8111-8119.

13. Beridze, T.G. (1986). *Satellite DNA.* Springer Verlag, Berlin.

14. Bernatzky, R., and S.D. Tanksley (1986a). Towards a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* 112:887-898.

15. Bernatzky, R., and S.D. Tanksley (1986b). Majority of random cDNA clones correspond to single loci in the tomato genome. *Mol Gen Genet* 203:8-14.

16. Bethesda Research Laboratories (1986). BRL pUC host: *E. coli* DH5α^{TM} competent cells. *Bethesda Res. Lab. Focus* 8(2):9.

17. Bhatty, R.S., S.L. McKenzie, and A.J. Finlayson (1968). The proteins of rapeseed (*Brassica napus* L.) soluble in salt solutions. *Can J Biochem* 46:1191-1197.

18. Bolivar, F. , R.C. Rodriguez, P.L. Greene, M.C. Betlach, H.C. Heyneker, H.W. Boyer, J.H. Crosa and S. Falkow (1977). Construction and characterisation of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2:95.

19. Bonierbale, M., R. Plaisted, and S.D. Tanksley (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* 120:1095-1103.

20. Botstein, D., R. White, M. Skolnick,and R.W. Davis (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331.

21. Boulter, M.E., E. Croy, P. Simpson, R. Shield, R.R.D. Croy, and A.H. Shirsat (1990). Transformation of *Brassica napus* L. (oilseed rape) using *Agrobacterium tumiefaciens* and *Agrobacterium rhizogenes*–a comparison. *Pl Sci* 70:91-99.

22. Bowland, J.P., D.R. Clandinin, and L.R. Wetter (1965). Rapeseed meal for livestock and poultry – a review. Canadian Department of Agriculture, publication 1257.

23. Brennicke, A., V. Hemleben (1983). Sequence analysis of the cloned *Cucumis melo* highly repetitive satellite DNA. *Z. Naturforsch Teil C* 38:1062-1065.

24. Bunting, F.S. (1986). Oilseed rape in perspective. In: *Oilseed Rape*, Scarisbrick, D.H., and R.W. Daniels, eds. Collin, London.

25. Burr, B., R.A. Burr, K.H. Thompson, M.C. Albertsen, and C.W. Stuber (1988). Gene mapping with recombinant inbreds in maize. *Genetics* 118(3):519-526.

26. Burr, B., E. Evola, R.A. Burr, and J.S. Beckman (1983). The application of restriction fragment length polymorphism to plant breeding. In: *Genetic Engineering: Principles and Methods*, Setlow, J.K., and A. Hollaender eds. Plenum Press, New York and London, pp 45-59.

27. Capesius, I. (1983). Sequence of the cryptic satellite DNA from the plant *Sinapis alba*. *Biochim Biophys Acta* 739:276-280.

28. Cameron, J.R., E.Y. Loh, and R.W. Davis (1978). Evidence for transposition of dispersed repetitive DNA families in yeasts. *Cell* 15:687-701.

29. Chang, C., J.L. Bowman, A.W. DeJohn, E.S. Lander, and E.M. Meyerowitz (1988). Restriction fragment length polymorphism linkage map for *Arabidopis thaliana*. *Proc Natl Acad Sci (USA)* 85:6856-6860.

30. Chen, B.Y., W.K. Heneen, and R. Johsson (1988). Resynthesis of *Brassica napus* L. through interspecific hybridisation between *B. alboglabra* Bailey and *B. campestris* L. with special emphasis on seed colour. *Pl Breed* 101:52-59.

31. Chomczynski, P. and N. Sacchi (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156-159.

32. Chung, C.T., S.L. Niemela and R.H. Miller (1989). One-step preparation of competent cells *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci (USA)* 86:2172-2175.

33. Close, P.S., R.C. Shoemaker, and P. Keim (1989). Distribution of restriction site polymorphism within the chloroplast genome of the genus *Glycine*, subgenus *Soja*. *Theor Appl Genet* 77:768-776.

34. Crouch, M.L., K.M. Tenbarge, A.E. Simon, and R. Ferl (1983). cDNA clones for *Brassica napus* seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. *J Mol Appl Genet* 2:273-283.

35. Crouch, M.L., and I.M. Sussex (1981). Development and storage-protein synthesis in *Brassica napus* L. embryos *in vivo* and *in vitro*. *Planta* 153:64-74.

36. Domoney, C., T.H.N. Ellis, and D.R. Davies (1986). Organisation and mapping of legumin genes in *Pisum*. *Mol Gen Genet* 202:280-285.

37. DeLisle, A.J., and M.L. Crouch (1989). Seed storage protein transcription and mRNA levels in *Brassica napus* L. during development and in responce to exogenous abscisic acid. *Plant Physiol* 91:617-623.

38. Delseny, M., M. Laroche and P. Penon (1986). Methylation pattern of radish (*Raphanus sativus*) nuclear ribosomal RNA genes. *Plant Physiol* 76:627-632.

39. Deumling, B. (1981). Sequence arrangement of a highly methylated satellite DNA of a plant *Scilla*: A tandemly repeated inverted repeat. *Proc Natl Acad Sci (USA)* 78:338-432.2

40. Dvorak, J., P.E. McGuire, and B. Cassidy (1988). Apparent source of the A genomes of wheat inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. *Genome* 30:680-689.

41. Ellis, T.H.N., and W. Cleary (1988). Trisomy : a useful adjunct to RFLP mapping in pea. *Chromosoma (Berl.)* 96:91-94.

42. Ellis, T.H.N., D.R. Davies, J.A. Castleton, and I.D. Bedford (1984). The organisation and genetics of rDNA length variants in peas. *Chromosoma (Berl.)* 91:74-81.

43. Evans, I.M., L.N. Gatehouse, J.A. Gatehouse, J.N. Yarwood, D. Boulter, and R.R.D. Croy (1990). The extensin gene family in oilseed rape (*Brassica napus* L.)–characterisation of sequences of representative members of the family. *Mol Gen Genet* 223(2):273-287.

44. Ericson, M.L., J. Rodin, M. Lenman, K. Glimelius, L.G. Josefsson, and L. Rask (1986). Structure of the rapeseed 1.7S storage protein, napin, and its precursor. *J Biol Chem* 261(31):14576-14581.

45. Feinberg, A.P., and E. Vogelstein (1984). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 137:266-267.

46. Flavell, R.B. (1982). Sequence amplification, deletion and rearrangement: major sources of variation during species divergence. In: *Genome Evolution*, Dover, G.A. and R.B. Flavell, eds. Academic Press, London.

47. Flavell, R.B. (1980). The molecular characterisation and organisation of plant chromosomal DNA sequences. *Annu Rev Plant Physiol* 31:569-596.

48. Gatehouse, L.N., I.M. Evans, J.A. Gatehouse, and R.R.D. Croy (1990). Characterisation of a rape (*Brassica napus* L.) extensin gene encoding a polypeptide relatively rich in tyrosine. *Pl Sci* 71:223-231.

49. Gebhardt, C., E. Ritter, T. Debener, U. Schachtschabel, B. Walkemeier, H. Uhrig, and F. Salamini (1989). RFLP analysis and linkage mapping in *Solanum tuberosum*. *Theor Appl Genet* 78:65-75.

50. Geldermann, H. (1975). Investigations on inheritance of quantitative characters in animals by gene markers. 1. Methods. *Theor Appl Genet* 46:319-330.

51. Golberg, R.B., G.A. Galau, R.J. Britton and E.H. Davidson (1973). Non-repetitive DNA sequence representation in sea urchin messenger RNA. *Proc Natl Acad Sci (USA)* 70:3516-3520.

52. Goodman, H., M. Olson, and B. Hall (1977). Nucleotide sequence of a mutant eukaryotic gene : the yeast tyrosine inserting ochre suppressor Sup 4-0. *Proc Natl Acad Sci (USA)* 74:5453-5457.

53. Grodziker, T., J. Williams, P. Sharp, and J. Sambrook (1974). Physical mapping of temperature sensitive mutations of adenoviruses. *Cold Spring Harbor Symp Quant Biol* 39:439-446.

54. Greer, M.A. (1950). Nutrition and goiter. *Physiol Rev* 30:513-548.

55. Grellet, F., D. Delcasso, F. Panabieres and M. Delseny (1986). Organisation and evolution of a higher plant alphoid-like satellite DNA sequence. *J Mol Biol* 187:495-507.

56. Greunbaum, Y., T. Naveh-Many, H. Cedar and A. Razin (1981). Sequence specificity of methylation in higher plant DNA. *Nature* 292:860-862.

57. Gusella, J.F. (1986). DNA polymorphism and human disease. *Ann Rev Biochem* 55:831-854.

58. Havey, M.J.,and F.J. Muehlbauer (1989a). Linkages between restriction fragment length, isozyme, and morphological markers in lentils. *Theor Appl Genet* 77:395-401.

59. Havey, M.J., and F.J. Muehlbauer (1989b). Variability for restriction fragment lengths and phylogenies in lentil. *Theor Appl Genet* 77:839-843.

60. Heaney, R.K., E.A. Spinks, A.B. Hanley, and G.R. Fenwick (1986). Technical Bulletin : Analysis of Glucosinolates in Rapeseed. Agricultural and Food Research Council, Food Research Institute, Norwich, Colney Lane, Norwich NR4 7UA.

61. Helentjaris, T., D. Weber, and S. Wright (1988). Identification of the genomic location of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. *Genetics* 118:353-363.

62. Helentjaris, T. (1987). A genetic linkage map for maize based on RFLP's. *Trends in Genet* 3(8):217-221.

63. Helentjaris, T., M. Slocum, S. Wright, A. Schaefer, and J. Nienhuis (1986a). Construction of genetic linkage maps in maize and tomato using restriction fragment polymorphisms. *Theor Appl Genet* 72:761-769.

64. Helentjaris, T., D.F. Weber, and S. Wright (1986b). Use of monosomics to map cloned DNA fragments in maize. *Proc Natl Acad Sci (USA)* 83:6035-6039.

65. Helentjaris, T., K Gretchen, M. Slocum, C. Siedenstrang, and S. Wegman (1985). Restriction fragment length polymorphism as probes for plant diversity and their development as tool for applied plant breeding. *Pl Mol Biol* 5:109-118.

66. Helentjaris, T., and R. Gesteland (1983). Evaluation of random cDNA clones as probes for human restriction fragment polymorphism. *J Mol Appl Genet* 2:237-247.

67. Hepburn, A., F.C. Belanger and J.R. Mathies (1987). DNA methylation in plants. *Developmental Genet* 8(5/6): 475-494.

68. Hughes, W.G. (1988). Personal communication. Plant Breeding International Cambridge Ltd, Maris Lane, Trumpington, Cambridge CB2 2LQ.

69. Jensen, J. (1989). Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. *Theor Appl Genet* 78:613-618.

70. Josefsson, E. (1973). Studies of the biochemical background to differences in glucosinolate content in *Brassica napus* L. III. Further studies to localise metabolic blocks. *Physiol. Plant* 29:28-32 (Cited in Schnug, 1990).

71. Joseffson, E. (1971). Studies of the biochemical background to differences in glucosinolate content in *Brassica napus* L. II. Administration of some sulphur-35 and carbon-14 compounds and localisation of metabolic block. *Physiol Plant* 24:161-175 (cited in Schnug, 1990).

72. Kato, A., K. Yakura, and S. Tanifuji (1984). Sequence analysis of *Vicia faba* repeated DNA, the *Fok* I repeat element. *Nucleic Acids Res* 12:6415-6426.

73. Kam-Morgan, L.N.W., B.S. Gill, and S. Muthukrishnan (1989). DNA restriction fragment length polymorphisms: a strategy for genetic mapping of D genome of wheat. *Genome* 32:724-732.

74. Keim, P., R.C. Shoemaker, and R.G. Palmer (1989). Restriction fragment length polymorphism diversity in soybean. *Theor Appl Genet* 77:786-792.

75. Kessler, C., T.S. Neumaier and W. Wolf (1985). Recognition sequences of restriction endonucleases and methylases–a review. *Gene* 33:1-102.

76. Kit, S. (1961). Equilibrium sedimentation in density gradient of DNA preparations from animals tissues. *J Mol Biol* 3:711-716.

77. Kondra, Z.P., and B.R. Stefansson (1970). Inheritance of the major glucosinolates of rapeseed (*Brassica napus*) meal. *Can J Pl Sci* 50:643-647.

78. Koukalova, B., J. Reich, R. Matyasek, V. Kuhrova, and M. Bezdek (1989). A *Bam* HI family of highly repeated DNA sequences of *Nicotiana tabacum*. *Theor Appl Genet* 78:77-80.

79. Kubelik, A.R., J. K. Williams, and K.J. Livak (1991). *in press.*

80. Kustervansomeren, M.A., R.A. Samson, and J. Visser (1991). The use of RFLP Analysis in the classification of the black *Aspergilli*-Reinterpretation of the *Aspergillus niger* aggregate. *Curr Genet* 19(1):21-26.

81. Lakshmikumaran, M. and S.A. Ranade (1990). Isolation and characterisation of a highly repetitive DNA of *Brassica campestris*. *Plant Mol Biol* 14:447-448.

82. Landers, E.S., and P. Green (1987). Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci(USA)* 84:2363-2367.

83. Landers, E.S., P. Green, J. Abrahamson, A. Barlow, M. Daly, S. Lincoln, and L. Newberg (1987). MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181.

84. Landry, B.S., R.V. Kesseli, B. Farrara, and R.W. Michelmore (1987a). A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance, and morphological markers. *Genetics* 116:331-337.

85. Landry, B.S., R. Kesseli, H. Leung, and R.W. Michelmore (1987b). Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphism in lettuce (*Lactuca sativa* L.). *Theor Appl Genet* 74:646-653.

86. Landry, B.S., and R.W. Michelmore (1985). Selection of probes for restriction fragment length analysis from plant genomic clones. *Pl Mol Biol Rep* 3(4):174-179.

87. Larsen, P.D. (1981). Glucosinolates. In: *The Biochemistry of Plants, A Comprehensive Treatise (Volume 5, Secondary Plant Products, Conn, E.E., ed.)*, Stumpf, P.K., and E.E. Conn, eds.. Academic Press, London.

88. Leclerk, R.F., A. Siegel (1987). Characterisation of repetitive elements in several *Cucurbita* species. *Plant Mol Biol* 8:4321-4325.

89. Lee, M., E.B. Godshalk, K.R. Lamkey, and W.W. Woodman (1989). Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. *Crop Sci* 29:1067-1071.

90. Lee, D., L. Turner, D.R. Davies, and T.H.N. Ellis (1988). An RFLP markers for r_b in pea. *Theor Appl Genet* 75:362-365.

91. Lein, K.A. (1972a,b). As cited in: Schnug, E. (1990).

92. Logemann, J., J. Schell, L. Willmitzer (1987). Improved method for the isolation of RNA from plant tissues. *Anal Biochem* 163:16-20.

93. Maniatis, T., E.F. Fritsch, and J. Sambrook (1982). *Molecular Cloning.* Cold Spring Harbor Laboratory, Cold spring Harbor, New York.

94. Manicom, B.Q., M. Bar-Joseph, A. Rosner, H. Vigodsky-Haas, and J.M. Kotze (1987). Potential applications of random DNA probes and restriction fragment length polymorphisms in the taxonomy of the fusaria. *Phytopathology* 77:669-672.

95. Manuelidis, L. (1982). Repeated DNA sequences and nuclear structure. In: *Genome Evolution*, Dover, G.A., and R.B. Flavell, eds. Academic Press, London.

96. Marck, C. (1988). DNA Strider – A C-programme for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acid Res.* 16(5):1829-1836.

97. Martin, B., J. Nienhuis, G. King, and A. Schaefer (1989). Restriction fragment length polymorphisms associated with water use efficiency in tomato. *Science* 243:1725-1728.

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

99. May, C.E., and R. Appels (1987). Variability and genetics of spacerDNA sequences between the ribosomal-RNA genes of hexaploid wheat (*Triticum aestivum*). *Theor Appl Genet* 74:617-624.

100. McCough, S.R., G. Kochert, Z.H. Yu, Z.Y. Wang, G.S. Khush, W.R. Coffman, and S.D. Tanksley (1988). Molecular mapping of rice chromosomes. *Theor Appl Genet* 76:815-829.

101. McNaughton, I.H. (1976). Turnip and relatives. In: *Evolution of crop plants.* Simmonds, H.W. ed. pp 45-48. Longman, New York.

102. Messing, J. (1979). A multipurpose cloning system based on single-stranded DNA bacteriophage M13. *Recomb DNA Tech Bull* 2(2):43.

103. Miller, K. (1987). Gel electrophoresis of RNA. *Focus* 9(3):14-15.

104. Murphy, D.J., J.N. Keen, J. O'Sullivan, D.M.Y. Au, E. Edwards, P.J. Jackson, I. Cummings, T. Gibbons, C.H. Shaw, and A.J. Ryan (1991). A class of amphipathic proteins assiciated with lipid storage bodies in plants. Possible similarities with animal serum apoliproteins.*In press.*

105. Nam,H-G, J. Giraudat, B. den Boer, F. Moonan, W.D.B. Loos, B.M. Hauge, and H.M. Goodman (1989). Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana. Pl Cell* 1:699-705.

106. Nienhuis, J., T. Helentjaris, M. Slocum, B. Ruggero, and A. Schaefer (1987). Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. *Crop Sci* 27:797-803.

107. Olson, M., L. Hood, and D. Botstein (1989). A common language for physical mapping of the human genome. *Science* 245:1434-1435.

108. Osborne, T.C., D.C. Alexander, and J.F. Fobes (1987). Identification of Restriction fragment length polymorphism linked to genes controlling soluble solids content in tomato fruit. *Theor Appl Genet* 73:350-356.

109. Ott, J. (1985). *Analysis of human genetic linkage.* John Hopkins University Press, Baltimore, Maryland, USA.

110. Peacock, W., E.S. Dennis, M.M. Rhoades, and A.J. Pryor (1981). Highly repeated DNA sequence limited to knob heterochromatin in maize. *Proc Natl Acad Sci (USA)* 78:4490-4494.

111. Petes, T.K., and D. Botstein (1977). Simple Mendelian inheritance of the reiterated ribosomal DNA of yeast. *Proc Natl Acad Sci (USA)* 74:5091-5095.

112. Polans, N .O., N.F. Weeden, and W.F. Thompson (1985). Inheritance, organisation, and mapping of *rbcS* and *cab* multigene families in pea. *Proc Natl Acad Sci (USA)* 82:5983-5087.

113. Reddy, A.S., V. Srivastava and S. Guhamukherjee (1989). A tandemly repeated DNA sequence from *Brassica juncea. Nucleis Acids Res* 17(14):5849.

114. Reeds, K.C., and D.A. Mann (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res* 13(20):7207-7210.

115. Rees, H.G., G. Jenkins, A.G. Seal, and J. Hutchinson (1982). Assays of the phenotypic effects of changes in DNA amounts. In: *Genome Evolution*, Dover, G.A. and R.B. Flavell, eds. Academic Press, London.

116. Rimpau, J., D.B. Smith, and R.B. Flavell (1980). Sequence organisation in barley and oats chromosomes revealed by interspecies DNA/DNA hybridisation. *Heredity* 44:131-149.

117. Rimpau, J., D.B. Smith, and R.B. Flavell (1978). Sequence organisation analysis of the wheat and rye genomes by interspecific DNA/DNA hybridisation. *J Mol Biol* 123:327-359.

118. Ritter, E (1990). (A computer programme for analysing segregation data). As cited in Gebhardt *et al.*, 1989.

119. Rivard, S.R., M. Cappadocia, G. Vincent, N. Brisson, and B.S. Landry (1989). Restriction fragment length polymorphism (RFLP) analyses of plants by *in vitro* anther culture of *Solanum chacoense* Bitt. *Theor Appl Genet* 78:49-56.

120. Rivin, C.J., E.A. Zimmer, C.A. Cullis, V. Walbot, T. Huynh, and R.W. Davis (1983). Evaluation of genomic variability at the nucleic acid level. *Pl Mol Biol Rep* 1:9-16.

121. Rodin, J., M.L. Ericson, L.G. Joseffson, and L. Rask (1990). Characterisation of a cDNA encoding a *Brassica napus* 12S protein (cruciferin) subunit. *J Biol Chem* 265(5):2720-2723.

122. Rosen, B, C. Hallden, and W.K. Heneen (1988). Diploid *Brassica napus* somatic hybrids: Characterisation of nuclear and organellar DNA. *Theor Appl Genet* 76:197-203.

123. Saaghai-Marrof, M.A., K.M. Soliman, R.A. Jorgensen, and R.W. Allard (1984). Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and polulation dynamic. *Proc Natl Acad Sci (USA)* 81:8014-8018.

124. Saiki, R.K., D.H. Gelfand, S.J. Stoffe, R. Higuchi, G.T. Horn, K.B. ullis, and H.A. Erlich (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.

125. Sambrook, J., T. Maniatis, and E.F. Fritsch (1989). *Molecular Cloning.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

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

127. Schnug, E. (1990). Glucosinolates – Fundamental, environmental and agricultural aspects. In: *Sulphur Nutrition and Sulphur Assimilation in Higher Plants–Fundamental, environmental, and Agricultural Aspects.* Rennenberg, H., C. Brunold, L.J. DeKok and I. Stulen, eds. SPB Publishing, The Hague.

128. Schwenke, K.D., B. Raab, P. Plietz, and G. Damaschun (1983). The structure of the 12S globulin from rapeseed (*Brassica napus* L.). *Die Nahrung* 27:165-175.

129. Scofield, A., and M. Crouch (1987). Nucleotide sequence of a member of the napin storage protein family from *Brassica napus. J Biol Chem* 262(25):12202-12208.

130. Scutt, C. P. (1990). *The molecular biology of self-incompatibility in Brassica.* PhD Thesis, University of Durham, South Road, Durham DH1 3LE, U.K..

131. Sears, E.R. (1954). The aneuploids of common wheat. *University of Missourri Res Bull* 572:1-59.

132. Sharp, P.J., S. Chao, S. Desai, and M.D. Gale (1989). The isolation, characterisation, and application in the *Triticeae* of a set of wheat RFLP probes identifying each homeologous chromosome arm. *Theor Appl Genet* 78:342-348.

133. Shirsat, A.H. (1984). *Gene for legumin: A major storage protein of Pisum sativum.* PhD Thesis, University of Durham, South Road, Durham DH1 3LE, UK.

134. Simoens, C.R., J. Gielen, M. Van Montagu and D. Inze (1988). Characterisation of highly repetitive sequences of *Arabidopsis thaliana*. *Nucleic Acids Res* 16(4):6753-6766.

135. Simon, A.E., K.M. Tenbarge, S.R. Scofield, R.R. Finkelstein and M.L. Crouch (1985). Nucleotide sequence of a complementary DNA clone of *Brassica napus* 12S storage protein shows homology with legumin from *Pisum sativum*. *Pl Mol Biol* 5:191-201.

136. Singer, M.F. (1982). Highly repeated sequences in mammalian genome. *Int Rev Cytol* 76:67-112.

137. Slocum, M.K., S.S. Figdore, W.C. Kennard, J.Y. Suzuki,and T.Y. Osborn (1991). The linkage rearrangement of restriction fragment length polymorphism loci in *Brassica oleraceae*. *Theor Appl Genet. (in press.)*

138. Smith, J.P. Jr. (1977). *Vascular Plant Family.* Mad River Press, Eureka, Kansas, USA.

139. Soller, M., and J.S. Beckman (1983). Genetic polymorphism in varietal identification and genetic improvement. *Theor Appl Genet* 67:25-33.

140. Song, K., T.C. Osborn, and P.H. Williams (1990). *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 3. Genome relationships in *Brassica* and related genera and the origin of *B. oleraceae* and *B. rapa* (syn.*campestris*). *Theor Appl Genet* 79:497-506.

141. Song, K.M, T.C. Osborn, and P.H. Williams (1988a). Brassica taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 2. Preliminary analysis of subspecies within *B. rapa* (syn.*campestris*) and *Brassica oleraceae*. *Theor Appl Genet* 76:593-600.

142. Song, K., T.C. Osborn, and P.H. Williams (1988b). *Brassica* taxonomy based on nuclear DNA restriction fragment length polymorphisms (RFLPs). 1. Genome evolution of diploid and amphidiploid species. *Theor Appl Genet* 75:784-794.

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

144. Suiter, K., J. Wendel, and J. Case (1983). LINKAGE-1: a Pascal computer programme for the detection and analysis of genetic linkage. *J Hered* 74:203-204.

145. Tanksley, S.D., N.D. Young, A.H. Paterson and M.W. Bonierbale. (1989) RFLP mapping in plant breeding:new tool for an old science. *Bio/Technology* 7:257-264.

146. Tanksley, S.D., J. Miller, A. Paterson, and R. Bernatzky (1988a). Molecular mapping of plant chromosomes. In: *Chromosome Structure and Function*, Gustafson, J.F., and R. Appels, eds. Plenum Press, New York, pp 157-173.

147. Tanksley, S.D., R. Bernatzky, N.L. Lapitan, and J.P. Prince (1988b). Conservation of gene repertoire but not gene order in pepper and tomato. *Proc Natl Acad Sci (USA)* 85:6419-6423.

148. Tanksley, S.D., and J. Hewitt (1988c). Use of molecular markers in breeding for soluble solids content in tomato – a re-examination. *Theor Appl Genet* 75:811-823.

149. Tanksley, S.D. (1983). Molecular markers in plant breeding. *Pl Mol Biol Rep* 1:3-8.

150. Tanksley, S.D., H. Medina-Filho, and C.M. Rick (1982). Use of naturally-occuring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. *Heredity* 49:11-25.

151. Thomas, P.S. and M.N. Farquhar (1978). Specific measurement of DNA in nuclei and nucleic acids using diaminobenzoic acid. *Anal Biochem* 89:35-44.

152. Thompson, K.F., and W.G. Hughes (1986). Breeding and varieties. In: *Oilseed Rape*, Scarisbrick, D.H., and R.W. Daniels, eds. Collins, London, pp 32-82.

153. Tremousaygue, D., F. Grellet, M. Delseny, R. Delourme, and M. Renard (1988). The large spacer of a nuclear ribosomal RNA genes from radish : organisation and use as probe in rapeseed breeding. *Theor Appl Genet* 75:298-304.

154. U, N. (1935). Genomic analysis in the *Brassicae*, with special reference to the experimental formation of *Brassica napus*. *Japan J Bot* 7:389-452.

155. Underhill, T. (1990). As cited in: Schnug, 1990.

156. Vanyushin, B.F. and A.N. Belozerskii (1960). Nucleotide composition of the deoxyribonucleotides of higher plants. (As cited in: Hepburn *et al.*, 1987).

157. Vosberg, H.P. (1989). The polymerase chain reaction: An improved method for the analysis of nucleic acids. *Hum Genet* 83:1-15.

158. Walbot, V., and R. Goldberg (1979). Plant genome organisation and its relationship to classical plant genetics. In: *Nucleic acids in plants*, T.C. Hall and J. Davies, eds.. CRC Press, Boca Raton, Florida, USA.

159. Walton, M. (1988). RFLP information package. CERES, NPI, 417 Wakara Way, Salt Lake City, UT 84108, USA.

160. Watkin, P. (1988). Restriction fragments length polymorphism (RFLP): application in human chromosome mapping and genetic disease research. *BioTechnique* 6(4):310-319.

161. Welsh, J. and M. McClelland (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* 18(24):7213-7218.

162. Welsh, J., and M. McClelland (1991). Genomic fingerprints produced by PCR with consensus tRNA gene primers. *Nucleic Acids Res* 19(4):861-866.

163. Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18(22):6531-6535.

164. Wu, T. and R. Wu (1987). A new rice repetitive DNA shows sequence homology to both 5S and tRNA. *Nucleic Acids Res* 15:5913-5923.

165. Young, N.D., and S.D. Tanksley (1989). RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. *Theor Appl Genet* 77:353-359.

166. Young, N.D., J.C. Miller, and S.D. Tanksley (1987). Rapid chromosomal assignment of multiple genomic clones in tomato using primary trisomics. *Nucleic Acids Res* 15(22):9334-9348.

167. Zamir, D., and Y. Tadmor (1986). Unequal segregation of nuclear genes in plants. *Bot Gaz* 147:355-358.