The Potential Use of Sugarcane Varieties for the Identification of Genetic Markers

by

Julie Megan Barnes

Submitted in fulfilment of the academic
requirements for the degree of
Master of Science
in the
Department of Botany,
University of Natal

Pietermaritzburg
1996

PREFACE

The experimental work described in this thesis was carried out in the Biotechnology Department of the South African Sugar Association Experiment Station (SASEX), Mt Edgecombe, from February 1995 to December 1996, under the supervision of Prof. F.C. Botha.

These studies represent original work by the author and have not been submitted in any form to another University. Where use was made of the work of others, it has been duly acknowledged in the text.

ACKNOWLEDGEMENTS

My sincere thanks go to my supervisor, Prof. Frikkie Botha (SASEX), for his unfailing advice, support and encouragement throughout the duration of this study, and for remembering that we are human beings before being scientists.

My thanks and appreciation are also extended to the following:

Dr Barbara Huckett, Mornay Harvey and Dr Stuart Rutherford (Biotechnology Department, SASEX) for the use of some of their DNA samples;

The staff of the Plant Breeding Department (SASEX) for assistance in the collection of sugarcane varieties;

The staff of the Plant Pathology Department and the Entomology Department (SASEX) for information on disease and eldana resistance;

Darren Vorster (Biotechnology Department, SASEX) for writing "RAPD-Retrieve" according to my often demanding specifications;

Dr Stuart Rutherford for assistance and suggestions with the statistical analysis;

Dr Barbara Huckett for the critical evaluation of this manuscript;

My colleagues in the Biotechnology Department (SASEX) for advice and encouragement whenever it was needed; and

The South African Sugar Association and the Foundation for Research Development for financial support.

ABSTRACT

The use of genetic markers that are linked to specific traits in sugarcane has the potential to increase the efficiency of the selection of improved varieties. Conventionally, markers are identified by analysing the segregation of potential markers and traits in the progeny of single crosses. However, this approach is not practical for sugarcane breeding programmes where replicated, well characterized progenies do not exist. The objective of this project was to investigate the potential of using commercial varieties for identifying markers associated with some of the important traits in sugarcane. This approach would be far more effective than dealing with single progenies since the traits of commercial varieties have already been characterized.

The DNA of fifty commercial varieties of sugarcane was amplified by RAPD PCR using forty-one arbitrary decamer primers. Analysis of the resulting banding profiles, obtained by agarose gel electrophoresis, yielded fifty-four reliable polymorphic fragments. Two approaches were used to identify putative markers linked to the traits of resistance to eldana, sugarcane mosaic virus, and smut: (1) a correlation approach which attempted to identify whether the presence of any polymorphisms could be used to imply the existence of a particular phenotypic state, and (2) multiple regression analysis, in order to determine whether polymorphisms could be used to predict the performance of the varieties for each of the traits. Both approaches appeared to identify associations between polymorphisms and the traits, although multiple regression analysis yielded the most informative results and was able to assign statistical values to the associations.

Using multiple regression, the best predictive model was obtained for sugarcane mosaic virus resistance. This model consisted of four polymorphisms and had an r^2 of 0.401. By dividing the resistance ratings into three groups (resistant, intermediate and susceptible), 52% of the varieties were correctly classified and only 2% of the varieties were predicted in opposite groups (i.e. predicted susceptible when actually resistant, and vice versa). The predictive model for eldana resistance consisted of four polymorphisms and had an r^2 of 0.347. This model classified 30% of the varieties in the correct group of three while none of the varieties were predicted in opposite

groups. The predictive model for smut resistance consisted of three polymorphisms and had an r^2 of 0.316. This model classified 30% of the varieties in the correct group of three while 2% of the varieties were predicted in opposite groups.

Further analysis of sugarcane varieties using additional polymorphisms has the potential to identify markers linked to important traits. These markers could be used for marker-assisted selection to increase the efficiency of selecting for improved sugarcane genotypes for commercial release.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Allele-specific associated primers	ASAPs
Arbitrarily-primed polymerase chain reaction	AP-PCR
Base pairs	bp
Bulked segregant analysis	BSA
Lambda phage	λ
Logarithm of the odds ratio	LOD
Marker-assisted selection	MAS
Near isogenic lines	NILs
Polymerase chain reaction	PCR
Quantitative trait locus	QTL
Random amplified polymorphic DNA	RAPD
Restriction fragment length polymorphism	RFLP
Sequence characterized amplified region	SCAR
Single dose restriction fragment	SDRF
South African Sugar Association Experiment Station	SASEX
Sugarcane mosaic virus	SCMV

CHAPTER 1

INTRODUCTION

The sugarcane selection programme at the South African Sugar Association Experiment Station (SASEX), Mt Edgecombe, takes about 14 years from the time crosses are made between appropriate parent varieties to the commercial release of a variety. This prolonged time period is largely due to the reliance of the selection process on phenotypic characters. The phenotype of sugarcane is a complex interaction of the genotype of the variety and the environmental conditions under which it was grown (Skinner *et al.*, 1987). This can be an important source of error since the greater the environmental variance, the more difficult it is to distinguish genotypic variation on phenotypic grounds. Since the same varieties perform differently in different trials and in different years, superior varieties are selected from many consecutive trials rather than from a single conclusive trial. If genetic markers can be found that are linked to specific traits in sugarcane, the breeding and selection programme would have a means to select for traits at the genetic level, in the absence of environmental influence. The accurate, early selection of traits would mean that the unnecessary carriage of undesirable clonal genotypes to successive stages of the selection programme would be minimised.

The linkage of genetic markers to particular genes can easily be established in diploids. Consequently, genetic maps showing the chromosomal location of agronomically important genes have been developed for most major crops. Sugarcane, on the other hand, is a genetically complex crop, and modern varieties are highly polyploid interspecific hybrids. This complexity has limited the genomic mapping of this crop in contrast to diploid species. However, in the last few years, the theory and practice of mapping and molecular marker technology has progressed to the stage where molecular marker maps of the ancestral species of sugarcane (Al Janabi *et al.*, 1994; Da Silva *et al.*, 1995) and commercial sugarcane varieties (D'Hont *et al.*, 1994; Grivet *et al.*, 1996) have been produced with relative ease. This has led to various expectations for marker technology with considerable potential value for the

improvement of this crop. One of these is the identification of markers that are tightly linked to traits of interest.

Studies on genetic linkage between markers and particular genes are invariably based on populations, derived from planned crosses, that segregate for the trait of interest. Comparative analysis of markers and the phenotypic diversity in a simple progeny generally results in the association of markers with the trait of interest. However, due to the large interaction between the genotype of sugarcane and the environmental conditions under which it was grown, extensive field trials over several years are required before the traits of such a progeny can be accurately evaluated in this crop. The general lack of well characterized segregating populations in sugarcane means that the identification of markers linked to large numbers of agronomically important traits using conventional linkage analysis is somewhat limited.

The sugarcane varieties that are currently cultivated are derived mainly from the interspecific hybridization of *Saccharum officinarum* and *S. spontaneum* that was carried out in the late 1800's. *S. barberi* and *S. sinense* were also used in the breeding of sugarcane varieties, however the contribution of these species was very limited (Berding & Roach, 1987), and they are themselves the product of natural interspecific hybridization between *S. officinarum* and *S. spontaneum* (Daniels & Roach, 1987). The progeny obtained from the early crosses between *S. officinarum* (the "noble" canes) and *S. spontaneum* was subsequently nobilized by systematic backcrossing with *S. officinarum* to allow the hybrids to rapidly recover the more favourable characters from *S. officinarum*. The process of nobilization is characterized by asymmetric chromosome transmission. During a cross between *S. officinarum* and *S. spontaneum*, the female *S. officinarum* generally transmits two haploid chromosome sets while the male *S. spontaneum* transmits one. This 2n+n transmission is repeated in the first backcross, and results in the initial increase in chromosome number in the hybrids. After a few generations of nobilization, further variety improvement involved the intercrossing of varieties produced from this process. An effect of the 2n+n chromosome transmission is that modern cultivated varieties only have between 5 to 10% of their chromosomes from *S. spontaneum*. Furthermore, the genetic base of these varieties is narrow in relation to the potential range of genetic material available. The reason for this is that due to the early

success sugarcane breeders obtained with interspecific hybridization and nobilization, only a few of the available clones of *S. officinarum* and *S. spontaneum* were used as a source of germplasm.

The narrow genetic base of modern sugarcane varieties prompted the suggestion that these varieties can essentially be considered as a large progeny derived from a cross between *S. spontaneum* and *S. officinarum* (Lu *et al.*, 1994a, 1994b). The analysis of a population of modern varieties for genetic markers that are linked to various traits is a far more practical approach than the analysis of a true progeny, since the traits of varieties have already been characterized during the selection programme. Taking into account the narrow genetic background of sugarcane varieties, the low number of meioses since the original interspecific crosses, and the apparent low rate of recombination in sugarcane (D'Hont *et al.*, 1994), then most of the chromosomes initially contributed by the progenitors may have conserved the initial arrangements between markers and useful genes (Lu *et al.*, 1994a).

The objective of this study was to investigate the potential of identifying genetic markers associated with particular traits in sugarcane, using a collection of varieties rather than a progeny derived from a specific cross. By treating the varieties as one progeny, *S. officinarum* and *S. spontaneum* were considered as the parents of the progeny and were included in the analysis. This was done to allow the origin of any potential markers, and hence the traits to which they may be linked, to be traced back to the two progenitors of sugarcane varieties. The selection programme at SASEX selects for many *S. spontaneum* characters such as hardiness, disease resistance and ratooning ability. Thus, the few *S. spontaneum* chromosomes that are left (5-10%) may be carrying some very important genes, for which markers would be very useful.

It is interesting to note that during the course of this study a report was published on the use of varieties for the identification of genetic markers in rice (Virk *et al.*, 1996). The markers obtained in that study were used to predict the performance of various quantitative traits in other rice varieties, and according to the authors, this approach has potential for marker-assisted selection in breeding programmes. Since Virk *et al.* (1996) argue that there is no

reason why the principles that are applied to a segregating population cannot be applied to a collection of varieties as well, the results of the rice study can be regarded as favourable support for the use of varieties to identify markers in sugarcane.

There is a wide variety of agronomically important traits that can be used in a study aimed at finding genetic markers. The ideal traits for which to find markers would be those showing monogenic segregation, however, the genetic complexity of sugarcane has meant that simple Mendelian inheritance has limited applicability in sugarcane (Hogarth, 1987). Disease resistance traits are probably the most suitable candidates for finding markers in sugarcane since, if not simply inherited, then fewer genes are likely to be involved in the expression of these traits than with complex traits such as crop yield or sucrose production. Markers for disease resistance would be particularly useful, since one of the major limitations of sugar production in South Africa is the susceptibility of available sugarcane varieties to pathogens (fungal, bacterial and viral) and pests. Inadequate resistance to pathogens and pests directly limits sucrose production, and also restricts the output of improved varieties from the plant breeding programme.

Resistance to the sugarcane borer *Eldana saccharina*, sugarcane mosaic virus, and the fungal pathogen smut (*Ustilago scitamineae*) were selected for this study. Although these three traits all involve resistance to pests or diseases, the actual mechanisms and genetic bases of resistance are likely to be quite varied and this may have affected the relative ease with which potential markers were found. Resistance of sugarcane to eldana is probably a fairly complex trait. Factors such as stalk surface wax components, flavonoid components of nodal budscales, and fibre content are believed to contribute to the observed levels of resistance to this pest (Nuss *et al.*, 1986; Rutherford *et al.*, 1995). In contrast, many cases of resistance to plant viruses are under very simple genetic control, although vector transmission of viruses can also affect the apparent resistance of the plant (Fraser, 1990). Resistance to sugarcane mosaic virus may thus be a far more simple trait than eldana resistance. Although the genetic basis of resistance to smut in sugarcane is unknown, many cases of resistance to fungal diseases are under the control of a number of genes. However, the actual number of genes that are involved is probably smaller than for typical polygenic characters (Geiger & Heun, 1989).

The random amplified polymorphic DNA (RAPD) technique was used to amplify DNA polymorphisms in a collection of sugarcane varieties. Two methods were then investigated in an attempt to identify associations between the RAPD polymorphisms and the three traits. The first of these was a simple correlation analysis, the aim of which was to determine whether the presence of any polymorphisms could be used to imply the existence of a particular phenotypic state in the varieties. The second method utilized multiple regression analysis in order to determine whether polymorphisms could be used to predict the performance of the varieties for each of the traits.

CHAPTER 2

A HISTORY OF SUGARCANE CULTIVATION AND BREEDING

A knowledge of the history of the cultivation and breeding of sugarcane is important in understanding how sugarcane varieties today function on a genetic level, and how this affects and challenges genetic research and the development of methods for the genetic improvement of these varieties.

2.1 Taxonomy and Origin of Sugarcane and its Direct Relatives

Sugarcane belongs to the *Saccharum* genus, which is a member of the grass family, Gramineae. Other genera, such as *Erianthus*, *Miscanthus*, *Sclerostachya* and *Narenga*, are closely related to *Saccharum* and together constitute an interbreeding group which is believed to have been involved in the origin of sugarcane (Daniels & Roach, 1987). This group has been described as the "*Saccharum* complex".

There are six recognized species of *Saccharum*, i.e. *S. robustum*, *S. spontaneum*, *S. officinarum*, *S. sinense*, *S. barberi* and *S. edule* (Daniels & Roach, 1987). *S. spontaneum* and *S. robustum* are both wild species which contain little or no available sucrose. *S. robustum* is indigenous to New Guinea and adjacent islands, and is proposed to have been derived from introgression of *S. spontaneum* with other related genera. *S. spontaneum* has a wider indigenous range, extending from southern Asia, Indonesia and Melanesia to Africa and the Mediterranean, and is believed to be a product of evolutionary introgression between various members of the *Saccharum* complex. The remaining four species are all cultivated forms. *S. edule* is a small group of clones of probable intergeneric origin and is restricted to New Guinea and its neighbouring islands. It produces aborted inflorescences which are used as a food by the native inhabitants of the islands, and due to its sterility it reproduces vegetatively. *S.*

officinarum (the "noble" canes), *S. sinense* and *S. barberi* have been cultivated since prehistoric times for sugar production in different parts of the world, up until the advent of man-made hybrids.

S. officinarum is most likely to have originated from the Indonesian and New Guinea islands, and is believed to have evolved from *S. robustum* (Daniels & Roach, 1987). Both *S. barberi* (the Indian sugarcanes) and *S. sinense* (the Chinese sugarcanes) are thought to have arisen due to the introgression of *S. officinarum* with *S. spontaneum* (Daniels & Roach, 1987, Lu *et al.*, 1994b). The evolution of the *Saccharum* species must have taken place many thousands of years ago, and seems to be due to the hybridization of the different ancestral species, and possibly related genera, followed by their migration and geographic isolation in particular areas (Stevenson, 1965).

2.2 History of Sugarcane Cultivation

Sugarcane is known to be one of the oldest cultivated plants in the world. The noble canes (*S. officinarum*) were undoubtedly grown as a food plant in the native gardens of New Guinea and neighbouring islands since earliest times. There are also records of sugarcane and sugar production in India and China since several centuries B.C., although the cultivation of sugarcane in these areas probably extends much further back than this (Stevenson, 1965; Berding & Roach, 1987). These early growers improved their varieties by selecting clones with characters such as a sweeter taste or less fibrous stalks, as well as decorative characters such as striped canes and coloured leaves. This is evidenced by the wide range of variation in the noble canes seen today. Sugarcane was spread from the Pacific islands and India to Persia, the Mediterranean, West Africa, America, the West Indies and other sugarcane growing areas initially by the early explorers and subsequently by organized collecting expeditions (Stevenson, 1965; Berding & Roach, 1987).

Sugarcane improvement in those times involved the collection, evaluation and substitution of naturally occurring clones of predominantly *S. officinarum*, exploiting the natural variation to obtain improved yields due to better adaptation or greater disease resistance. Until the end of

the nineteenth century it was assumed that sugarcane was sterile, and cane plants were vegetatively propagated from stem cuttings. This belief was largely due to the male-sterility of the varieties with which planters in the western hemisphere were familiar. Although there were undoubtedly male-fertile varieties in the Old World, the willingness to accept sugarcane as it was probably resulted in the absence of specific attempts at hybridization.

2.3 Initiation of Sugarcane Breeding

It was in 1858, in Barbados, that the fertility of sugarcane was first reported. However, the seedlings that were obtained subsequent to this discovery were regarded as curiosities and were not evaluated for commercial use. It was only in 1888, when the fertility of sugarcane was rediscovered independently in both Java and Barbados, that the production of seedlings became accepted as a means of producing better sugarcane varieties. This resulted in the rapid establishment of organized breeding work, initially in Java and Barbados, and soon after in other parts of the world, leading to the sugarcane breeding and selection programmes in use today in most sugarcane producing areas.

Initially, commercial interest focussed on intercrossing the noble varieties. While proving to be effective in producing clones with better sucrose yields than naturally occurring nobles, this method was not particularly successful in improving the vigor and resistance to most diseases in sugarcane (Berding and Roach, 1987). At the time that sugarcane fertility was discovered, sereh disease was causing extensive damage to sugarcane in Java. This stimulated the collection of material other than *S. officinarum* (including *S. barberi* and *S. sinense*) for interspecific hybridization in an attempt to obtain varieties with resistance to sereh and other diseases. The first attempt at crossing *S. officinarum* with the wild *S. spontaneum* (which was resistant to sereh) failed. However, interest in interspecific crossing remained, and the crossing of *S. officinarum* with *S. spontaneum* and subsequent backcrossing with *S. officinarum* was soon recognized as a means of diluting the negative effects of the wild germplasm. This method became known as nobilization.

S. officinarum generally has the qualities of high sucrose content and low fibre which, together with their large size and good appearance, evidently led to their being referred to as the "noble canes", in contrast to the less spectacular attributes and poorer quality of most canes of the other species. However, the clones of *S. officinarum* do not generally show high levels of resistance to many of the major diseases of sugarcane. *S. spontaneum*, on the other hand, while possessing very low levels of sucrose, has the attributes of resistance to many pests and diseases, tolerance to environmental stress, and adaptability. Nobilization thus became established as a method of maintaining the desirable qualities of *S. officinarum*, retaining the hardiness and disease resistance of *S. spontaneum*, while minimising the otherwise inferior characters of the wild cane (Stevenson, 1965; Berding & Roach, 1987).

2.4 Features and Consequences of Nobilization

Cytological studies by Bremer (1961) showed that nobilization is characterized by asymmetric chromosome transmission. In a cross between *S. officinarum* (2n=80) as the female parent and *S. spontaneum* (2n=40-128) as the male parent, *S. officinarum* generally transmits two haploid chromosome sets while *S. spontaneum* transmits one. This 2n+n transmission is repeated in the first backcross, whereas from the second backcross and in intercrosses, chromosome transmission becomes normal (n+n). Modern varieties, therefore, have chromosome numbers ranging from 100 to 130, of which only 5-10% come from *S. spontaneum*, a characteristic that was recently demonstrated in a study using fluorescent *in situ* hybridization with different coloured probes for *S. officinarum* and *S. spontaneum* DNA, respectively (D'Hont *et al.*, 1996). The asymmetric chromosome transmission has an important effect on the outcome of nobilization since it allows the progeny to recover the desirable characters from the *S. officinarum* parent more rapidly than with normal n+n transmission. Another aspect of the 2n+n transmission is the increase of chromosome number in the hybrids, with the production of allopolyploids of AAB constitution. In addition, nobilized hybrids have a small number of univalents at meiosis (Bremer, 1961) which result in the formation of aneuploids.

It is important to note that although asymmetric chromosome transmission is the norm rather than the exception in *S. officinarum* x *S. spontaneum* hybrids, rare instances of n+n progeny have been reported from these interspecific crosses, and these often involve *S. spontaneum* with 2n=80 (Sreenivasan *et al.*, 1987). A recent report using genomic *in situ* hybridization has shown that n+n transmission can also occur in crosses involving *S. spontaneum* with 2n=64 (D'Hont *et al.*, 1996). It should be emphasized, however, that n+n transmission is an unusual occurrence.

At meiosis, these nobilized hybrids form mainly bivalents and and a few univalents, while multivalents are rare (Bremer, 1961; Stevenson, 1965). This type of chromosome association in a complex hybrid indicates that there can be little or no pairing between the chromosomes of the different *Saccharum* species, and as a consequence, interspecific genetic exchange between the chromosomes must be negligible. This implies that assortment of characters in the hybrids will result from within-species recombination and genetic segregation involving whole chromosomes (Price, 1967), which means that limitations are imposed due to linkage between desirable and undesirable genes. Evidence for recombination between *S. officinarum* and *S. spontaneum* chromosomes has recently been demonstrated in a cultivated sugarcane variety (D'Hont *et al.*, 1996; Grivet *et al.*, 1996). However, these studies indicate that interspecific recombinations remain infrequent events. During selection, therefore, variety improvement in sugarcane hybrids is likely to be due to retention of *S. spontaneum* chromosomes with favourable net effects, while those with unfavourable net effects are rejected.

By crossing and backcrossing to *S. officinarum*, *S. spontaneum* contributes only a quarter of its haploid complement to the hybrids by the second backcross generation, provided no chromosomes are lost at meiosis. However, aneuploids with fewer than the expected number of chromosomes predominate in backcrossing (Berding & Roach, 1987), which means that the *S. spontaneum* contribution can be reduced even further. It is therefore likely that a dosage effect occurs during nobilization whereby the effects of *S. spontaneum* genes are repressed by sufficient *S. officinarum* genes (Price, 1967). This, together with the lack of interspecific chromosome pairing, may largely account for the difficulty in transfer of desirable attributes

from *S. spontaneum* to commercial hybrids.

2.5 The Narrow Genetic Base of Sugarcane Varieties

Although the interspecific hybridization programmes that were initiated in the 1890's in Java revolutionized sugarcane breeding, the genetic base of modern sugarcane varieties is narrow in relation to the potential range of genetic material available. During the early work with nobilization, only a few of the large number of potential clones of *S. officinarum* were used for breeding purposes, while the use of non-*officinarum* sources was even more restricted (Berding & Roach, 1987). *S. spontaneum* was predominantly used, although once again these consisted of only a few of the available clones. Even fewer clones of *S. barberi* and *S. sinense* were involved. The involvement of *S. robustum* germplasm is very limited and is restricted to a few commercial varieties in Hawaii (Sreenivasan *et al.*, 1987). Following the early successes of interspecific hybridization and nobilization, efforts at incorporating additional sources of germplasm were infrequent, and probably more out of curiosity than need. Today's sugarcane varieties, therefore, are mainly the result of three or four generations of nobilization involving mainly *S. officinarum* and *S. spontaneum*, followed by the intercrossing of the resultant hybrids and selection of varieties with improved characters (Bremer, 1961; Berding & Roach, 1987).

Lu *et al.* (1994b) showed that the genome of *S. spontaneum* is the most variable, while that of *S. officinarum* is the most conserved. Consequently, due to the 2n+n transmission of nobilization, much of the genetic diversity has been lost in modern varieties (supported by Harvey & Botha, 1996), and what diversity remains is principally due to genes from *S. spontaneum* (shown in Lu *et al.* 1994a).

2.6 Inheritance in Sugarcane

Due to their high polyploidy and interspecific origin, modern sugarcane varieties are genetically very complex and have high aneuploid chromosome numbers (Sreenivasan *et al.*, 1987). A consequence of this is that there have been relatively few attempts to study

characters thought to be simply inherited, and it would appear as though simple Mendelian inheritance has limited applicability in sugarcane (Hogarth, 1987). The genetic complexity of sugarcane varieties today could indicate that the inheritance of most characters would be of a complicated nature, even for those which behave in a simple Mendelian manner in diploids. However, as mentioned previously, *S. officinarum* x *S. spontaneum* hybrids seem to have essentially regular chromosome pairing at meiosis, with the formation of predominantly bivalents and a few univalents, while multivalent associations are rare. Sugarcane varieties can therefore be said to behave essentially like diploids. While it is only in the complete absence of non-homologous pairing at meiosis that diploid inheritance can be assumed, there is evidence to suggest that inheritance in sugarcane approximates diploidy and that deviation from this assumption may not be too serious (Hogarth, 1987).

2.7 Sugarcane Breeding and Selection Programmes

The aims of sugarcane breeding programmes are to produce varieties with improved characteristics that will prove to be profitable for both growers and millers. Important characters include improved cane yield, higher sucrose content, lower fibre, tolerance to drought and cold, disease resistance, rapid and reliable germination of cuttings, and quick and prolonged ratooning. Focussed breeding programmes in most sugarcane-producing areas around the world have led to significant contributions to characters such as these.

Before they can be commercially released as new sugarcane varieties, the progeny from crosses between appropriate parent varieties have to be screened for suitable characters. The selection process consists of several different stages, and at SASEX, Mt Edgecombe, this takes about 14 years to complete. At each stage, varieties with unsuitable characters are discarded and the survivors are tested in larger plots in order for their performance to be evaluated more reliably (Table 2.1). Those selections which prove to be superior to existing commercial varieties are propagated and then released to growers.

Table 2.1 Sugarcane selection programme at SASEX. Selection criteria in the various stages involve screening for various diseases and the sugarcane borer eldana, agronomic evaluations, and yield and sucrose determinations. Selection is carried out on different farms representing the diverse conditions that exist within the South African sugarcane belt.

Year number	Stage	Approx. number of clones
1	Potted seedlings (from 600 crosses)	180 000
2	STAGE 1: Single stools	130 000
3	STAGE 2: Single lines	12 000
4	Plant crop + 1st ratoon	
5	STAGE 3: Observation trial	1 200
6	2 rows, 2 replications; plant crop	
7	STAGE 4: Primary variety trial	240
8	5 rows, 3 replications; plant crop + 2 ratoons	
9	STAGE 5: Secondary variety trial	35
10	Final evaluation phase	
11	5 rows, 3 replications; plant crop + 2 ratoons	
12	Trials at all regional testing sites	
13		
14	Bulking plots	1
	Release	

The prolonged time period before a new sugarcane variety can be commercially released is largely due to the reliance of the selection process on phenotypic characters. The phenotype of sugarcane is a complex interaction between the genotype of the variety and the environmental conditions under which it was grown (i.e. P=G+E) (Skinner *et al.*, 1987). This can be an important source of error in selection programmes since the greater the environmental variance, the more difficult it is to distinguish genotypic variation on phenotypic grounds, which is essential to variety improvement through breeding. Since the same varieties perform differently in different environments and in different years, superior varieties are

selected from many consecutive trials in a number of locations, rather than from a single conclusive trial.

Although genotype x environment interactions occur in all crops, sugarcane seems to be a more variable crop than others (Johnston *et al.*, 1995). Consequently, variety selection programmes for sugarcane require more trials to have the same accuracy as trials for other crops.

CHAPTER 3

MARKER-FACILITATED GENETIC ANALYSIS FOR CROP IMPROVEMENT: A REVIEW

Markers can fulfil a number of functions in plant breeding and crop improvement. In general, however, the plants that have benefitted the most from the application of DNA marker technology are diploids. Polyploids are genetically more complex than diploids, and it is this complexity that has, until recently, impeded genetic studies in polyploid plant species. Since many polyploids are economically important as major crops (e.g. wheat, potato, and sugarcane), the application of marker technology to these plants would be highly beneficial.

Together with the types of markers available for genetic research, the application of markers to plant breeding programmes will be discussed below. In the appropriate sections, an account of the development of marker technology in sugarcane will be given. This should provide an indication of the status of marker research in sugarcane, as well as emphasize the complexity involved in applying marker technology to this crop.

3.1 The Development of Markers

In the absence of much understanding of the genetic mechanisms controlling characters of agricultural or economic interest, initial plant breeding efforts focussed on developing new varieties by selecting plants with desirable phenotypes. Conventional plant breeding programmes achieve this by generating an F_2 population, screening the plants for desirable traits, followed by repeated backcrossing, selfing and testing (Winter & Kahl, 1995). This process is time-consuming and costly. Furthermore, the phenotype of a plant is not necessarily an accurate measure of its genetic potential, since the environment interacts with the genotype and can mask its effects (Tanksley *et al.*, 1989). In addition to this, factors such as multigenic (quantitative) inheritance of traits, and partial and complete dominance can confound the

expression of genetic traits (Tingey & del Tufo, 1993). The development of markers has resulted in the improvement of the breeding process by allowing the acceleration of the generation of new varieties, and the connection of phenotypic traits with the genetic loci responsible for them.

3.1.1 Morphological Markers

Since plant breeding initially depended on the observation and measurement of phenotypic characters, it became obvious that certain phenotypic differences between parents were often co-inherited but did not necessarily co-segregate in all the progeny. Thus, although some characters seemed linked together, they were occasionally split by genetic recombination, and the extent of the recombination was interpreted as an indication of the extent of linkage (Winter & Kahl, 1995). This led to the development of genetic maps. Initially, these maps were based on morphological characters, including natural or induced mutations such as dwarfism, albinism and altered leaf morphology. Such characters were tested for linkage and assembled into maps that reflected the arrangement of the genes on the chromosomes.

Morphological markers, however, have limited applicability in terms of their use for marker-assisted selection, whereby selection for a gene of interest is conducted by selecting for the presence of a tightly linked marker to that gene. This is primarily due to the limited availability of morphological markers in many crop species, their often recessive inheritance, the influence of the environment on their expression, the undesirable effects many morphological markers have on the phenotype of the plant (e.g. deleterious mutants), and the inability to score multiple morphological marker loci in a single segregating population due to the epistatic or pleiotropic effects of such loci (Tanksley & Orton, 1983; Kelly, 1995).

3.1.2 Isozyme Markers

The disadvantages of morphological markers led to the development of isozymes as markers. Isozymes are allelic variants of specific enzymes having the same specificity but differing in attributes such as pH optima and kinetic properties (Tanksley & Orton, 1983). The different forms of these enzymes can be separated by electrophoresis and identified by specific staining. Unlike most morphological markers which act in a dominant-recessive manner, the co-

dominant nature of the alleles at most isozyme loci allows heterozygotes to be distinguished from homozygotes. Alternative alleles at isozyme marker loci usually have no associated deleterious effects on phenotype, and they rarely exhibit epistatic or pleiotropic interactions, allowing an almost limitless number of segregating markers to be monitored in a single population (Tanksley & Orton, 1983).

The ability to use isozymes as markers makes them ideal to use in linkage studies and the construction of genetic maps of plant chromosomes. Once isozymic genes have been mapped and the linkage map expands, desirable genes which are tightly linked to isozyme markers can be detected in segregating populations. Genomic maps based on isozymes have been developed in a number of plant species, including tomato (Tanksley & Rick, 1980) and maize (Goodman *et al.*, 1980). Despite their advantages over morphological markers, however, isozymes are not without their shortcomings. The numbers of isozyme markers available are limited, and their expression is often restricted to specific tissues or developmental states (Tanksley & Orton, 1983; Winter & Kahl, 1995). Consequently, they are unable to fulfil the full potential of genetic mapping in plant breeding.

3.1.3 Restriction Fragment Length Polymorphisms

In 1980, the development of the first map of the human genome based on molecular markers was described (Botstein *et al.*, 1980). The technology used for this map involved the digestion of genomic DNA with restriction enzymes, followed by hybridization with specific probes. The polymorphisms produced by this technique are known as restriction fragment length polymorphisms (RFLPs), and they have subsequently been used in the development of genomic maps of many other organisms, including many important plant species, e.g. tomato (Bernatzky & Tanksley, 1986), lettuce (Landry *et al.*, 1987), *Arabidopsis thaliana* (Chang et al., 1988), rice (Saito *et al.*, 1991), and sugarcane (Da Silva *et al.*, 1993).

Digestion of genomic DNA with restriction endonucleases generates a mixture of restriction fragments, the number and length of which depend on the distribution of recognition sites in the genome. If individuals differ in the presence of a recognition sequence at a distinct genomic location, differently sized restriction fragments will be generated at that particular

locus by the enzyme. Separation of the restriction fragments by electrophoresis, followed by blotting onto a membrane and hybridization to a specific, labelled DNA probe allows the visualization of these fragments. If probes incorporating highly repetitive DNA sequences are used, they will hybridize with a large number of the restriction fragments and result in a smear on the autoradiograph. For this reason, single copy DNA probes that will detect only one defined genomic fragment each are generally used (Beckmann & Soller, 1986; Tanksley *et al.*, 1989), although multilocus probes can also be used in certain applications (Winter & Kahl, 1995). When two individuals differ in a restriction site that affects the length of a DNA fragment homologous to the probe, the fragments will appear as differently sized bands on the autoradiograph.

RFLPs have a number of properties which make them amenable for use as genetic markers (Beckmann & Soller, 1986; Tanksley *et al.*, 1989). Since an RFLP heterozygote will show two discrete allelic bands, they possess a co-dominant mode of inheritance. This is a desirable attribute, particularly for recessive traits, since it makes it possible to determine whether a linked trait is present in a homozygous or heterozygous state in an individual. RFLP markers are generally phenotype-neutral, are free of environmental influence, have a high level of allelic variation, and are apparently free of epistatic and pleiotropic effects. Furthermore, they are detectable in all tissues and at any stage of plant development.

Although there is a virtually unlimited number of probe/enzyme combinations available, only one out of several RFLP markers provides a polymorphism (Winter & Kahl, 1995). This is a significant problem, particularly in closely related cultivated breeding lines where RFLPs can become scarce. The development of RFLP markers is technically complex, time consuming, and expensive, and involves safety considerations due to the use of radioactive probes. These aspects, coupled with the requirement for large amounts of DNA and cloned and characterized probes, has raised questions as to whether RFLP technology is compatible with large-scale crop improvement programmes, despite its usefulness in generating genetic markers.

3.1.4 Random Amplified Polymorphic DNA Markers

The polymerase chain reaction (PCR) is a powerful technique that has revolutionized molecular biology, and various modifications of the original method have been designed to suit a range of needs. One such variation has been used to generate a type of molecular marker known as random amplified polymorphic DNAs (RAPDs), and these have been able to overcome many of the limitations of RFLP markers described above.

PCR, initially developed by Saiki *et al.* (1985), is a technique designed to amplify specific sequences of DNA. The method is based on three simple steps that are repeated in a cyclic fashion: (1) denaturation of the DNA template into single strands; (2) annealing of primers to each strand for new strand synthesis; and (3) extension of new DNA strands from the primers using DNA polymerase. PCR uses two specific oligonucleotide primers that are complementary to opposite strands of the target sequence and flank the region to be amplified (Figure 3.1). In addition to the two primers, PCR uses a cocktail of a DNA polymerase enzyme and deoxyribose nucleoside triphosphates (dNTPs) which gets mixed with the DNA to allow amplification. Following hybridization of the primers, DNA synthesis by the polymerase proceeds across the region between the primers, effectively doubling the amount of that DNA segment. Since the extension products are complementary to and capable of binding primers, each successive cycle doubles the amount of DNA synthesized in the previous cycle, resulting in the exponential accumulation of the specific target fragment. PCR is performed by incubating the reaction components and the DNA to be amplified at three temperatures in succession, each temperature being specific for the particular step in the cycle.

Saiki *et al.* (1985) originally described the use of the Klenow fragment of *Escherichia coli* DNA polymerase I for the extension of the annealed primers. However, this DNA polymerase is temperature-sensitive. Consequently, the high temperature required in the denaturation step to yield single stranded DNA inactivated the polymerase, and fresh polymerase had to be added after each denaturation step for new DNA strands to be synthesized. PCR technology was drastically improved with the replacement of *E. coli* DNA polymerase with a thermostable DNA polymerase purified from the thermophilic bacterium *Thermus aquaticus* (*Taq*) (Saiki et al., 1988). This heat-resistant polymerase is unaffected by the denaturation step and does

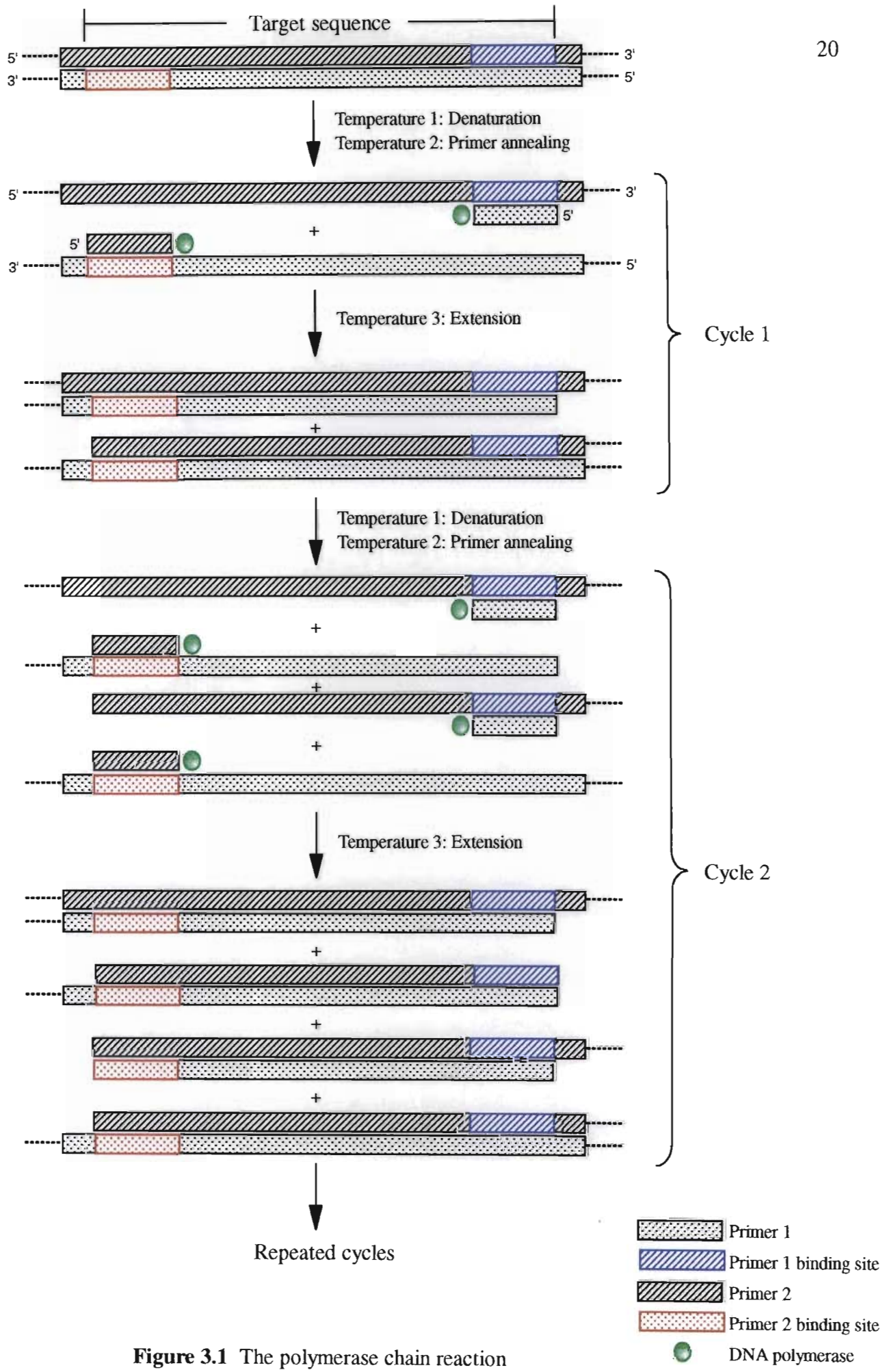


Figure 3.1 The polymerase chain reaction

not have to be replenished at each cycle. In addition to making the procedure amenable to automation, the overall performance of PCR is substantially improved by the use of *Taq* polymerase since the specificity, yield, sensitivity and length of targets that can be amplified are increased (Saiki *et al.*, 1988).

In 1990, two independent but similar studies described a simple method based on PCR whereby single primers of arbitrary nucleotide sequence can be used to reproducibly amplify segments of genomic DNA from a wide variety of species, generating polymorphic genomic profiles (Welsh & McClelland, 1990; Williams *et al.*, 1990). Although this method uses single primers (as opposed to two primers in classical PCR), some of the complementary sequences on the genomic DNA will be within a few hundred base pairs of each other on opposite strands and in opposite directions, allowing sequences between these positions to be PCR-amplifiable. An advantage of this technique is that primers are chosen without regard to specific nucleotide sequence information of the template DNA. Thus, no prior knowledge of the genome of the organism to be investigated is required.

Due to its nature, this technique has been termed arbitrarily primed PCR (AP-PCR) (Welsh & McClelland, 1990), or RAPD assay (after Random Amplified Polymorphic DNA) (Williams *et al.*, 1990). On average, each primer will direct the amplification of several discrete loci in the genome. Polymorphisms among the amplification products of different individuals are frequently detectable. These are due to either changes in the sequence of the primer binding site which prevent the association of primer, or from changes which alter the size or prevent the successful amplification of the target DNA. Generally, size variants are rarely detected and individual amplification products represent one allele per locus (Williams *et al.*, 1990). Amplification products are thus transmitted as dominant markers. Although primers for this technique are of random sequence, they are usually biased to consist of at least 50% G+C, and are usually 9 or 10 nucleotides in length (Williams *et al.*, 1990). Amplification products are easily separated using agarose gel electrophoresis and visualized by ultraviolet illumination of ethidium bromide stained gels (Williams *et al.*, 1990), although polyacrylamide gel electrophoresis and autoradiography can also be used (Welsh & McClelland, 1990).

Williams and co-workers (1990) noted that the polymorphisms generated by the RAPD technique are useful as genetic markers and can be used to construct genetic maps. Since their development, RAPDs have been used successfully in many plant species to develop comprehensive genetic maps and complement already existing RFLP and isozyme maps, e.g. in sugar beet (Uphoff & Wricke, 1995), celery (Yang & Quiros, 1995), grape (Lodhi *et al.*, 1995), maize (Ajmone Marsan *et al.*, 1993), *Arabidopsis thaliana* (Reiter *et al.*, 1992) and sugarcane (Al-Janabi *et al.*, 1993; Da Silva *et al.*, 1995).

The use of RAPD markers has a number of advantages over methods involving RFLPs or isozymes. There is no requirement for DNA sequence information, the isolation of cloned DNA probes or preparation of hybridization filters is not required, the protocol is relatively quick and simple to perform and is amenable to automation, the use of radioactivity is not necessary, and since the RAPD technique is an amplification-based assay, only nanogram quantities of DNA are required as opposed to microgram quantities for RFLPs. Whereas RFLP analyses often fail to detect variation between individuals of the same species or closely related breeding material, RAPDs can detect far greater levels of polymorphism, making them more valuable for construction of intraspecific maps and in gene tagging research in commercially adapted genotypes (Foolad *et al.*, 1993; Haley *et al.*, 1994b). Since RAPD fragments can represent repetitive DNA (Williams *et al.*, 1990), a reason for the increased level of polymorphism detected by RAPDs could be that fragments are amplified in genomic regions which are not accessible to RFLP analysis due to the presence of repetitive DNA sequences.

The main limitation of RAPD markers is their dominant nature, resulting in the inability to determine whether a locus is homozygous or heterozygous. This is particularly important when dealing with recessive traits. This can be overcome by obtaining two closely linked RAPD markers, each specific for one of the alleles, and using them as a pair (Williams *et al.*, 1993). Amplification of both markers of the pair would be diagnostic for a heterozygous genomic region. Alternatively, a RAPD band can be excised and used as a probe to establish sequence similarity of two segregating RAPD bands of different molecular weights, thereby allowing that pair of bands to be scored as a single co-dominant marker (Williams *et al.*,

1993). RAPD bands can also be converted to co-dominant markers by converting them to SCARs, discussed below.

There is some debate as to the reliability of the RAPD technique, since difficulties have been obtained in reproducing RAPD patterns in different laboratories and in different thermocyclers. However, a number of authors (Waugh & Powell, 1992; Ajmone Marsan *et al.*, 1993; Sobral & Honeycutt, 1993) have reported reliable and reproducible RAPD markers. By standardizing laboratory conditions and optimizing components of the RAPD reaction, irreproducibility of RAPDs may, to a large extent, be solved. RAPD fragments have even been shown to be stable across various meiotic events. For example, Haley *et al.* (1994a) screened an extensive collection of common bean genotypes and observed that a RAPD marker and its linkage to a rust resistance gene appeared to have remained intact during different introgression and meiotic events. Huckett and Botha (1995) screened sugarcane varieties involved in the genealogy of two commercial varieties, and found that specific RAPD fragments could be reliably identified across the seven generations of the genealogy.

A slight, but nevertheless important, lack of reproducibility can be caused by imprecise matches between the oligonucleotide primer and template DNA at the low annealing temperature (35°C) usually employed in RAPD reactions. If concern exists as to the reliability of RAPD markers, it is relatively easy to convert the RAPD assay to a more reliable secondary PCR assay. Paran & Michelmore (1993) were able to increase the reliability of RAPD markers by sequencing the two ends of a RAPD fragment and synthesizing two longer primers (24mers) homologous to each end. These primers included the original decamer primer sequence, and when used in a traditional PCR reaction (annealing temperature of 50°C-65°C), they amplified the RAPD fragment specifically and reliably. These markers are referred to as sequence characterized amplified regions (SCARs), and the use of more stringent annealing conditions and longer primers than for RAPDs results in an assay that is more reproducible and only detects one locus.

Although SCAR primers usually retain the presence/absence polymorphism of the original RAPD fragment, they sometimes amplify fragments linked to both alleles at the locus of

interest. In some cases these fragments are of distinct sizes when resolved on agarose gels (Paran & Michelmore, 1993; Horvath *et al.*, 1995), allowing the amplification products to be scored as co-dominant markers and thereby differentiating between homozygous and heterozygous states. In other cases the fragments are of identical size but upon digestion with various restriction enzymes, polymorphisms in fragment length can be found between the two genotypes, yielding co-dominant markers (Ohmori *et al.*, 1996). An alternative approach to obtaining co-dominant markers involves obtaining two different RAPD fragments, each specific for one of the alleles at the locus of interest. By converting these to SCARs and using the primers together, co-amplification of both fragments indicates a heterozygous state, while single fragments indicate their respective homozygous states (Nair *et al.*, 1995).

Gu *et al.* (1995) developed a similar approach to SCARs which not only increases the reliability of RAPD markers, but facilitates the screening of large numbers of samples. The longer primers used for SCARs sometimes detect fragments in both genotypes, as mentioned above. However, by modifying the primer length to be shorter than the 24mers typically used for SCARs, Gu *et al.* (1995) were able to reliably amplify a single DNA fragment (as with SCARs), but only in the individuals possessing the appropriate allele. Since these primers (termed allele-specific associated primers, or ASAPs) maintain the original presence/absence polymorphism observed for the RAPD, the need to separate amplified DNA fragments by electrophoresis is eliminated. Instead, samples possessing the appropriate allele can be identified in a quick plus/minus assay by direct staining of the DNA in the reaction tubes or microtitre plates with ethidium bromide, making the procedure applicable to large-scale screening. Since plants homozygous for the allele cannot be distinguished from heterozygous plants, ASAPs, like RAPDs, are dominant markers. To overcome this, two ASAP markers, one for each allele, could be developed and two separate amplification reactions would then be performed to identify both alleles.

DNA amplification fingerprinting (DAF) (Caetano-Anollés *et al.*, 1991; Caetano-Anollés & Bassam, 1993) is a variation of the RAPD technique. This method uses arbitrary oligonucleotide primers as short as five nucleotides, but typically seven or eight nucleotides in length to direct the enzymatic amplification of discrete portions of genomic DNA. Complex

banding patterns, or fingerprints, are obtained which can be resolved by polyacrylamide gel electrophoresis and silver staining.

With both DAF and RAPDs, two different primers can be used in pairs to obtain additional amplified fragments (Caetano-Anollés *et al.*, 1991; Klein-Lankhorst *et al.*, 1991; Williams *et al.*, 1993). Another technique to enhance the detection of polymorphic DNA involves the digestion of genomic DNA with restriction enzymes prior to amplification. While Williams and co-workers (1993) found that this pretreatment simplified the pattern of amplified products, permitting the interpretation of patterns that would otherwise be too complex, Caetano-Anollés and co-workers (1993) found a significant increase in the detection of polymorphic DNA without decreasing fingerprint complexity. The use of paired primers and the predigestion of DNA with restriction enzymes are two methods which can be used to increase the information content of amplified fingerprints from a limited collection of primers.

3.1.5 Microsatellite Markers

Microsatellites (or simple sequence repeats) are DNA sequences in the genome of many higher organisms, including plants, composed of the tandem repetition of a simple, short (2-5 base pairs in length) sequence motif, e.g. $(CA)_n$. These regions evolve faster than the surrounding DNA and are consequently highly polymorphic, producing a powerful tool for genetic mapping and genetic diagnostics.

Various approaches can be used to obtain microsatellite markers. One of these involves digesting genomic DNA with restriction enzymes, electrophoretically separating the DNA, blotting the DNA onto a membrane, and hybridizing with a labelled simple repetitive oligonucleotide probe such as $(AAT)_6$ (Li & Niwa, 1996). A second approach involves the PCR amplification of genomic DNA using specific microsatellite primers (e.g. $(GACA)_2$) (Tautz, 1989). Since microsatellites are randomly dispersed throughout the genome and are highly variable, highly polymorphic fragment profiles are produced. A third approach involves cloning and sequencing microsatellites and designing PCR primers from the flanking conserved DNA sequences (Cregan *et al.*, 1994). These will amplify the intervening microsatellite sequences, which can then be resolved on agarose or polyacrylamide gels.

Multiple allelic length variants can be identified at most microsatellite loci, providing a source of co-dominant markers.

3.2 Genetic Maps

A detailed genetic map contains a vast amount of information that plant breeders can use to identify, manipulate and complement traits to their maximum advantage. As mentioned above, morphological and isozyme markers formed the basis of the first maps. Later, the development of markers such as RFLPs and RAPDs changed the concept of linkage mapping in terms of plant breeding and greatly enhanced the potential of genetic markers to assist breeding programmes. The goal of genetic mapping is to generate an extensive map for segregating populations quickly and at the breeders' convenience in order to manipulate traits with greater speed and precision. This relies on access to numerous, easily scored genetic markers. The ideal marker should be quick and easy to screen, and have a co-dominant nature. Since such a marker has yet to be developed, a choice has to be made from among those that are available. Today, detailed maps have been developed for many plant species (O'Brien, 1990).

3.2.1 Construction of Genetic Maps

Once a suitable cross has been made between two parental lines, a genetic linkage map is constructed by identifying markers that are polymorphic between the two parents, followed by analysing the segregation of the markers based on their presence or absence in the progeny. This results in the assignment of the markers to various linkage groups in the map. Linkage analysis can be carried out by a computer program such as MAPMAKER (Lander *et al.*, 1987), which tests segregation ratios and calculates linkage relationships based on LOD scores (LOD is an abbreviation for logarithm of the odds ratio, and is a statistical measure of the likelihood that two genes, or markers, are linked).

A variety of mapping populations can be used to generate linkage maps. The conventional method is to use an F_2 population for the segregation analysis (e.g. Landry *et al.*, 1987; Redoña & Mackill, 1996b). The dominant nature of RAPD markers means that individuals

with two copies of an allele (homozygotes) are not distinguishable from individuals with only one copy of an allele (heterozygotes). The disadvantage of using an F_2 population for mapping with dominant markers is the difficulty in demonstrating linkage between markers linked in repulsion (i.e. markers residing on separate chromatids). Therefore, if dominant markers such as RAPDs are used for an F_2 population, only markers linked in coupling (i.e. residing on a single chromatid) should be considered to minimize error (Williams *et al.*, 1993).

In order to overcome this limitation of mapping with dominant markers, backcross populations (where all alleles from the recurrent parent are homozygous, all alleles from the donor parent are heterozygous) or recombinant inbred populations (where nearly all loci are homozygous) can be used (Reiter *et al.*, 1992; Lin & Ritland, 1996). Doubled haploid lines, where all individuals are homozygous (Dion *et al.*, 1995; Kjær *et al.*, 1995), and haploid megagametophytes of some gymnosperms (Kaya & Neale, 1995; Plomion *et al.*, 1995) are also suitable, since dominance-related effects are absent and linkage data can be obtained directly.

Other populations that can be used for producing linkage maps are aneuploid lines, including nullisomic/tetrasomic and monosomic series, which have been useful for mapping polyploids such as hexaploid wheat, and addition, substitution and translocation lines (Tanksley & Orton, 1983).

3.2.2 Mapping Clonally Propagated Crops

Breeding methods used for most annual crops generally involve the development of homozygous lines (inbreds) to use as parents, and the generation of F_1 hybrid varieties which are used to obtain segregating F_2 or backcross populations. It is these populations, or recombinant inbred lines, that are therefore used for estimating recombination percentages to construct linkage maps. However, since the breeding methods used for clonally propagated crops are quite different from the strategy described above, gene mapping approaches are also different (Mehlenbacher, 1995). The most common approach to breeding clonally propagated crops is sometimes called "complementary hybridization". The parent clones for each cross are chosen such that the weaknesses of one are matched by the strengths of the other. Since

the parent clones are relatively heterozygous, a high level of segregation is apparent in the F_1 progeny and the task is then to identify from among the thousands of F_1 seedlings the few that have potential as new cultivars, optimally combining the strengths of both parents and none of their weaknesses. Progeny evaluation is generally stepwise, with the number of selected clones decreasing in each step and the number of vegetatively propagated plants of each clone increasing.

Since a high degree of segregation is observed in the first generation progeny of clonally propagated plants, simultaneous segregation of many loci allows linkage map construction from the F_1 population (Mehlenbacher, 1995). An advantage of mapping with clonally propagated plants is their perennial nature. Since the same plants can be used over a period of years, many additional markers can be placed on the original map.

3.2.3 Integration of Classical and Molecular Maps

In order to fully exploit the potential of genome mapping and allow more efficient use of markers in breeding programmes, classical and molecular maps should be integrated. A good example of how this has been done is provided by rice (Kishimoto *et al.*, 1993). The classical linkage map, established by rice geneticists over the past several decades, includes genes corresponding to important morphological, physiological and agronomic characteristics. However, the lack of linked DNA markers which can be used for marker-assisted selection means that this map is not readily applicable to rice breeding. In comparison, the molecular genetic map (Saito *et al.*, 1991) contains a large number of DNA markers which can be used as "landmarks" for mapping and tagging genes. Kishimoto and co-workers (1993) showed that by aligning the molecular and classical linkage maps of rice, the linear order of phenotypic markers along the linkage groups on the classical map is generally consistent with the order on the molecular linkage map. Similarly, Shoemaker and Specht (1995) integrated the soybean classical map consisting of morphological, pigmentation and isozyme markers, with the molecular map consisting of RFLP and RAPD markers. Many of the maps produced today consist of a variety of the different types of markers (e.g. Tanksley *et al.*, 1992; Akkaya *et al.*, 1995; Rajapakse *et al.*, 1995).

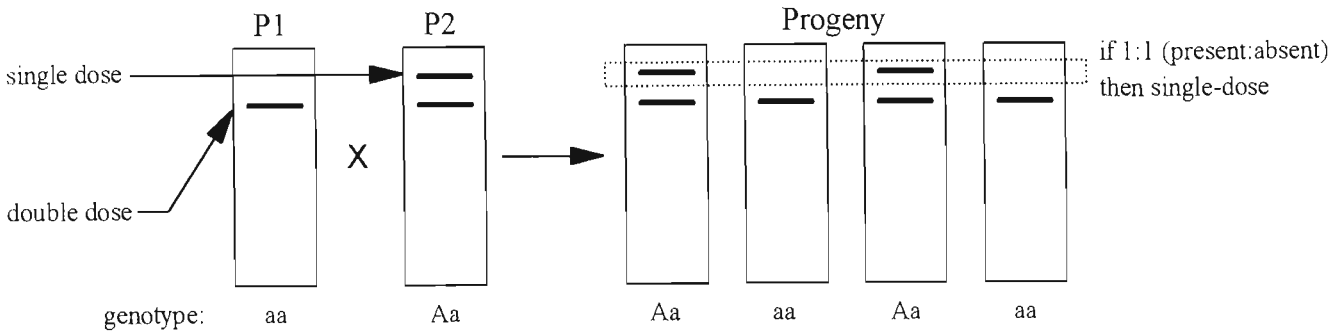
3.2.4 Mapping in *Saccharum*

The elevated ploidy levels and the cytogenetic complexity of interspecific *Saccharum* hybrids have limited classical genetic studies in sugarcane (Hogarth, 1987). One exception to this was the work of Glaszmann *et al.* (1989), where isozyme variation was used to identify biochemical markers of potential utility in sugarcane genetics and breeding. This illustrated how markers could be used as an efficient means of identifying linkage groups in a sugarcane genome (Glaszmann *et al.*, 1989). Large numbers of DNA markers can now be acquired with high efficiency on essentially any previously uncharacterized genome (Sobral & Honeycutt, 1993). Consequently, during the last few years the application of DNA markers has promoted the development of the theory and practice of direct genetic mapping of even the most complex of polyploids (Da Silva & Sobral, 1996). Molecular markers now have the potential to help unravel the complex genetics of *Saccharum* species.

While genetic linkage maps in cultivated diploid species are constructed by analysing populations derived from crossing inbred lines, no equivalent populations exist for polyploids. Linkage maps based on RFLP markers have been constructed in simple polyploids by resorting to diploid relatives in which mapping is more straightforward (e.g. potato, Bonierbale *et al.*, 1988; Gebhardt *et al.*, 1989). However, for some complex polyploids such as sugarcane, there are no closely related diploid species to use for mapping.

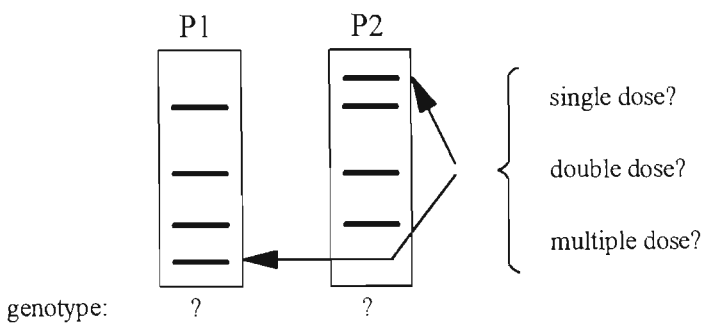
Mapping in polyploid species is inherently more complex than in diploids, basically due to the difficulty in identifying genotypes based on their banding phenotypes. To avoid this difficulty in mapping, Wu *et al.* (1992) proposed a method for mapping polyploids based on the segregation of single-dose restriction fragment length polymorphisms (SDRFs). These markers are present in one parent, absent in the other, and segregate 1:1 (presence:absence) in the progeny. Polymorphic fragments with other segregation ratios cannot be mapped, because there is no statistical method for determining their linkages (Figure 3.2). The single-dose marker approach has been able to extend linkage mapping from inbred lines of diploid species to virtually any cross between two heterozygous individuals of any ploidy level (Da Silva & Sobral, 1996).

A. Diploid Analysis



B. Polyploid Analysis

Parent analysis



Progeny test

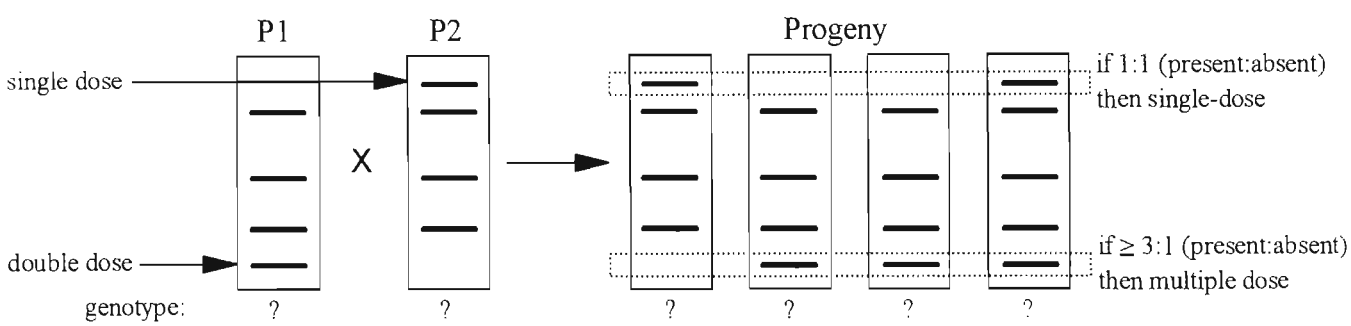


Figure 3.2 Comparison of DNA polymorphisms between diploid and polyploid parents and among their progeny (from Moore & Irvine, 1991).
(A) Diploid analysis. Parent and progeny genotypes are recognised as either homozygous (aa or AA) or heterozygous (Aa) by the number of hybridizing fragments. Linkage analysis is usually based on segregation of the genotypes, but it could be based on segregation of the single-dose fragments.
(B) Polyploid analysis. The genotype of polyploids is not known. If a polymorphic band segregates as 1:1 (present:absent) in either diploid or polyploid progeny populations, then that allele is a single dose. However, if the polymorphic band segregates as >3:1 (present:absent), then it is due to multiple doses in one of the parents. The progeny test for distinguishing segregation ratios of 1:1 from >3:1 will have a 98% confidence level on a progeny population of 75 individuals.

Using the single-dose marker approach, Da Silva *et al.* (1993) constructed a SDRF linkage map of *S. spontaneum* SES 208. Wu *et al.* (1992) suggested that a haploid population derived from a highly heterozygous plant would be the most efficient mapping population for SDRFs, since potentially every SDRF can be mapped whereas in the progeny of a cross, only the polymorphic ones (present in one parent and absent in the other) are usable. Thus, Da Silva and co-workers (1993) initially compared two mapping populations for the construction of the linkage map, i.e. a population of 80 haploid plants derived from anther culture of SES 208, and 90 plants derived from the cross between SES 208 and ADP 85-0068, a doubled haploid of SES 208. A disadvantage of using a haploid population obtained through anther culture is that selection of gametes can occur during the *in vitro* propagation due to the effects of compounds in the medium and other growing conditions. If the segregation distortion is great enough to influence the detection of SDRFs, the genetic map will be seriously biased. Da Silva and co-workers (1993) found a strong segregation distortion in the SES 208 haploids, and therefore used the cross population to construct the linkage map. The rationale for using that particular population was that, on average, half of the RFLP bands that represent single "alleles" present in SES 208 should be absent in the doubled haploid ADP 85-0068. Individual bands present in the SES 208 and absent in the doubled haploid were scored for presence or absence in the progeny. Those RFLP bands that conformed to a 1:1 segregation ratio were used for linkage analysis with MAPMAKER, resulting in a map of 216 markers.

Since only single-dose markers can be used for mapping in polyploids, a high number of polymorphic fragments are required to saturate the map with those that segregate 1:1. Sobral and Honeycutt (1993) optimized the RAPD reaction to produce a high output of these PCR-based markers in *S. spontaneum* and showed that single-dose RAPD fragments can be used in linkage analyses. This makes the RAPD method particularly suitable for genetic mapping in sugarcane. This same group then constructed a 208 marker linkage map of *S. spontaneum* SES 208 using single-dose RAPD markers (Al-Janabi *et al.*, 1993). Estimates of genome size and coverage of both the SDRF and RAPD studies were in close agreement, and based on the marker segregation data both concluded that this *S. spontaneum* representative is an autopolyploid.

Da Silva and co-workers (1995) integrated the RAPD map with the SDRF map of *S. spontaneum* SES 208 to form a single extensive molecular marker linkage map that will facilitate the application of DNA markers to sugarcane breeding. Since mapping output is higher with PCR-based markers than with RFLPs, and automation of marker acquisition is possible with PCR, PCR-based markers are more amenable for use in breeding programmes. However, information regarding homologous relationships among PCR markers obtained with the same primer is not available without Southern hybridization data, so SDRFs are preferred to obtain information about homology of linkage groups. Duplex and triplex loci can be mapped in autopolyploids (Da Silva *et al.*, 1995; Da Silva & Sobral, 1996), and this methodology was used to extend the *S. spontaneum* map (Da Silva *et al.*, 1995). Once the linkage relationships between the multiple-dose markers and single-dose markers were calculated, groups of pairing homologues were identified using multiple-dose markers as well as probes that detected multiple SDRFs. For example, linkage groups X and Y were considered homologous when X contained SDRFs from probes A and B, and Y also contained different SDRFs from the same probes A and B. Evidence of homology can also be obtained when a double-dose marker shows linkage to SDRFs from both X and Y. This approach organized the 64 linkage groups in the *S. spontaneum* map into eight chromosome homology groups, which agreed with the conclusion that SES 208 is an auto-octoploid (Da Silva *et al.*, 1995).

Al-Janabi and co-workers (1994) initiated the mapping of *S. officinarum* and *S. robustum* (the putative wild ancestor of *S. officinarum*) on the basis of a cross between these two species. A total of 83 single-dose RAPD polymorphisms were identified, and linkage analysis of these revealed seven linkage groups in the *S. officinarum* genome with 17 linked and 33 unlinked markers, and five linkage groups in the *S. robustum* genome defined by 12 linked markers and 21 unlinked markers. This data is limited in comparison to the 527 marker linkage map of *S. spontaneum* (Da Silva *et al.*, 1995). However, analysis of more progeny with a larger number of markers would provide a more detailed map.

D'Hont *et al.* (1994) initiated genome mapping in sugarcane varieties by analysing plants derived from the selfing of variety SP 701006, using isozymes and RFLPs. Linkage analysis

of single-dose markers placed 94 markers into 25 linkage groups. Eighteen of these involved *S. spontaneum* specific markers, thereby tentatively marking *S. spontaneum* chromosomes, while the remaining four involved *S. officinarum* specific markers. On the basis of probes in common (as described for the *S. spontaneum* map), the 25 linkage groups were assembled into eight tentative groups of pairing homologues.

Grivet *et al.* (1996) constructed a genetic map for sugarcane variety R570. Using the self progeny of this variety, linkage analysis of single dose markers generated from RFLP probes and one isozyme resulted in 408 markers being placed in 96 linkage groups. On the basis of common probes, these were tentatively assembled into 10 basic groups of pairing homologues. The origin of a selection of markers was investigated, leading to the identification of 80 *S. officinarum* and 66 *S. spontaneum* derived markers. Of the 96 linkage groups, 38 had two or more species-specific markers. Among these, 19 were *S. officinarum* specific, 15 were *S. spontaneum* specific, while the remaining four appeared to have a double origin. This provides evidence that, while recombination is possible between *S. spontaneum* and *S. officinarum* chromosomes, such events are likely to be rare (Grivet *et al.*, 1996). The occurrence of recombination between *S. officinarum* and *S. spontaneum* chromosomes was recently demonstrated for the first time using genomic *in situ* hybridization (D'Hont *et al.*, 1996). This technique makes it possible to distinguish the chromosomes contributed by these two species, as well as the chromosomes containing sequences from both species, and about 10% of the chromosomes of sugarcane variety R570 were identified as recombinant between *S. officinarum* and *S. spontaneum*. These authors do emphasize, however, that interspecific recombinations remain infrequent events in sugarcane crosses.

3.3 Applications of Genetic Markers to Crop Improvement

The various functions of genetic markers in crop improvement are discussed below in order to provide a broad overview of the general utility of markers. However, in the context of this study, emphasis will be placed on the use of markers for the identification of associations between markers and traits (both qualitative and quantitative), with the goal of marker-assisted selection in mind.

3.3.1 Estimation of Genetic Diversity and Varietal Identification

The improvement of crops depends upon the existence, nature and extent of the genetic variability available for manipulation. Information on these aspects of genetic variability will allow the design of strategies for maximum gain from breeding programmes, as well as ensuring that germplasm resources represent a wide range of genetic variability.

Genetic markers can be used to measure similarity (genetic distance) among individuals within natural or breeding populations. For example, Haley *et al.* (1994b) used RAPD markers to assess variability between and within gene pools of common bean, and found that variability followed a "three-tiered" pattern: variability between gene pools > between races of the same gene pool > within races. Genetic diversity has been examined in sugarcane. Harvey and Botha (1996) showed that RAPDs are suitable for the determination of genetic diversity in sugarcane varieties, and their results suggest that there is limited DNA diversity between the commercial varieties in comparison to the ancestral species *S. spontaneum* and *S. officinarum*. Similar results were obtained by Lu *et al.* (1994a) using RFLPs.

Genetic markers can also be used as a source of phylogenetic information. Orozco-Castillo *et al.* (1994) used RAPD markers to examine the distribution of genetic diversity of *Coffea* species, and found that dendrograms constructed from RAPD profiles were consistent with the known history and evolution of *Coffea arabica*. Vierling and Nguyen (1992) conducted a study to determine the extent of inter- and intra-specific genetic diversity of diploid wheat genotypes using RAPD markers, and were able to establish the relationships of genotypes within each species. Lu *et al.* (1994b) used RFLPs to analyse relationships among the ancestral species of sugarcane, and their results proved to be consistent with traditional phylogenetic hypotheses based on morphological and biochemical data.

A consequence of the general lack of diversity within the sugarcane varieties currently grown commercially is that breeders are trying to improve the traits of existing varieties through the use of a very limited genetic background. An understanding of the diversity and phylogenetic relationships within the *Saccharum* complex (the interbreeding group of sugarcane and its wild relatives) can be of primary importance to breeders, since recommendations can then be made

to introgress new germplasm into commercial varieties in an attempt to increase genetic diversity and obtain improved traits.

Markers can be used to "fingerprint" strains and varieties of agricultural importance, providing a tool for strain identification as well as a means for protecting breeders' rights to novel varieties. This has been illustrated in rice where matrices, generated from scoring RAPD bands as present or absent, were reproducible and amenable for identification of each single plant line (Wang *et al.*, 1994). Harvey and Botha (1996) assessed the PCR amplification of sugarcane DNA using microsatellite and telomere primers. The fingerprints generated by these primers for different varieties were highly polymorphic, and based on these results the suggestion was made that this methodology would be useful for the rapid and reliable identification of sugarcane varieties. Parental analysis can also be accomplished with genetic markers. Büscher *et al.* (1994), working with grapevine varieties, and Huckett and Botha (1995), working with sugarcane, reported that the presence of RAPD fragments in a particular variety that was not matched in either of the supposed parents could be used as evidence for invalid paternal identity.

3.3.2 Identification of Markers Linked to Important Major Genes for Marker-Assisted Selection

One of the immediate benefits of genome mapping to plant breeding is the use of genetic markers linked to genes to indirectly select for important qualitative traits. Such linked markers have been identified in a number of important crop species using linkage analyses, a few examples of which are listed below. Jin *et al.* (1993) were able to obtain morphological markers linked to two leaf rust resistance genes in barley. Ganal *et al.* (1995) used a high density RFLP map of tomato to identify a marker very closely linked to a gene conferring resistance to potato cyst nematode. Qi *et al.* (1996) identified a RAPD marker associated with a powdery mildew resistance gene in wheat.

The lagging behind of Mendelian genetics in sugarcane has been mentioned previously. Due to its genome complexity, qualitative segregation of traits seems to be exceptional. However, Daugrois and co-workers (1996) investigated inheritance of resistance to rust in sugarcane

using the self progeny of variety R570, the same population that was used to build a large-scale genetic map (Grivet *et al.*, 1996). A 3:1 distribution of resistance to susceptibility was observed in the progeny, which is the expected segregation ratio for a simplex dominant resistance gene. Linkage analysis using the genetic map revealed a highly significant association of rust resistance with a specific marker, and this was confirmed using a second sample of the self progeny of R570. This is the first documented report of monogenic inheritance for disease resistance in sugarcane.

In order to effectively identify tightly linked markers for several traits, the map in question should have a high level of saturation and the markers should be distributed throughout the genome. This can become a laborious and time-consuming process. However, other approaches have been established which are able to bypass the necessity of a genetic linkage map of the entire genome for identifying markers linked to traits.

The use of near-isogenic lines (NILs) is one such alternative (Young *et al.*, 1988). Most NILs have been developed by introgression, with the repeated backcrossing of a line carrying a gene of interest to a cultivated line chosen because of its otherwise favourable properties. After several generations of repeated backcrossing and selecting progeny based on the phenotype of the target gene, the genome of selected individuals will consist almost exclusively of that of the recurrent parent, with a small segment of foreign DNA flanking the target gene persisting from the donor parent. The product of introgression, therefore, is a pair of NILs which are identical except for the region near the target gene. The high probability that any polymorphism detected between a pair of NILs will be in the DNA surrounding the introduced gene, provides a powerful means of identifying markers linked to the trait of interest. Once markers have been identified, tight linkage of the marker to the gene can be verified by linkage analysis in a segregating F_2 population.

Using the NIL approach in tomato, Young *et al.* (1988) were able to identify two RFLP markers tightly linked to a gene conferring resistance to tobacco mosaic virus, while Ohmori *et al.* (1995) found five RAPD markers tightly linked to the tomato mosaic virus resistance gene, and Martin *et al.* (1991) identified three RAPD markers that were tightly linked to a

Pseudomonas resistance gene. The NIL approach for targeting genetic markers has been used in other crops as well. Diers *et al.* (1992) found RFLP markers linked to phytophthora resistance loci in soybean, Bai *et al.* (1995) identified two RAPD markers tightly linked to a gene for black root rot resistance in tobacco, and Young and Kelly (1996) obtained two tightly linked RAPD markers that appeared to flank an anthracnose resistance gene in common bean. Caetano-Anollés *et al.* (1993) were able to obtain markers tightly linked to the supernodulation locus in soybean using NILs and DNA amplification fingerprinting with DNA that had been predigested with restriction enzymes. In the absence of digestion, they failed to obtain polymorphisms between the NILs, thus illustrating how the identification of linked markers can be assisted without resorting to extensive screening of arbitrary primers.

Since NILs are often the product of co-ordinated plant breeding programmes, they are convenient to use for rapidly identifying markers linked to traits of interest. However, NILs are not available for all the important traits in every crop, and since it takes several generations of backcrossing to create NILs it is generally not feasible to develop them solely for the purpose of gene tagging for more than a few traits of interest (Tingey & del Tufo, 1993; Haley *et al.*, 1994a)

Another approach, described by Michelmore and co-workers (1991) and termed bulked segregant analysis (BSA), is based on similar principles but avoids the time consuming generation of NILs. The only prerequisite for BSA is the existence of a population resulting from a cross that segregates for the trait of interest. Individuals are sorted into two groups based on the expression of that trait, e.g. a resistant group and a susceptible group for a disease resistance trait, and the DNA from the individuals is bulked to form two pools. The two pools will differ at loci linked to the trait while the remaining genetic background will be randomized due to contributions of all the other unselected alleles from each of the individuals in the bulk. Thus, any polymorphisms between the two bulks are expected to be closely linked to the target locus as this is the only region in the DNA samples that differs significantly between the two bulks. Linkage of the markers to the trait of interest can be confirmed with conventional linkage analysis.

Michelmore *et al.* (1991) were able to identify three RAPD markers linked to a gene for resistance to downy mildew in lettuce using this approach. Other examples where BSA has been used include the identification of a RAPD marker linked to a scab resistance locus in apples (Yang & Krüger, 1994), two RAPD markers linked to the gene conferring resistance to pea enation mosaic virus in pea (Yu *et al.*, 1995), a RAPD marker linked to leaf rust resistance in barley (Poulsen *et al.*, 1995), and a RAPD marker linked to a gene conferring sunflower rust resistance in sunflower (Lawson *et al.*, 1996).

NILs and BSA can therefore be used to identify markers associated with a particular phenotype without screening entire mapping populations with every probe or primer. The entire population need only be tested when polymorphisms between the NILs or bulks are detected, in order to obtain proof of linkage before such markers can be used for marker-assisted selection. These markers can then be placed on existing maps.

All the examples listed above involve simply inherited, qualitative traits, i.e. traits under monogenic control. However, most agronomically important traits such as yield or growth rate are quantitative. They show a continuous distribution in a segregating population and are controlled by the joint action of several genes, with each gene contributing only a proportion of the expression of the trait. Quantitative traits are far more complex than qualitative, or monogenic traits and will be dealt with in the following section. However, it should be mentioned that markers linked to genetic loci underlying quantitative traits (known as quantitative trait loci, or QTLs) can be identified and are potentially useful in marker-assisted selection.

Marker-assisted selection (MAS) involves the use of genetic markers as tags for various traits. Usually, a great deal of time and effort is required to screen cultivars for desirable phenotypic characters. Selecting plants that express recessive alleles, genes that have only minor effects on plant phenotype, or genes whose expression is strongly modified by the environment can become limiting factors in plant improvement. This can be simplified if desirable genes could be selected, rather than selecting for the effects of their expression. Markers that are tightly linked to the desirable genes can allow the efficient selection of the genes, and therefore the

characters they control.

The use of markers to screen for disease resistance genes is a particularly good example of the utility of MAS (Tanksley *et al.*, 1989; Michelmore, 1995). Typically, clones in a breeding programme are screened for the presence of disease resistance genes by field trials under natural disease pressure, or by artificial inoculation in glasshouse or pot trials. However, simultaneous or sequential screening of plants with several different pathogens can be difficult or impractical. Quarantine restrictions can also pose a problem for screening for resistance to foreign pathogens. Variability in aggressiveness or availability of the pathogen under natural conditions, and the requirement to conduct some disease screens at particular locations, times of year, or stages of plant development are all detrimental to the efficiency of selecting resistant plants based on phenotype. In contrast, detecting resistance genes on the basis of their linkage to markers can be carried out in the absence of the pathogen and at the seedling stage, as soon as the plants are large enough to yield sufficient DNA. The screen is independent of any environmental conditions, and resistance to as many pathogens as required can be carried out simultaneously, provided that markers are available. It should be mentioned that, despite the advantages of MAS, markers should not totally substitute for a disease screen. Occasional testing with the pathogen will be necessary to ensure that the resistance gene is still present and effective.

Most studies to identify markers linked to specific traits are only carried out using a single cross. However, the general utility of MAS will depend on the expression of the markers in other genetic backgrounds. When a marker for a barley mild mosaic virus resistance gene, identified from doubled haploid lines from a single cross, was tested in a number of barley cultivars, it could detect resistant cultivars carrying the specific gene with complete accuracy (Ordon *et al.*, 1995). Nair *et al.* (1995) obtained two SCAR markers that were linked to a gall midge resistance gene in rice. These were detected from recombinant inbred lines derived from a single cross, however when individuals from a different cross were assayed, the resistance-specific band was amplified in three of the susceptible lines. This was probably due to recombination between the marker and the gene. Markers should ideally retain linkage with the gene and be clearly expressed and functional across a broad range of genetic backgrounds.

MAS is not likely to replace traditional selection methods, but will rather supplement them in the development of new and better crop varieties. While markers can be used to select plants with desirable genes from among the many seedlings in the initial stages of selection, the performance of these plants should be evaluated over a wide range of environmental conditions at a later stage in the field. The extent to which MAS will be used will depend on the relative cost and the expected return compared to conventional selection, and it will likely prove to be most useful for traits that are either expensive or difficult to evaluate by conventional methods.

3.3.3 Characterization of Quantitative Traits

As mentioned above, many important characters in plants are the consequence of the joint action of several genes. Such characters are described as quantitative or polygenic, and the genetic loci that are involved in their expression are referred to as quantitative trait loci (QTLs). The number of genes and the genetic bases of the observed variation make manipulations and improvements in quantitative traits slow and difficult. Many of these genes have effects so small that they are overshadowed by environmental effects and cannot be easily evaluated. However, markers and marker-based techniques have provided opportunities to gain insight into the complex genetic organization of these traits by allowing them to be resolved into their individual genetic components.

Paterson and co-workers (1988) showed that a high density linkage map of DNA markers can be used to detect, map and measure the effects of genes underlying polygenic traits using linkage analysis carried out by specially designed computer software such as MAPMAKER-QTL (Lander & Botstein, 1989; Paterson *et al.*, 1991). Thus, the number and effects of QTLs underlying a character can be investigated, and each QTL can be studied as a discrete entity, and its individual and interactive properties measured. The ability to detect a QTL using this approach is a function of the magnitude of the QTL's effect on the character, the size of the population being studied, and the density of markers across the genome. As the proportion of the total variance explained by a QTL decreases, the number of progeny which must be studied in order to detect the QTL increases (Paterson *et al.*, 1991). Thus, only QTLs with sufficiently large effects will be detected, and the number of QTLs reported for a specific trait must be considered as a minimum estimate since many QTLs with much smaller effects may

be present. In addition, an apparent QTL represented by a single marker may actually consist of two or more closely linked QTLs, each with small phenotypic effects in the same direction (Paterson *et al.*, 1988). This would cause a further underestimation of the number of loci involved.

Paterson *et al.* (1988) identified the most likely chromosomal positions, effects, and strength of evidence for individual QTLs underlying each of three traits in tomato, i.e. fruit weight, soluble solids concentration, and pH. Since then, studies have been conducted in several other crops involving the dissection of many different quantitative traits. To list a few examples, Bubeck *et al.* (1993) identified QTLs associated with gray leaf spot resistance in maize, while Schön *et al.* (1993) were able to characterize QTLs affecting resistance against European corn borer in maize. Leon *et al.* (1995) identified markers linked to QTLs affecting oil percentage in sunflower seed, Lebreton *et al.* (1995) identified QTLs for drought response in maize, and Redoña and Mackill (1996a) mapped QTLs for seedling vigour in rice using RFLPs.

Linkage analysis of moderate to saturated genetic maps of sugarcane will permit the direct investigation of the number and nature of QTLs affecting such characters as fibre and sucrose percentages and cane tonnage. This can allow the analysis of the individual and combined effects of specific genes in selected genetic backgrounds, which in turn will increase the efficiency of sugarcane breeding and selection programmes. An extension of the mapping study of *S. officinarum* and *S. robustum* (Al-Janabi *et al.*, 1994) resulted in the first published report of a QTL study conducted directly on a population derived from crossing complex polyploid species (Sills *et al.*, 1995). Phenotypic data from replicated progeny trials of the *S. officinarum* x *S. robustum* cross provided the opportunity to investigate for associations between the 83 single-dose markers on the map and various quantitative traits (number of stalks, number of tassled stalks, number of stalks with smut, and average diameter of two stalks). QTL analysis revealed between two to four markers significantly associated with the expression of each trait analysed. A broader sampling of the genome with additional markers and the use of a greater number of progeny may reveal additional QTLs for the traits. In addition, for broad use in marker-assisted selection, genetic markers need to be associated with traits across years and locations. However, this study serves as an important preliminary QTL

evaluation in complex polyploids.

BSA can also be used to quickly identify markers linked to quantitative traits (Michelmore *et al.*, 1991). The individuals at the extremes of the distribution of the variation (i.e. expressing high or low scores for the trait) are likely to differ at most of the loci controlling the trait. These individuals can be used to form the bulks, which will then be screened for polymorphisms. Examples of where this approach has been used to identify markers linked to QTLs include erucic acid level in *Brassica napus* (Jourdren *et al.*, 1996) and quantitative disease resistance in barley (Kutcher *et al.*, 1996). As with qualitative traits, linkage analysis of segregating markers in the whole population must be carried out to confirm linkage to the relevant traits.

BSA has great potential to identify markers in sugarcane since near-isogenic lines are absent in this crop, and the high ploidy level of sugarcane varieties confounds the conventional identification of markers. The potential use of BSA to identify markers linked to a quantitative trait in sugarcane has been investigated. Msomi and Botha (1994) applied BSA to a population segregating for fibre, and were able to identify eight RAPD polymorphisms between the high and low fibre groups. Six of these were identified as single-dose fragments, due to 1:1 segregation in the population. Potential linkage to the fibre trait was investigated by analysing the highest and lowest 20% of the individuals in the population for presence or absence of the fragments, and this showed strong correlations between the putative markers and the trait. Linkage of these markers to fibre still needs to be confirmed conventionally, and this will be facilitated by the single-dose nature of the markers. Linkage analysis will also allow the calculation of how much fibre variation can be ascribed to each of the putative markers. By testing such markers in other crosses segregating for the same trait, the effects of genetic background on the markers, and hence their usefulness for MAS, can be determined. The major limitation of applying BSA to sugarcane is the requirement for populations segregating for specific traits. The extended time period required to accurately evaluate the phenotypic traits of sugarcane clones has already been discussed, and this would restrict the efforts of a breeding programme whose aim it would be to rapidly obtain tightly linked markers.

While the mapping of QTLs with genetic markers has provided insight into the genetic organization of complex traits, these markers could also be used for MAS of QTLs in breeding programmes (Michelmore, 1995). MAS has considerable potential for selecting loci underlying quantitative traits, since selection for individual markers is far easier than selection for quantitative traits, which is impaired by genetic and environmental noise.

In several instances of quantitative traits, only a few loci (e.g. two or three) control the majority of the genetic variation. It has been suggested that these quantitative traits should be considered as oligogenic rather than polygenic (Michelmore, 1995), and these traits can be treated as though they were qualitative in terms of MAS (Dudley, 1993). With traits that are truly quantitative (controlled by a large number of loci), the probability of identifying marker loci for all the QTLs is low, particularly if some of the QTLs only have small effects. The most useful markers will be those that are linked to one or a few major genetic regions controlling the trait. However, expression of some traits may be controlled by several minor regions. The application of MAS to these traits is unclear since the question arises as to whether selection with markers will be as effective as selection based on phenotype (Dudley, 1993). The usefulness of MAS for quantitative traits will therefore have to be determined on a case-by-case basis. Factors that will determine the usefulness of QTL-marker associations for MAS include their prevalence in the germplasm, and genotype x environment interactions. For example, only some of the QTLs detected by Bubeck and co-workers (1993) for gray leaf spot resistance were present in all three maize populations tested, and the QTLs were inconsistent over the three environments. Some of the QTLs detected in tomato by Paterson *et al.* (1991) were detected in all three environments, while some were only identified in a single environment. Only some of the markers linked to erucic acid level in *Brassica napus* were detected when tested against a second cross (Jourdren *et al.*, 1996). As with qualitative traits, MAS of quantitative traits will only be feasible if it is reliable and less expensive or tedious than direct selection.

An unconventional but nevertheless very interesting approach for the marker-assisted selection of quantitative traits was recently described by Virk *et al.* (1996). Rather than using the progeny from a planned cross, they used 48 accessions (varieties) of rice that represented a

wide range of the available diversity. These were evaluated for six quantitative traits, and screened for RAPD polymorphisms. Associations between RAPD markers and the various traits were established using the multiple regression approach. Significant associations were found between RAPD markers and all six traits. The usefulness of this approach was then illustrated by performing multiple regression analysis on only 40 of the accessions. Regression parameters thus estimated were used to predict the quantitative variation in the remaining eight accessions. In each case, association of the quantitative variation with markers provided good predictions of the performance of the eight "unknown" accessions. The multiple regression analysis was also able identify which markers showed particularly strong associations with QTLs. For example, one of the markers was able to describe 49.8% of the variation for the trait "culm number". The apparent advantages of using diverse germplasm instead of single progenies for QTL identification are that QTLs that vary across a wide spectrum of biodiversity rather than just between two parental lines can be detected, and that QTLs for any quantitative trait can be studied in the same investigation rather than analysing separate populations that segregate for each of the traits (Virk *et al.*, 1996).

3.3.4 Introgression of Favourable Genes

Gene introgression is generally conducted by crossing a parent line with a specific desirable trait (the donor parent) with an otherwise agronomically superior cultivar (the recurrent parent), followed by repeated backcrossing to the recurrent parent and screening of the progeny for the target character. After several cycles, the donor genome (which may include several undesirable traits) will be diluted out and selected plants will be nearly identical to the recurrent parent with the exception of the portion of the genome containing the gene for the target character.

Genetic markers can be used to overcome major limitations of backcross breeding. Tanksley *et al.* (1989) showed that by selecting for markers that are tightly linked to the trait and selecting against markers situated throughout the rest of the donor genome, the reconstruction of the recurrent parent genotype can be accelerated by selecting offspring with the lowest amount of donor genome in every generation. A problem associated with backcross breeding is linkage drag: the transfer, not only of the gene of interest, but also of additional linked

genes which are not necessarily desirable. In traditional backcross breeding, linked segments usually remain large for many generations since there is no effective way to identify individuals with recombination between the trait of interest and linked regions. However, high density genetic maps can make it possible to directly select individuals that have experienced recombination near the gene of interest (Tanksley *et al.*, 1989). Since markers linked to QTLs can be identified, quantitative characters can also be manipulated in marker-based backcross selection (Tanksley *et al.*, 1989). Without this approach, it would be extremely difficult and time consuming to monitor the introgression of such genes in a breeding programme.

Since little or no additional effort is required to screen for multiple markers, the possibility exists to add many genes simultaneously to a cultivar. For example, various resistance genes, each producing resistance against different strains of a pathogen, are common in plants (e.g. Diers *et al.*, 1992). Similarly, various classes of genes, each with different mechanisms of resistance, exist (Michelmore, 1995). The accumulation (or pyramiding) of several resistance genes into one cultivar can be achieved in only a few generations with the use of markers, whereas with conventional breeding this is nearly impossible since the action of different resistance genes in combination cannot be distinguished (Michelmore, 1995; Winter & Kahl, 1995). This strategy will result in a more durable resistance than that afforded by single resistance genes (Kelly, 1995), and will allow breeders to respond more quickly to market demands as well as unexpected environmental pressures, such as the appearance of new pathogens.

3.3.5 Utilization of Exotic Germplasm

Wild species contain a high level of genetic variation. Since cultivated varieties are generally derived from only a few accessions of ancestral or wild species and have been subjected to repetitive selection, the range of variation for any given character can be much greater in exotic germplasm than among cultivated varieties. Exotic germplasm thus has great potential value for improvement of crop varieties.

Backcross breeding provides the most straightforward method of transferring genes from exotic germplasm into cultivated varieties. However, as mentioned above, traditional backcross

breeding is time-consuming and relatively ineffective in transferring complex genetic characters. The undesirable effects of linkage drag can also be particularly pronounced when breeding with exotic germplasm. Genetic markers can be used not only to locate genes controlling desirable characters in wild species, but also to select for individuals with little unwanted donor DNA (Tanksley *et al.*, 1989).

The narrow genetic base of sugarcane has already been described, and mention has been made how markers can provide information on the diversity and phylogenetic relationships between sugarcane and its wild relatives. Based on this information, recommendations can be made to introgress new germplasm into commercial varieties to improve their performance with regard to specific traits. Markers also have the potential to follow the introgression of new genes from wild germplasm into cultivated sugarcane.

3.3.6 Map-Based Gene Cloning

Conventionally, a gene is cloned on the basis of its known product or sequence. However, the products of many important plant genes are unknown. A good example is disease resistance genes, where many such genes have been identified by classical genetics, but the mechanisms by which they act and their gene products are largely unknown. If such genes could be isolated and cloned, greater understanding about their mode of action could be obtained, and it might also be possible to introduce them, via transformation, into other crop species for which resistance genes are not available. Map-based gene cloning is an approach for isolating genes without knowledge of the gene product or sequence.

An absolute prerequisite for map-based cloning of genes is the availability of tightly linked markers flanking the gene of interest. Restriction fragments between the markers can be cloned into suitable vectors such as yeast artificial chromosomes (YACs). The linked markers serve as the starting point, and allow one to walk by overlapping clones along the chromosome from the marker to the gene of interest (Winter & Kahl, 1995). In order to identify cloned segments of DNA that contain the gene of interest, candidate clones can be transformed into suitable heterologous or homologous host plants as a functional assay for gene activity. Several genes have been successfully cloned using the map-based approach (e.g. Michelmore,

1995).

3.3.7 Determining Gene Homologies Among Crop Species

Despite their apparent independence, a number of crop species share taxonomic relationships. For example, maize, sorghum and sugarcane belong to the family *Gramineae*, while the family *Solanaceae* includes members such as potato, tomato and pepper. Genetic markers can be used to determine the extent of conservation of chromosome content and gene orders among related species (Tanksley *et al.*, 1989). The use of a common set of probes to map different species has been termed comparative mapping.

To illustrate this, RFLP maps based on a common set of probes have been used to determine gene homology in tomato, potato and pepper (Bonierbale *et al.*, 1988; Tanksley *et al.*, 1988). These comparative linkage maps show a high degree of linkage conservation in tomato and potato, while the map of pepper is highly rearranged in comparison. This suggests that the substitution of chromosomes or chromosome segments between tomato and potato might be feasible. One of the outcomes of comparative mapping is the ability to investigate homolgy between genes from otherwise independent pools of resources. This will allow theories concerning the inheritance and mechanisms of gene action proposed for one species to be suggested for the other. For example, Ganal *et al.* (1995) found a marker very closely linked to a gene conferring resistance to potato cyst nematode in tomato. The close proximity of the marker makes this gene a prime target for map-based gene cloning. While the potato cyst nematode has no economic significance in tomato, it is an important pathogen in potato. By cloning the resistance gene, its functionality in potato can be investigated by transformation, and due to the almost complete linearity of the tomato and potato genomes, one can determine whether similar resistance genes occur at comparable map positions in potato.

Comparison between the genomes of maize and sugarcane has been accomplished by using maize genomic probes to hybridize to and map the sugarcane genome (Hulbert *et al.*, 1990; D'Hont *et al.*, 1994). These studies have shown that a large degree of similarity exists between the maps of maize and sugarcane. The chromosome organization of sorghum also appears to be conserved in sugarcane (Grivet *et al.*, 1994). The comparative mapping of

sugarcane with related diploids, such as maize and sorghum, can provide guidance for locating specific important genes in the sugarcane genome. This also has implications for map-based gene cloning of these genes, provided that closely linked markers can be found.

3.4 Challenges Confronting Plant Breeders and the Prospects of Marker Technologies

Conventional plant breeding methods have led to great success in crop improvement. To achieve continued success and meet the changing demands for superior varieties, a range of problems need to be overcome. These include limited availability of technical assistance with plant maintenance and manipulation, limited availability of greenhouse and field space, and the need to grow and evaluate many generations before a variety can be released. While genetic markers will certainly not solve all the problems associated with plant breeding, they have the potential to directly and indirectly facilitate the improvement of crops in various stages of breeding and selection programmes.

Marker-based techniques can provide valuable information on the nature and extent of genetic variability in populations. This information, together with a greater understanding of the nature of complex traits, will benefit breeding efforts at the planning stage by allowing efficient strategies to be designed for maximum gain. Perhaps the greatest potential to improve crops through the use of markers lies in marker-assisted selection, and this includes the use of markers to speed the introgression of desirable traits into cultivated germplasm. The use of markers for population screening and genotype evaluation will result in large savings of time, space, effort, and therefore money, by directly selecting desirable genotypes at the seedling stage. Marker-assisted selection is not likely to replace conventional breeding methods, but instead can supplement and streamline them, allowing more efficient use of time, space and resources.

Breeding programmes typically deal with thousands of plants, and for markers to be effective in these programmes, the automation of as many of the rate-limiting steps as possible is required. The first steps towards the routine use of automation have already been taken. For

example, Rafalski *et al.* (1994) devised a machine to routinely extract DNA. The preparation of DNA amplification-based markers such as RAPDs and microsatellites can be automated (Sobral & Honeycutt, 1993). Size separation of the amplification products can limit throughput, but assays which do not require gel separations have been developed (Gu *et al.*, 1995). The initial developmental costs of automating marker technologies will be large. However, automation has the potential to dramatically improve the efficiency of large scale genotyping of large populations, thereby allowing a large return on investment capital.

CHAPTER 4

MATERIALS AND METHODS

4.1 Selection of Varieties

The sugarcane selection programme at SASEX takes about 14 years from the time various crosses are made to the commercial release of a variety. During this time the varieties in the programme undergo a series of trials in which their phenotypes are evaluated. Only those varieties with satisfactory performances for all the evaluated traits will be released. However, some varieties may, in general, have good characteristics but are not released due to the presence of one or two undesirable traits. These varieties are set aside for use as parents in the breeding programme. Since they have passed through various stages of the selection programme, information on their phenotypes will be available.

Fifty varieties of sugarcane were selected for this study (listed below). These are either commercial varieties, or varieties that were considered unsuitable for commercial cultivation but are nevertheless used as parents in the sugarcane breeding programme at SASEX. As mentioned above, the benefit of using these varieties is that their phenotypes have been well characterized.

The following varieties were selected:

1.	74M659	2.	75E247
3.	76H333	4.	76M1101
5.	77L1720	6.	77W1241
7.	78F909	8.	79F1043
9.	79H181	10.	79L181
11.	79L1294	12.	80L432
13.	80W1459	14.	81L1308
15.	81W133	16.	81W447

17.	82F2907	18.	83F448
19.	J59/3	20.	N8
21.	N11	22.	N13
23.	N14	24.	N16
25.	N17	26.	N18
27.	N19	28.	N20
29.	N21	30.	N22
31.	N23	32.	N24
33.	N25	34.	NCo293
35.	NCo376	36.	77F790
37.	77L1143	38.	79M955
39.	CP57/614	40.	CB38/22
41.	N52/219	42.	75E1293
43.	76M1566	44.	N12
45.	84F2753	46.	77F637
47.	75L1157	48.	NM214
49.	80F2147	50.	75L1463

Several clones of *Saccharum officinarum* and *Saccharum spontaneum* were also used in the analysis in an attempt to trace the phenotypic characters to one of these two progenitors of the modern varieties. Since there is very little genetic variability within *S. officinarum* and much higher levels of variability within *S. spontaneum* (Lu *et al.*, 1994b), three clones of the former and six clones of the latter were chosen to represent the two "parent" species:

S. officinarum:	Black Cheribon	
	Black Tanna	
	Badila	
S. spontaneum:	Pasoerean	Coimbatore
	Mauritius	Tabongo
	Kloet	Nigeria

4.2 Choice of Phenotypes

With a crop such as sugarcane, the most useful phenotypes or traits to use in a study aimed at finding genetic markers associated with those traits, would be those with economic or agronomic advantages. For this study, levels of resistance to the sugarcane borer *Eldana saccharina*, sugarcane mosaic virus (SCMV), and the fungal pathogen smut (*Ustilago scitaminea*) were chosen, since ratings for these traits are readily available for all the varieties listed above. Due to the nature of the analysis, traits such as fibre content and sucrose measurements can be added at a later stage in order to identify useful markers associated with additional traits.

Since the phenotype of sugarcane is influenced to such a large extent by the environment, commercial varieties as well as those used in the breeding programme continue to be evaluated in trials in order for their phenotypic ratings to be updated and improved. Unless otherwise indicated, the resistance ratings that were available in 1995 were used in the analyses aimed at finding associations between markers and the traits. However, an updated list of eldana resistance ratings became available in 1996. Since many of the ratings varied slightly from the original list, the modified ratings were included in the multiple regression analysis for comparative purposes. This has been indicated in the relevant sections.

The ratings used to describe resistance to pests or pathogens in sugarcane range from 1 to 9, where 1 signifies a level of resistance such that signs of infection or damage are either very limited or absent altogether, and 9 signifies extreme susceptibility with signs of severe infection or damage. The manner in which the different resistance ratings are assigned to the sugarcane varieties differ. The disease ratings are assigned in a very simple manner. For SCMV, the percentage of infected stalks in a trial are converted to a value from 1 to 9. Similarly, for smut the number of whips in a particular trial are converted into a 1 to 9 rating. (In a case of smut infection, the top-most leaf of the stalk and the leaf roll dies and turns a black-brown colour. This is referred to as a "whip".) For the eldana ratings, a more complex route is taken. The total length of damage, total number of internodes with damage, total number of larvae and pupae, and total weight of larvae and pupae are determined for each stalk. Various statistical calculations are performed on these data which result in the measures

being combined into a single score from 1 to 9.

4.3 DNA Extraction

It has been found that RAPD polymorphisms can be detected in the banding profiles of individual sugarcane plants (stools) belonging to a single variety (Harvey *et al.*, 1994). However, by extracting DNA from stalks obtained from at least three separate stools per variety, these individual-specific differences can be eliminated.

DNA extraction was performed according to a protocol modified from Honeycutt *et. al.* (1992). The outer leaf sheaths were removed from the stalks of three individual stools and the young leaf roll tissue was cut into slices (6-10g in total), pooled, and homogenized for two minutes (Ultra-Turrax T-25) in 40ml ice-cold homogenization buffer (50mM Tris(hydroxymethyl)-aminomethane (Tris), pH 8.0, 5mM ethylene diamine tetra-acetic acid (EDTA), 0.5mM spermidine, 1% (w/v) polyethylene glycol (8000), 0.1% (v/v) 2-mercaptoethanol, 0.35M sucrose). The homogenate was filtered through four layers of mutton cloth and the filtrate centrifuged at 5000g for 20 minutes in a rotor precooled to 4°C (Beckman Avanti J-25I Centrifuge). The supernatant was discarded and the cell pellet resuspended in 10ml cold wash buffer (50mM Tris, pH 8.0, 25mM EDTA, 0.5mM spermidine, 0.1% (v/v) 2-mercaptoethanol, 0.35M sucrose) and placed on ice. To this suspension were added, sequentially, NaCl to a final concentration of 0.7M, sodium dodecyl sulphate (SDS) to a final concentration of 0.7% (w/v), and cetyltrimethylammonium bromide (CTAB) to a final concentration of 0.9% (w/v), with gentle agitation after each addition. The solutions were incubated at 60°C for 30 - 45 minutes and then allowed to cool at room temperature for 15 minutes. Proteins were removed by gentle mixing with an equal volume of chloroform:isoamyl alcohol (24:1), centrifugation of the emulsion at 3500g for 10 minutes at 4°C, collection of the aqueous phase and repetition of the extraction. After the final collection of the aqueous phase, DNA was precipitated by the addition of an equal volume of isopropanol, spooled out of solution with a glass hook, drained, and released into 1ml TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0). The DNA was allowed to dissolve at 4°C for 2 - 5 days, whereafter the main stock solutions were stored at -20°C.

4.4 DNA Assessment

4.4.1 Spectrophotometric Analysis

The concentration of each of the DNA extracts was determined with the use of an ultraviolet-visible recording spectrophotometer (Beckman, DU 7500). Ten-fold and twenty-fold dilutions of each of the DNA solutions were prepared using TE buffer, and the absorbance values (A) of these were recorded at 260nm and 280nm, respectively.

The purity of the DNA extracts was determined according to the following equation, with pure DNA having a value of 1.8 (Sambrook *et al.*, 1989):

$$\text{Purity} = A_{260nm} / A_{280nm}$$

The concentrations of the DNA extracts were determined with the following equation, using the constant $1A_{260nm}$ unit of double-stranded DNA = $50ng.\mu l^{-1}$ (Sambrook *et al.*, 1989):

$$\text{DNA concentration} = A_{260nm} \times \text{dilution factor} \times 50ng.\mu l^{-1}$$

4.4.2 Agarose Gel Electrophoresis

The spectrophotometric concentrations were confirmed using electrophoresis through agarose minigels. Agarose gels (0.8% [w/v]) were prepared with 0.5X TBE buffer (45mM Tris, 45mM boric acid, 1mM EDTA, pH 8.0), incorporating ethidium bromide at a concentration of $1\mu g.ml^{-1}$. Volumes representing 500ng DNA, calculated from the spectrophotometric concentrations, were mixed with gel loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 30% (v/v) glycerol) and loaded into the gel wells. Samples of 250ng, 500ng and 750ng uncut Lambda (λ) DNA (Boehringer Mannheim) were included for comparison. Electrophoresis was conducted at $5.6V.cm^{-1}$ in 0.5X TBE buffer, and this was followed by destaining in the electrophoresis buffer for 20 minutes with constant shaking.

The DNA in the gels was visualized under ultraviolet light, and the intensity of fluorescence of the genomic DNA was compared to that of the λ standards to allow discrepancies in the spectrophotometer readings to be corrected. In addition, the gels provided a visual measure of the purity and integrity of the DNA.

4.5 RAPD Amplification of the Genomic DNA

4.5.1 Preparation of Template DNA Solutions

The stock solutions of DNA were stored at -20°C. Intermediate solutions of the DNA were prepared at a concentration of 60ng.μl^{-1} using TE buffer. These solutions were stored at 4°C for up to three months. Working solutions were prepared at a concentration of 3ng.μl^{-1} by diluting the intermediate solutions 1:20 with TE buffer. The working solutions were stored at 4°C for a maximum of one month before being discarded, due to the reduced stability of DNA at low concentrations. The intermediate DNA solutions (60ng.μl^{-1}) were used as a means to avoid the frequent defrosting of the stock solutions, since repetitive freeze/thawing of DNA can cause it to become sheared.

4.5.2 Primers

The primers used in the PCR reactions were obtained from Operon Technologies (Alameda, California). These are random decamer primers with a G-C content of 60-70%. Working solutions of the primers were prepared by diluting each of the primers in ultra-pure, sterile water to a concentration of 6μM. These were stored at -20°C.

The primers were named according to the example OD05, where 'O' is obtained from Operon Technologies, 'D' specifies the series, and '05' indicates the primer number in the particular series. The following selection of 41 primers were screened for their ability to detect polymorphisms in the sugarcane DNA: OA08, OD01-OD20, and OF01-OF20.

4.5.3 PCR Protocol

The RAPD protocol used in this study had been previously optimized for the amplification of sugarcane genomic DNA (Sobral & Honeycutt, 1993).

The final volume of the PCR reaction mixtures was 21μl, and these contained 10mM Tris-HCl (pH 8.3), 10mM KCl, 4mM $MgCl_2$, 1.4μg acetylated bovine serum albumin, 0.2μM primer, 0.1mM of each of dATP, dTTP, dCTP, dGTP (Boehringer Mannheim), 1.4U AmpliTaq Stoffel fragment (Perkin-Elmer) and 21ng template DNA. Reaction mixtures were overlaid with 30μl mineral oil prior to thermal cycling. PCR amplification was carried out in a Hybaid

OmniGene Temperature Cycler using the following cycling profile: 1 cycle of 3 minutes at 94°C, 1 minute at 35°C and 2 minutes at 72°C with a 2.4°C.s^{-1} ramp; 40 cycles of 1 minute at 94°C, 1 minute at 35°C and 2 minutes at 72°C with a 2.4°C.s^{-1} ramp; and 1 cycle of 7 minutes at 72°C.

The DNA from the three *S. officinarum* clones and the DNA from the six *S. spontaneum* clones were pooled, respectively, to form two bulks representative of the two progenitors (or "parents") of modern sugarcane varieties. Initially, the DNA from the 50 varieties and the two "parent" bulks was amplified with a selection of four random primers to test the suitability of the DNA to RAPD analysis. Once this was confirmed, the DNA samples were amplified using the 41 random primers. Amplification of the DNA samples with each primer was done at least twice to test the reproducibility of the results. If a primer produced inconsistent banding profiles, it was excluded from the analysis.

Each set of RAPD reactions included a negative control. This consisted of the PCR reaction mixture with ultra-pure, sterile water instead of DNA. This was included to show whether amplification occurred in the absence of sugarcane DNA.

4.6 Resolution of the Amplified DNA

Electrophoresis through horizontal agarose gels was used to separate the amplified DNA fragments. Two percent (w/v) agarose gels (300mm x 200mm x 7mm) were prepared with 0.5X TBE buffer. Gel loading buffer (2μl) was added to each 21μl amplification sample, and 20μl of each sample was loaded into the wells of the agarose gels. Electrophoresis of the amplified DNA was conducted at 5.6V.cm^{-1} in 0.5X TBE buffer. The molecular weight markers used for determining the sizes of the amplified fragments were λ DNA digested with EcoRI and HindIII, and pBR322 digested with HaeIII (the λ DNA, plasmid and restriction enzymes were obtained from Boehringer Mannheim). Aliquots of each of these were loaded into the gel wells alongside the DNA samples prior to electrophoresis.

Following electrophoresis, the gels were stained in a 1μg.ml^{-1} solution of ethidium bromide

in electrophoresis buffer for 30 to 45 minutes, followed by a period of destaining in electrophoresis buffer for 20 to 30 minutes, both with constant shaking. The amplified fragments were visualized under ultraviolet light and the banding patterns recorded by photographing the gels with a Cohu Solid State Camera and a Sony Video Graphic Printer (UP-860CE).

4.7 Analysis of the RAPD Data

The RAPD banding profiles that proved to be consistent across two separate PCR amplifications were scored for polymorphisms. Initially, two methods of scoring were compared to determine which provided the best results for the objectives of this study. The amplification profiles obtained from two arbitrarily chosen primers (OD10 and OD15) were subjected to these two scoring methods, and the results were compared.

4.7.1 Visual Scoring of the RAPD Profiles

The gel photographs were scored visually by noting the presence or absence of specific bands across all the varieties. Only clear, distinct polymorphic bands were considered. To eliminate subjectivity, each set of profiles was scored on two independent occasions and the results compared for inconsistencies.

4.7.2 Automated Scoring of the RAPD Profiles

The efficiency of automated scoring was tested in an attempt to devise a scoring system as free from subjectivity as possible. An image documentation system (Screen Machine II, Fast Electronic GmbH, 1993) was used to scan the gel photographs in order to electronically capture the images of the gels. The gel images were then imported into an image analysis software package for gel applications (WinCam 2.1, Cybertech, 1994). This software was used to scan each of the lanes in the gel image and detect and mark the bands. The images were scored using the "automatic contrast enhancement" either on or off, and the "background subtraction" function either on or off. The automatic contrast enhancement option adjusts the contrast of the picture for optimal representation, while the background subtraction function subtracts a slowly changing background from the image, leaving sharp signals untouched.

These two functions were used in an attempt to enhance the images of the gels for optimal analysis.

4.8 Estimation of the Sizes of the Polymorphic Fragments

Screen Machine II was used to scan the photographs of the gels with polymorphic fragments, and the captured images were then imported into WinCam image analysis software. This package has an "Apply Standards" function which allows the determination of the properties of detected bands based on those of defined standards. By scanning the gel lanes containing the molecular weight markers, their bands were marked and the sizes of each were defined. The subsequent marking of "unknown" bands in the RAPD profiles resulted in their respective values being calculated, based on the molecular weight markers.

Two molecular weight markers were used, i.e. λ DNA digested with EcoRI and HindIII, which results in a series of bands ranging from 21 226 to 564 base pairs (bp), and pBR322 digested with HaeIII, which results in a series of bands ranging from 587 to 123bp. Consequently, larger fragments were estimated using the λ marker while smaller fragments were estimated using the pBR322 marker.

4.9 Associations Between RAPD Polymorphisms and Phenotypes

Two alternate methods for identifying associations between RAPD polymorphisms and specific traits (in this case, resistance to various pests or pathogens) were investigated and compared. The first involved a simple correlation approach which attempted to determine whether the presence of any polymorphisms implied the existence of a particular phenotypic state. The second approach utilized multiple regression analysis in order to determine whether polymorphisms could be used to predict the actual resistance ratings of the varieties.

4.9.1 Correlation Analysis Between Polymorphisms and Phenotypes

It was decided that for a polymorphism to be considered useful in a marker-assisted selection programme, the group of varieties with the particular polymorphism should contain greater

than 50% of the desirable (e.g. eldana resistant) varieties and less than 50% of the undesirable (e.g. eldana susceptible) varieties. Naturally, the greater the percentage of resistant varieties and the less the percentage of susceptible varieties in the group with the polymorphism, the more useful the polymorphism is likely to be. The ideal polymorphism will describe a group of varieties consisting of 100% of the resistant varieties and 0% of the susceptible varieties. Since this is fairly unlikely, particularly when dealing with quantitative traits, it may be possible to find groups of polymorphisms that, together, will account for as many resistant varieties as possible and as few susceptible varieties as possible. A group of polymorphisms may correspond to QTLs, with each locus accounting for different components of the phenotypic variation. Since the resistance ratings of the sugarcane varieties ranged from 1 to 9, varieties were considered to be resistant if they possessed a rating of 1, 2, or 3, and susceptible if they possessed a rating of 7, 8, or 9. Varieties with a rating of 4, 5, or 6 were classified as intermediate.

The first step of the correlation analysis involved identifying all the polymorphisms that grouped together more than half of the resistant individuals with less than half of the susceptible individuals for eldana resistance, SCMV resistance and smut resistance, respectively. Since this can become a tedious process when dealing with a large number of polymorphisms and more than one trait, use was made of a simple computer program called "RAPD-Retrieve" (D. Vorster, 1996, pers. comm.[1]). This program is able to automatically assess each polymorphism/trait combination, and lists those polymorphisms that group together the specified percentage of resistant or susceptible individuals. From the polymorphisms selected by "RAPD-Retrieve", the best three were chosen for each trait, i.e. those polymorphisms with the highest ratio between the percentage of resistant individuals and the percentage of susceptible individuals. The number of intermediate varieties that were included by the polymorphisms were disregarded.

The next step involved identifying additional potentially useful polymorphisms that, when added to those found in the first step, increased the percentage of resistant individuals selected

---

[1]Darren Vorster, Biotechnology Department, SASEX, Private Bag X02, Mt Edgecombe, 4300

while maintaining the percentage of susceptible individuals as low as possible. "RAPD-Retrieve" was also used to identify polymorphisms that could be used for this purpose. This program has a stringency variable (x), such that the output of the program will list all the polymorphisms whereby $\geq x\%$ of the varieties with the polymorphism will be resistant. By setting a suitably high stringency value "RAPD-Retrieve" can specify polymorphisms that, when used together with those identified in the first stage, may be able to maximize the number of resistant varieties accounted for while keeping the susceptible varieties to a minimum.

4.9.2 Multiple Regression Analysis

Multiple regression is a statistical method of predicting one variable from one or more other variables. There are a number of statistics software packages available that can carry out multiple regression analysis. In the Biotechnology Department, SASEX, multiple regression has been performed on near-infrared (NIR) values of sugarcane stalk surface wax extracts to quantify the contribution of wax components towards resistance to eldana in sugarcane (Rutherford & van Staden, 1996). Although the software used for this purpose was specifically designed for use with NIR spectroscopy values (InfraAlyzer Quantitative Calibration, Bran+Luebbe Analysing Technologies, 1987), it can be applied to other types of data as well.

For the multiple regression, each trait was treated as the dependent variable and the various RAPD polymorphisms as independent variables. Conventionally, the polymorphisms would be scored as 1 for presence and 0 for absence (Virk *et al.*, 1996). However, since the InfraAlyzer software does not allow values of 0 for the independent variables, values of 0.00001 were used to specify absence while 1 was used in the standard way to specify presence. The analysis was based on the model:

$$Y = B_0 + B_1X_1 + B_2X_2 + \ldots B_pX_p$$

which related the variation in the dependent variable (Y, representing the resistance ratings) to a function of the set of independent variables (X_1 to X_p, representing the polymorphisms). The B_1 to B_p terms are the regression coefficients, calculated by regression analysis, that specify the relationships between Y and X. B_0 is the equation intercept and p is the number of

polymorphisms.

Stepwise regression was performed on the polymorphism data. This approach begins with the X variable most highly correlated with Y. The partial correlations of all remaining X variables to Y are calculated with the previously selected variable fixed in the model, and the variable with the highest partial correlation is added to the model. Additional variables are then sequentially added in the same manner. However, each time a new variable is added to the model, each variable already in the model can be removed and replaced by any other variable that will improve the overall fit of the model.

The InfraAlyzer software can only work with 20 independent variables at a time. Since a total of 54 reproducible polymorphisms were obtained, these were arbitrarily divided into three groups of 18 and a separate regression was performed on each group. Using the approach outlined above, the best six polymorphisms with the highest r^2 were identified from each group and these were then combined to form a final group of 18 polymorphisms. Since this was carried out independently for each trait, three separate groups of the best 18 polymorphisms were obtained for eldana resistance, SCMV resistance and smut resistance, respectively. Using these final groups and starting with the best single polymorphism model (i.e. the polymorphism most highly correlated with the specific trait), models with progressively higher numbers of polymorphisms were assessed in order to obtain a significant fit of the polymorphism data to the trait values. As additional variables (polymorphisms) are added to a regression model, the correlation coefficient, and therefore r^2, and the F-ratio of the regression change accordingly. A general relationship between these parameters is shown in Figure 4.1. The F-ratio is a statistical method of judging the goodness of fit of a regression equation to the data, thus the most significant model will be that in which the F value is maximized.

The best model for each trait, determined according to the r^2 and F-ratio values of the regression, was used in a "leave out one at a time" approach in order to test the predictive potential of the model. Multiple regression analysis was applied to 49 of the 50 varieties at a time, with the prediction of the fiftieth. This allowed for the fitting of 50 models for each

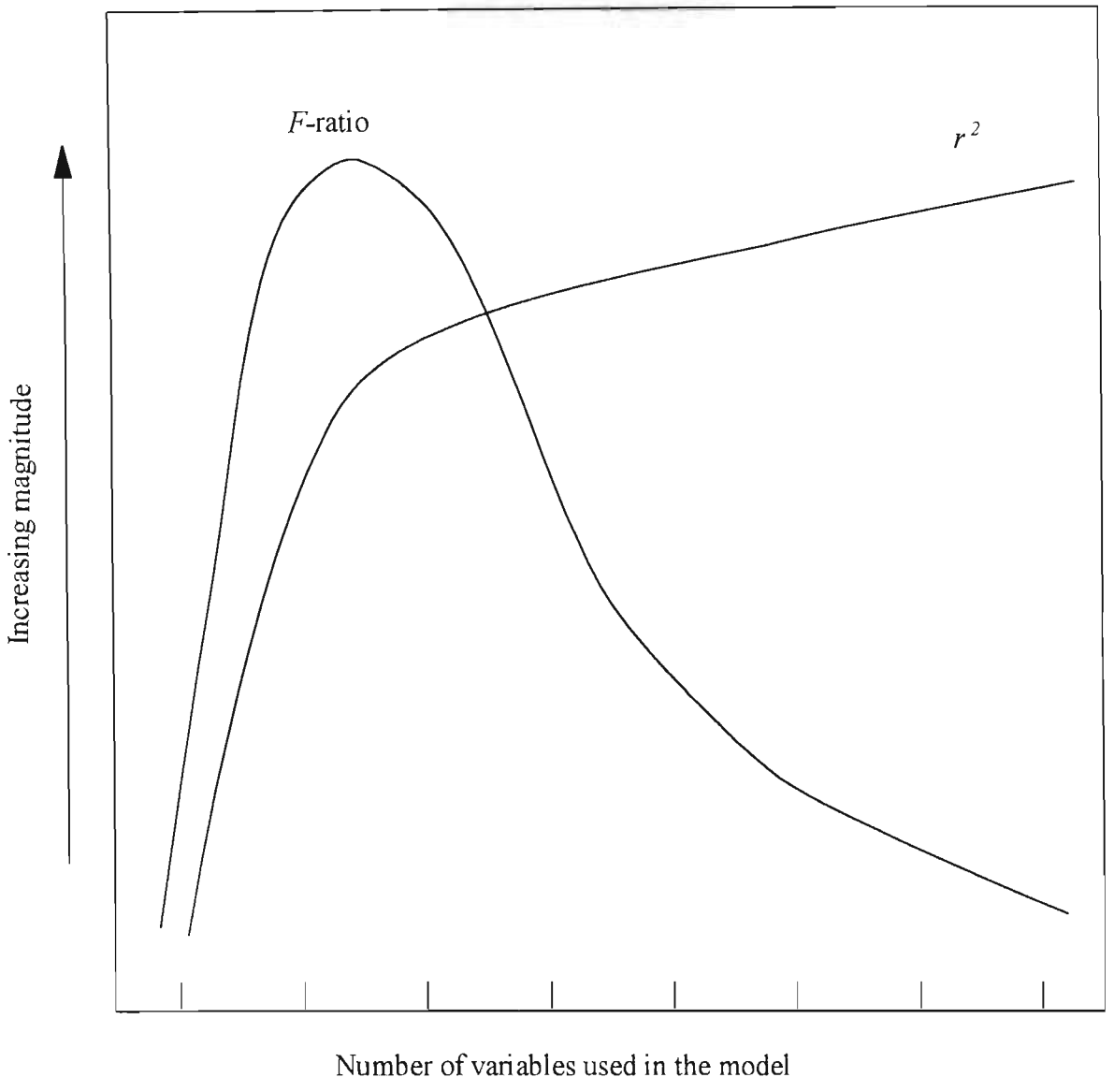


Figure 4.1 The general relationship between the correlation coefficient and the statistical significance (*F*-ratio) of the regression, as additional independent variables get added to the model

trait, and is the equivalent of using the chosen polymorphisms in a regression model calibrated from the 50 varieties to predict 50 unknown varieties. An indication of the actual performance of the selected polymorphisms was obtained by performing regression analysis on the actual ratings and predicted values for the 50 varieties. The actual ratings were treated as the independent variable with the predicted ratings as the dependent variable, and regression analysis was performed using the "Data Regression" option of the Lotus 1-2-3 computer package (Release 5, Lotus Development Corporation, 1994).

CHAPTER 5

RESULTS

5.1 Preparation of Template DNA

The concentrations of the genomic DNA extracts from the sugarcane varieties were determined with the use of a spectrophotometer, followed by the confirmation of the concentrations with agarose gel electrophoresis and comparison with known standards. The concentrations ranged from 315ng.μl^{-1} to 1128ng.μl^{-1}, with an average concentration of 572ng.μl^{-1}. Since 1000μl of DNA extract was produced for each variety and only 21ng of template DNA was required for each RAPD reaction, the quantity of DNA obtained was by far sufficient to fulfil the requirements of multiple RAPD reactions.

Figure 5.1 shows an example of an agarose gel that was used for the confirmation of DNA concentrations. As can be seen from this figure, the confirmation step is necessary since the spectrophotometer does not reliably determine the concentrations. However, the spectrophotometric concentrations provide a guide as to how much DNA to load onto the agarose gels; comparison of the genomic DNA with the standards then allows the spectrophotometric concentrations to be corrected, if necessary.

One of the main advantages of the RAPD technique as opposed to RFLPs is that the former only requires minute amounts of template DNA which need not be ultra-pure in terms of protein contamination (Waugh & Powell, 1992). The purities of the DNA extracts, as determined by the A_{260}/A_{280} ratios, ranged between 1.56 and 1.79 with an average purity of 1.69, showing that the DNA extracts were, in general, reasonably free of contaminants and of a sufficiently high quality for efficient RAPD amplification.

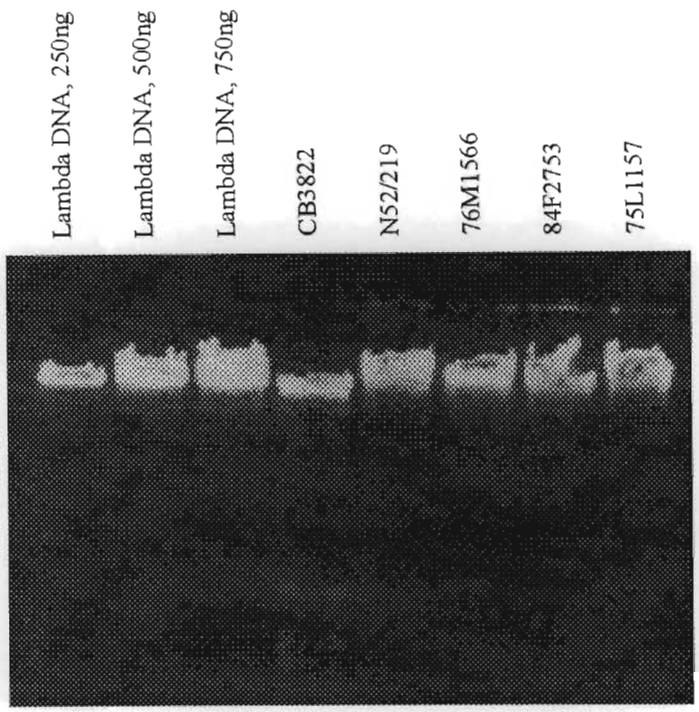


Figure 5.1 An agarose gel used for the confirmation of genomic DNA concentrations. The first three lanes contain the known Lambda DNA standards. Lanes 4-8 contain the genomic DNA of a selection of the sugarcane varieties. Volumes representing 500ng, calculated from the spectrophotometer readings of the DNA, were loaded onto the gel. The fluorescent intensity of the DNA samples as well as the size of the bands can be compared with those of the Lambda standards to allow discrepancies in the spectrophotometer readings to be corrected. The following estimates of the quantities loaded onto the gel were made: lane 4, 250ng; lane 5, 550ng; lane 6, 450ng; lane 7, 400ng; lane 8, 400ng. This provides an indication of the variation in the accuracy of the spectrophotometer readings.

The DNA samples on the gel are high molecular weight with no breakdown, or shearing. Breakdown of the DNA would be visualised as the smearing of the DNA bands due to the presence of a wide range of smaller fragments. The two smaller bands visible in some of the lanes, particularly lane 4, indicate the presence of RNA. The RNA, however, is unstable, degrades easily, and does not interfere with the RAPD reactions.

The integrity of the DNA extracts can be determined from the agarose gels used to confirm the concentrations, with excessive shearing of the DNA being seen as the smearing of the DNA in the gels. As can be seen from the examples shown in Figure 5.1, the genomic DNA extracts from the sugarcane varieties were of a high molecular weight with no breakdown.

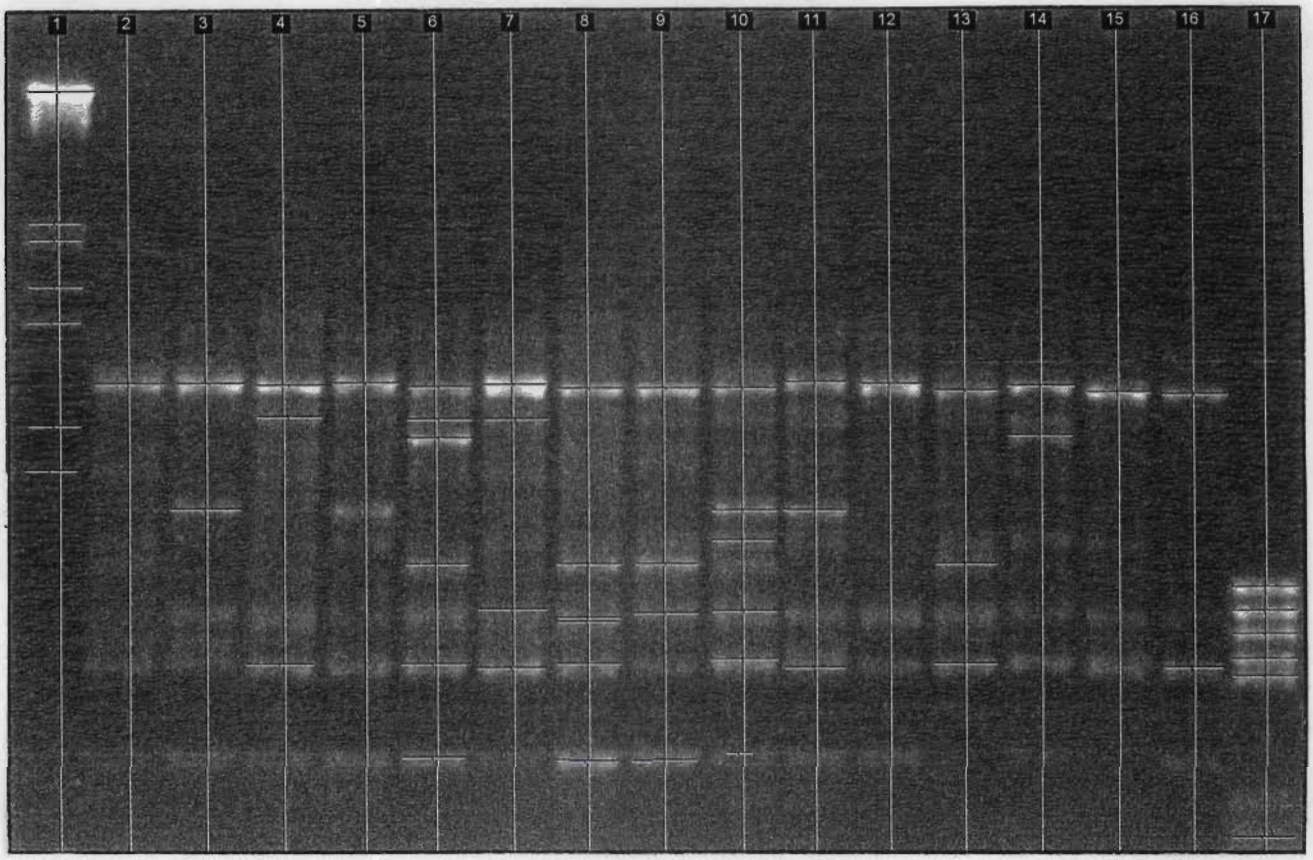
5.2 Comparison of Different Methods of Scoring RAPD Profiles

Visual scoring of the RAPD profiles from gel photographs was compared with automated scoring using WinCam image analysis software in an attempt to devise a scoring system as free from subjectivity as possible. The automated scoring produced unsatisfactory results, since certain bands were not detected by the software (Figure 5.2 A, B). The automatic contrast enhancement and the background subtraction function of WinCam were then used in an attempt to enhance the images of the gels, however neither of these options were able to improve the detection of bands (Figure 5.2 C, D). By adjusting the levels of automatic contrast enhancement and background subtraction, it may have been possible to obtain an image that was suitably enhanced for optimal band detection. However, the degree by which these levels would have to be adjusted would depend on the clarity and definition of each separate gel image to be scored. This would introduce a certain amount of subjectivity to the scoring system.

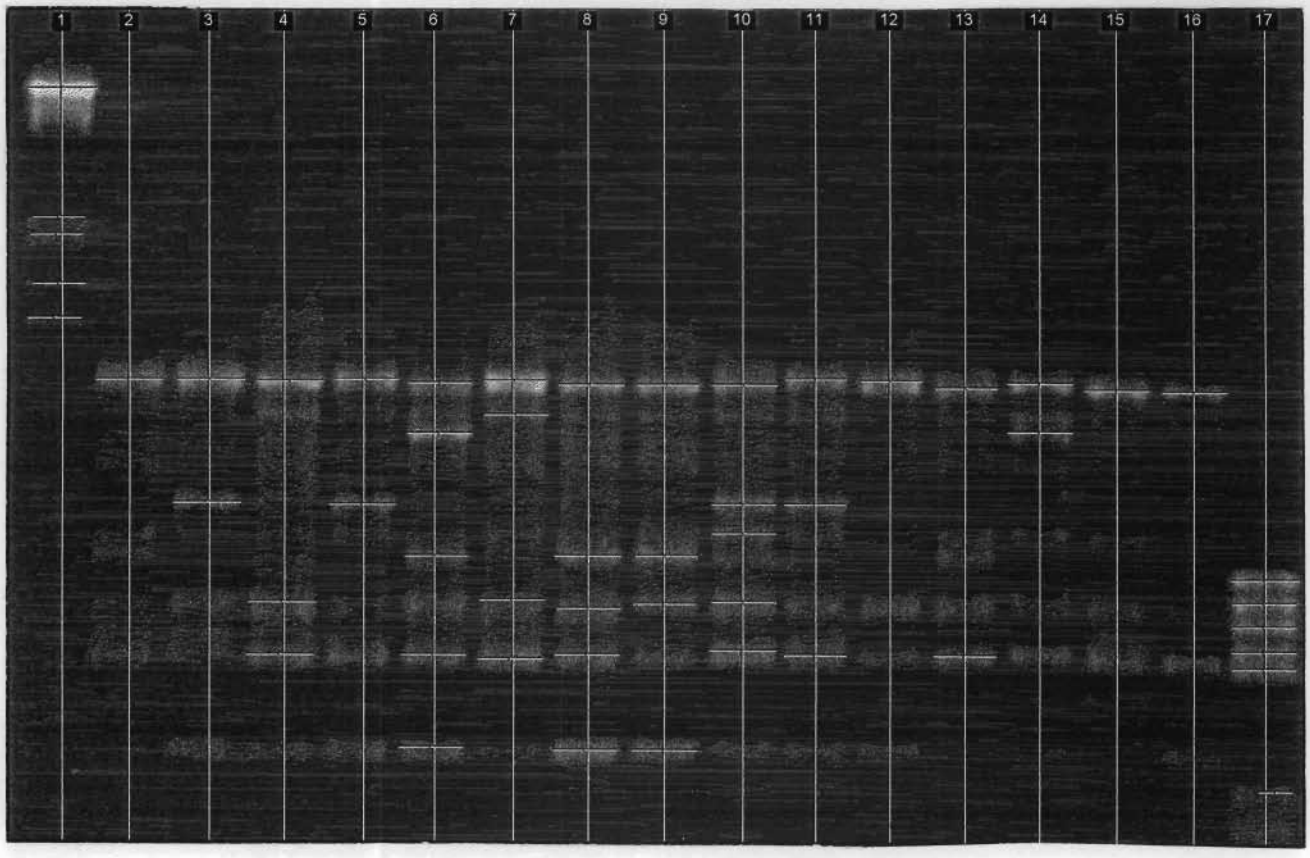
The RAPD profiles were scored visually on two independent occasions and the results compared for inconsistencies. No discrepancies were obtained between independent analyses of the same gel. It was therefore decided that visual analysis of the gel photographs would be the most effective method to use for scoring the RAPD profiles for polymorphisms. Only clear, distinct polymorphisms were considered; bands that appeared fuzzy as well as bands that were obscured by the co-migration of bands of similar size were disregarded. Examples of distinct, scorable polymorphisms are shown in Figure 5.3. As a contrast, the gel illustrated in this figure also shows the type of bands that were excluded from the analysis.

Overleaf: **Figure 5.2** An example of automated scoring of RAPD profiles using WinCam image analysis software

C


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

D


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

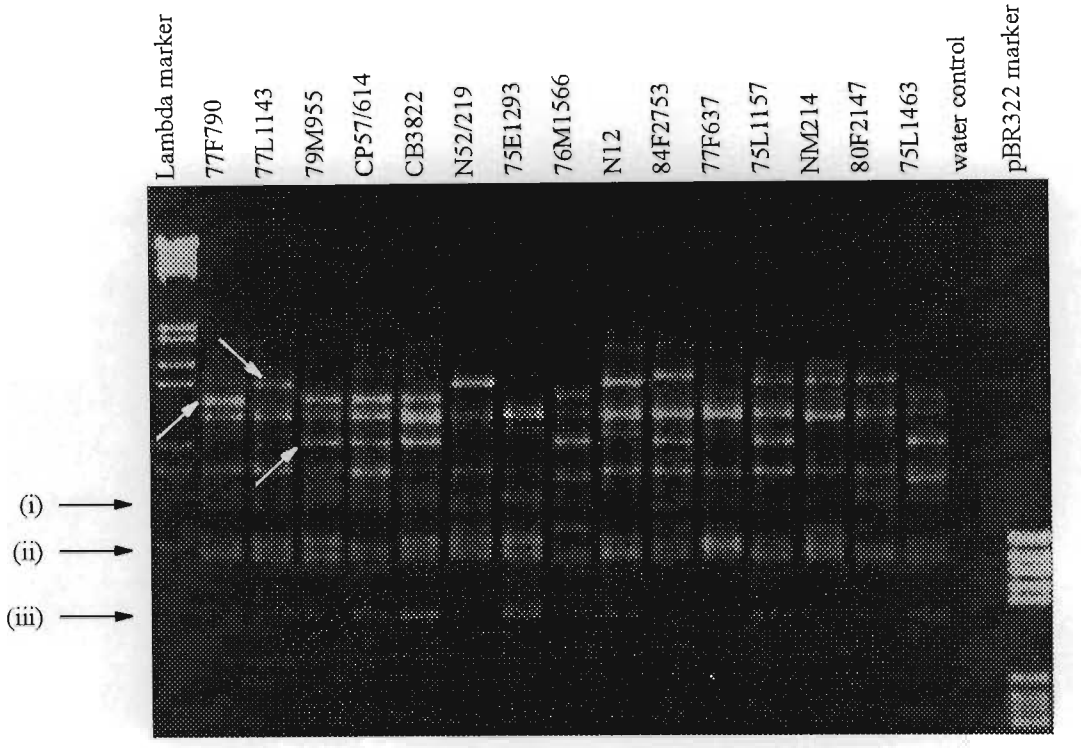


Figure 5.3 RAPD amplification profiles showing examples of the types of fragments that were considered for visual scoring. These profiles have been produced from primer OA08, and the varieties from which they have been amplified are indicated above their respective lanes. The white arrows show typical examples of clear, distinct polymorphisms that are easily scored. The black arrows indicate the types of polymorphisms that were disregarded.

(i) These fragments appear to be polymorphic, however they are too faint for their presence or absence to be accurately determined.

(ii) There appears to be three fragments of very similar size at this position. These would not be scored visually due to their close proximity and fuzzy appearance.

(iii) The intensity of this fragment varies across the different varieties, and where it appears to be absent the fragment is actually present, but very faint. This fragment would thus not be considered as a distinct polymorphism.

5.3 RAPD Amplification of the Sugarcane DNA

A total of 41 random primers were screened for their ability to detect reliable polymorphisms in the sugarcane genomic DNA. The results are summarized in Table 5.1. RAPD amplification of the DNA samples with each primer was done at least twice to test the reproducibility of the amplified loci, and this resulted in eight of the primers being excluded from further analysis due to the generation of inconsistent banding profiles. The remaining 33 primers amplified a total of 382 loci with an average of 11.6 loci per primer. Using the visual scoring approach as described above, 54 distinct, reproducible polymorphic loci were detected from 26 of the primers (average of 2.1 polymorphisms per primer). These fragments ranged in size from 169 to 1645 base pairs.

Table 5.1 Output of the RAPD amplification of sugarcane genomic DNA

Total number of primers screened	41
Number of primers producing reliable profiles	33
Total number of loci amplified from the reliable primers	382
Average number of loci per primer	11.6
Number of primers producing scorable polymorphisms	26
Number of polymorphic loci scored[a]	54

[a] Since only clear, distinct polymorphisms were considered, this figure should not be interpreted as being an indication of the extent of genetic variation in sugarcane varieties.

The number of loci and the frequency of polymorphisms varied between the different RAPD primers used to amplify the DNA of the varieties. For example, Figure 5.4 shows the amplification profiles of a selection of varieties obtained from primer OD15. A relatively large number of fragments were amplified (between 16 to 18) of which a high proportion (about 35%) are distinct, easily scorable polymorphisms.

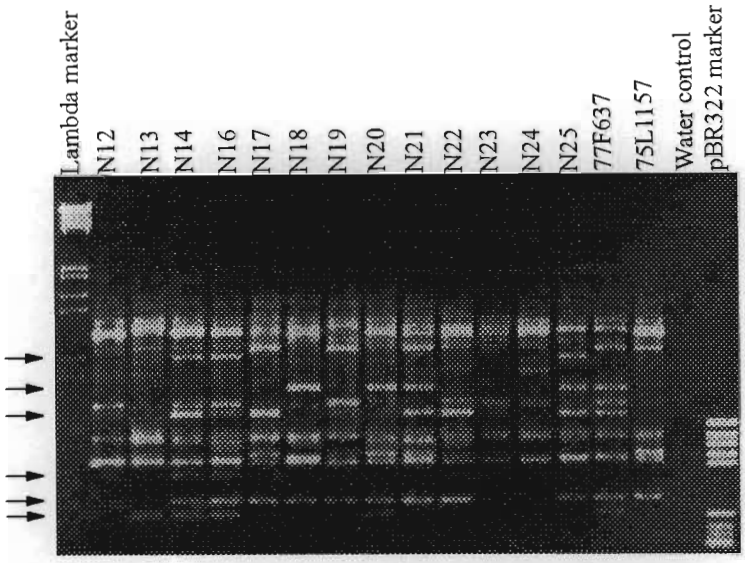


Figure 5.4 RAPD profiles of a selection of varieties amplified with primer OD15. Arrows indicate easily scorable polymorphisms.

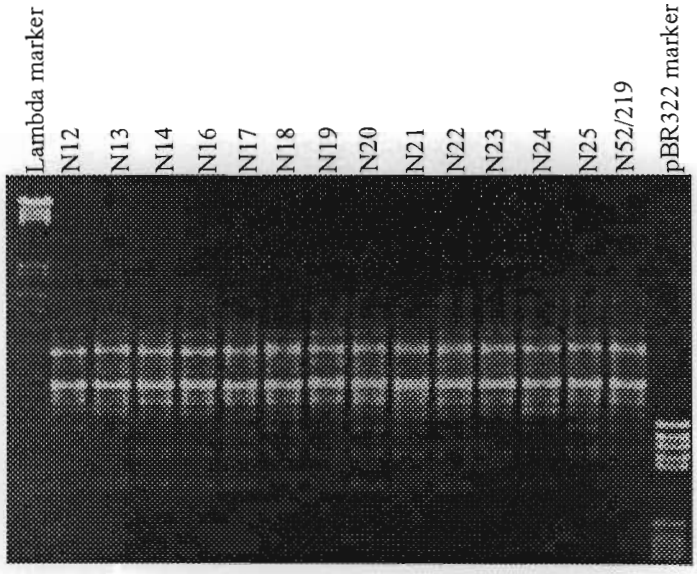


Figure 5.5 RAPD profiles of a selection of varieties amplified with primer OD08

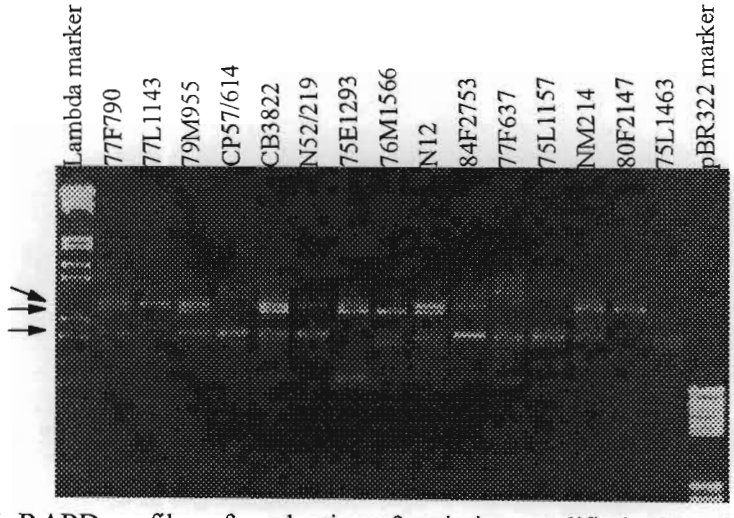


Figure 5.6 RAPD profiles of a selection of varieties amplified with primer OD17. Arrows indicate easily scorable polymorphisms.

In contrast to the above example, certain primers amplified fragments that showed very little or no variation between the different varieties. This indicates that the loci targeted by the particular primers are conserved across all the varieties. An example of this is shown in Figure 5.5, where the amplification profiles of the varieties produced by primer OD08 were identical.

While the primers used in both the examples mentioned above resulted in the amplification of several loci, some primers amplified comparatively few loci. This is illustrated in Figure 5.6, which shows the amplification profiles of a selection of varieties produced by primer OD17. About six loci were amplified, and only three of these were present in more than half of the varieties. It should be mentioned, however, that despite the low number of amplified loci, some of these are distinctly polymorphic (as indicated in the figure), and were consequently used in the analysis.

The inclusion of a negative control in the RAPD reactions was able to show that all fragments amplified during the reactions originated from the sugarcane DNA. This control consisted of the RAPD reaction mixture with ultra-pure, sterile water instead of DNA. Electrophoresis of the control along with the amplification products from the sugarcane DNA showed that no contaminating fragments were present in the amplified samples. For example, no fragments were detected in the water control lane in Figure 5.4.

5.4 Estimation of the Sizes of the Polymorphic Fragments

The sizes of the polymorphic fragments were determined with the "Apply Standards" function of the WinCam software. This allows the determination of the properties of detected bands based on those of defined standards. The molecular weight markers used for this purpose were λ DNA digested with EcoRI and HindIII, and pBR322 digested with HaeIII. These two markers are shown in Figure 5.7, along with the sizes of each of the fragments. Due to the respective sizes of the marker fragments, larger polymorphic fragments were estimated using the λ marker while smaller fragments were estimated using the pBR322 marker. Figure 5.8 shows how the sizes of three polymorphic fragments have been defined based on the λ marker.

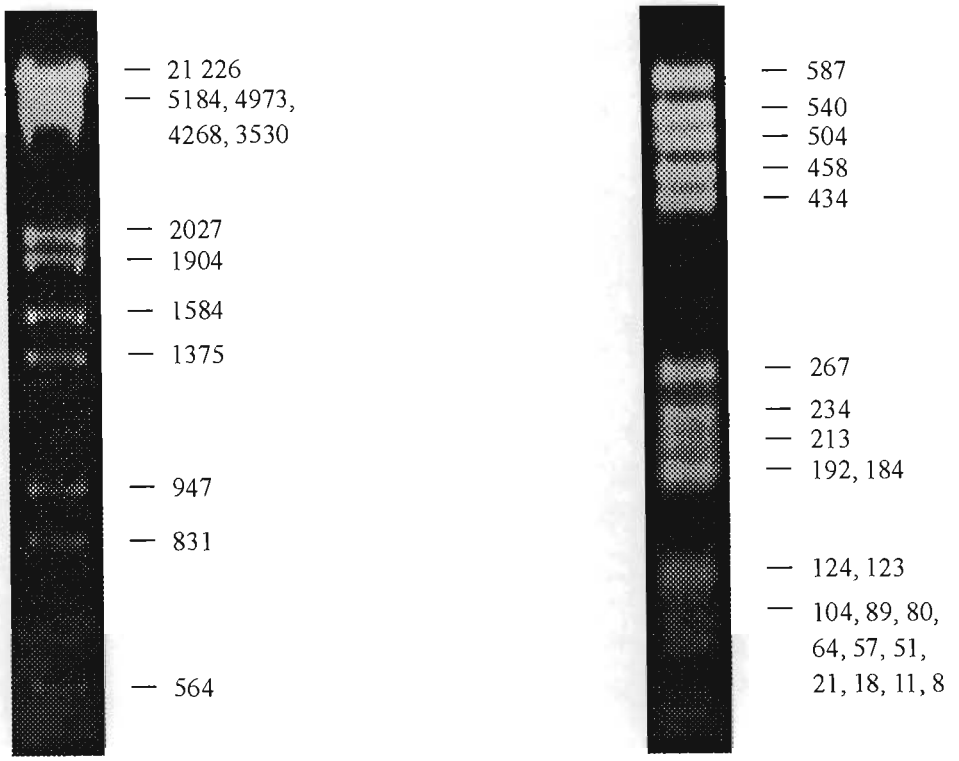
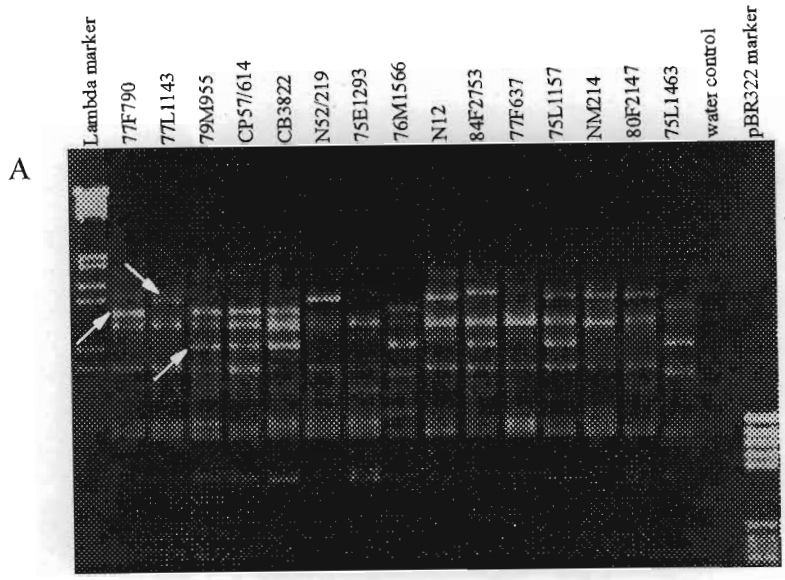


Figure 5.7 Molecular weight markers used to estimate the sizes of the RAPD fragments. The sizes of each of the marker bands have been indicated in base pairs (bp).
(A) Lambda DNA digested with EcoRI and HindIII. The largest fragments (21 226bp - 3 530bp) do not separate on the 2% (w/v) agarose gels used for the RAPD analysis.
(B) pBR322 digested with HaeIII. The fragments of 104bp and less are too small for clear resolution on 2% (w/v) agarose gels.



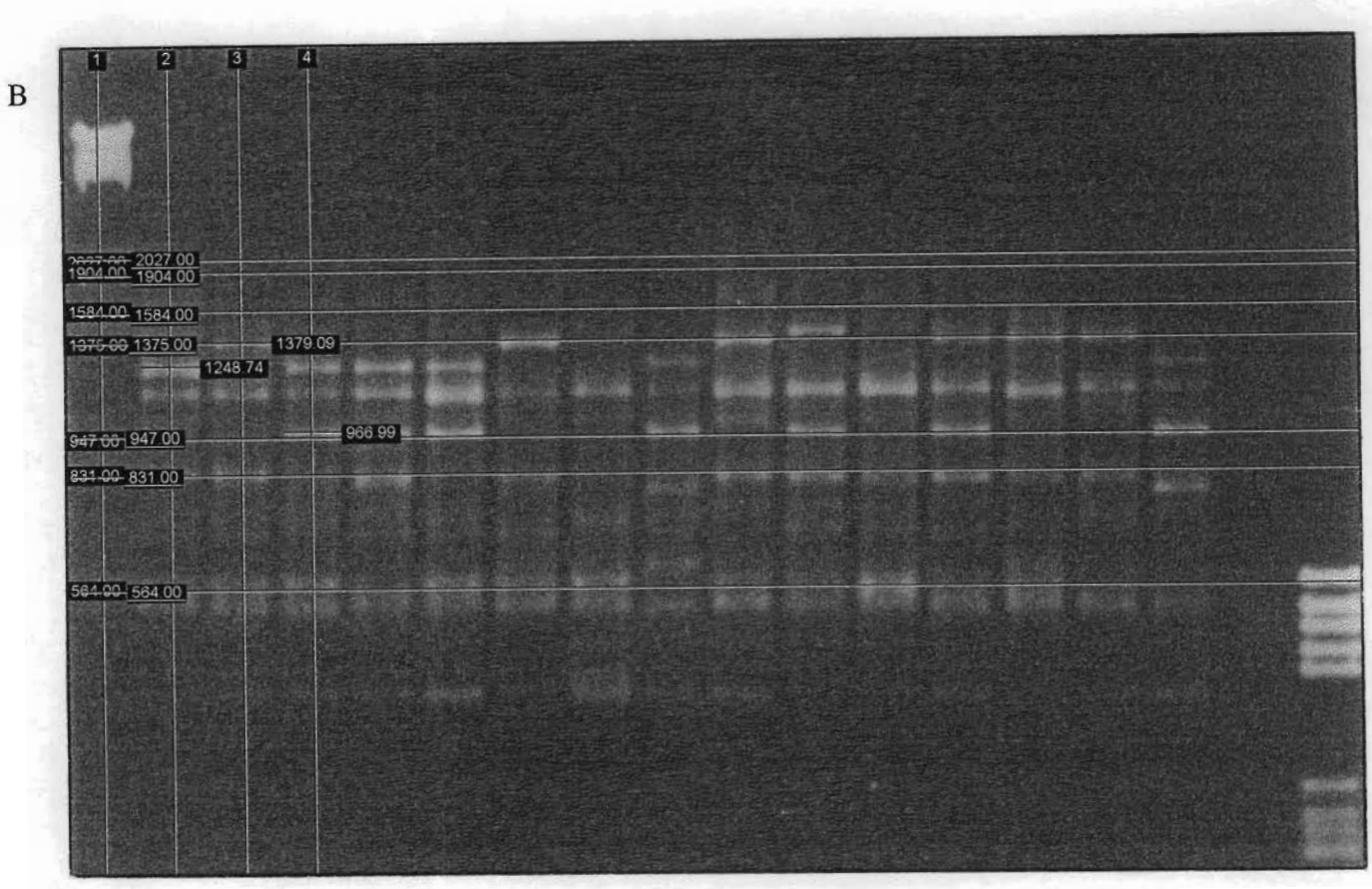


Figure 5.8 Estimation of the sizes of polymorphic fragments using WinCam image analysis software.

(A) Amplification profiles of a selection of varieties produced by primer OA08. The arrows indicate the polymorphisms that were used to illustrate the estimation of fragment size.

(B) WinCam output. The lane with the appropriate molecular weight marker is scanned and the sizes of the fragments defined. The sizes of all the fragments in an "unknown" lane can be calculated by marking the entire lane. Alternatively, as was done in the example illustrated above, single fragments can be marked for size estimation.

Polymorphic fragments were named according to the primer which amplified them, followed by the size (in base pairs) of the fragment, as a subscript. Consequently, the polymorphisms indicated in Figure 5.8 were named as follows: $OA08_{1379}$, $OA08_{1249}$ and $OA08_{967}$.

5.5 Origin of the Polymorphisms With Respect to *S. officinarum* and *S. spontaneum*

The amplification of the bulked *S. officinarum* and *S. spontaneum* DNA along with the DNA from the 50 varieties allowed the origin of the different polymorphisms to be traced to one of the two main progenitors of the modern varieties. Table 5.2 shows which of the polymorphisms were present in the *S. officinarum* DNA bulk, and which were present in the *S. spontaneum* DNA bulk. Certain polymorphisms were present in both the *S. officinarum* and *S. spontaneum* bulks, while others were absent in both; these are listed in Table 5.2 as well.

Table 5.2 Origins of the polymorphisms amplified from the 50 sugarcane varieties, as determined by the presence of the polymorphic fragments in the bulked DNA samples from *S. officinarum* and *S. spontaneum*, respectively[a]

***S. officinarum*-specific polymorphisms**	***S. spontaneum*-specific polymorphisms**	**Polymorphisms present in *S. officinarum* and *S. spontaneum***	**Polymorphisms absent in *S. officinarum* and *S. spontaneum***
$OD04_{629}$	$OD03_{984}$	$OD15_{305}$	$OD03_{512}$
$OD05_{921}$	$OD03_{553}$	$OD17_{1109}$	$OD04_{899}$
$OD06_{1455}$	$OD04_{806}$	$OF12_{1081}$	$OD06_{1645}$
$OD06_{354}$	$OD06_{1053}$		$OD10_{566}$
$OD09_{979}$	$OD15_{641}$		$OD10_{400}$
$OD09_{537}$	$OF06_{618}$		$OD14_{1316}$
$OD10_{1192}$	$OF09_{725}$		$OD14_{926}$
$OD10_{783}$	$OF10_{737}$		$OD15_{767}$
$OD11_{1085}$			$OD15_{270}$
$OD14_{356}$			$OA08_{1249}$
$OD17_{1058}$			$OA08_{967}$
$OD17_{863}$			$OF01_{722}$
$OD19_{343}$			$OF01_{653}$
$OD20_{933}$			$OF02_{584}$
$OD20_{506}$			$OF08_{755}$
$OA08_{1379}$			$OF10_{1508}$
$OF04_{610}$			$OF11_{635}$
$OF05_{585}$			$OF12_{750}$
$OF08_{1032}$			$OF12_{169}$
$OF20_{850}$			$OF14_{801}$
			$OF15_{1112}$
			$OF15_{654}$
			$OF15_{312}$

[a] Three clones were used to make up the *S. officinarum* bulk, while six clones were used for the *S. spontaneum* bulk. These are listed in Section 4.1 (p. 51).

5.6 Associations Between RAPD Polymorphisms and Phenotypes

Two alternate approaches were compared to determine the relative efficiency of each at identifying associations between the RAPD polymorphisms and the three traits (resistance to eldana, resistance to SCMV and resistance to smut).

5.6.1 Correlation Analysis Between Polymorphisms and Phenotypes

The polymorphisms that grouped together greater than 50% of the resistant individuals (resistance rating of 1, 2 or 3) with less than 50% of the susceptible individuals (resistance rating of 7, 8 or 9) for each trait were identified. From these, the best three for each trait were selected on the basis of the highest ratio between the percentage of resistant individuals and the percentage of susceptible individuals. These are presented in Table 5.3, along with their apparent origins according to Table 5.2.

Table 5.3 Single polymorphisms grouping the highest percentage resistant individuals with the lowest percentage susceptible individuals for the three traits

Polymorphic fragment		% Resistant varieties	% Susceptible varieties
Eldana resistance			
$OF04_{610}$ (-)[a]	(o)[b]	77.8	23.1
$OF12_{1081}$ (-)	(o/s)	77.8	23.1
$OF14_{801}$ (+)	(x)	100	30.8
SCMV resistance			
$OD15_{270}$ (-)	(x)	57.1	27.3
$OD17_{1109}$ (+)	(o/s)	66.7	36.4
$OA08_{1249}$ (-)	(x)	61.9	36.4
Smut resistance			
$OF01_{653}$ (-)	(x)	60.0	16.7
$OD17_{1109}$ (+)	(o/s)	70.0	33.3
$OA08_{1379}$ (-)	(o)	50.0	25.0

[a] (+) and (-) indicate whether the varieties were grouped together according to the presence or the absence of the polymorphic fragment, respectively.

[b] Origin of the polymorphic fragment, from Table 5.2. (o): *S. officinarum*, (o/s): *S. officinarum* and *S. spontaneum*, (x): neither *S. officinarum* nor *S. spontaneum*.

Additional potentially useful polymorphisms were then identified that would increase the percent resistant varieties selected while maintaining the percent susceptible varieties selected as low as possible. This was achieved using the stringency variable (x) of "RAPD-Retrieve", which results in a list of all the polymorphisms such that $\geq x\%$ of the varieties with the polymorphism will be resistant. Initially, various different values for the stringency variable were tested. With values of 60% or lower, it was found that polymorphisms which selected too many intermediate and susceptible varieties were identified, while values of 80% or greater resulted in the identification of very few polymorphisms. The optimal stringency value appeared to be 70%, since this resulted in the identification of several polymorphisms which each selected a satisfactory number of resistant varieties. This value was therefore chosen for the identification of additional polymorphisms.

Each polymorphism identified in the second step was used in combination with each polymorphism from the first step, and the varieties selected by the pairs of polymorphisms were assessed. A variety was considered to be selected if it possessed one or both of the polymorphisms in the pair. In general, three effects on the selected varieties were obtained.

(i) The varieties selected by the second polymorphism had already been selected by the first polymorphism.

(ii) The second polymorphism selected too many new susceptible varieties and too few new resistant varieties.

(iii) The second polymorphism selected few or no susceptible varieties as well as resistant varieties that had not already been selected by the first polymorphism.

Those pairs of polymorphisms that fell into category (iii), above, were chosen. A selection of these with the highest ratios between the percentage of resistant individuals and the percentage of susceptible individuals are listed in Table 5.4. Since no high stringency (70%) polymorphisms were obtained for eldana resistance, the polymorphisms in Table 5.4 are those that appear to be correlated with SCMV resistance and smut resistance only.

The combination of two or more high stringency polymorphisms with those identified in the

first step did not result in any further improvements on the selection of varieties, since this invariably resulted in the inclusion of too many susceptible varieties.

Table 5.4 Pairs of polymorphisms grouping the highest percentage resistant individuals with the lowest percentage susceptible individuals for two of the traits

Polymorphic fragments		% Resistant varieties	% Susceptible varieties
SCMV resistance			
$OD15_{270}$ (-)[a] $OD11_{1085}$ (-)	(x)[b] (o)	65.5	28.6
Smut resistance			
$OF01_{653}$ (-) $OD10_{1192}$ (-)	(x) (o)	64.0	20.0
$OF01_{653}$ (-) $OF06_{618}$ (+)	(x) (s)	76.0	26.7
$OF01_{653}$ (-) $OD04_{629}$ (+)	(x) (o)	72.0	26.7

[a] (+) and (-) indicate whether the varieties were grouped together according to the presence or the absence of the polymorphic fragment, respectively.

[b] Origin of the polymorphic fragment, from Table 5.2. (o): *S. officinarum*, (s): *S. spontaneum*, (x): neither *S. officinarum* nor *S. spontaneum*.

5.6.2 Multiple Regression Analysis

5.6.2.1 Prediction of eldana resistance

Using stepwise regression and starting with the best single polymorphism model, models with progressively higher numbers of polymorphisms were assessed in an attempt to obtain a significant fit of the polymorphism data from the 50 varieties to the eldana resistance ratings. Stepwise regression was performed until the seven-polymorphism model was obtained, and the following polymorphisms were selected in sequence: $OF14_{801}$, $OA08_{1249}$, $OD03_{984}$, $OD14_{1316}$, $OF06_{618}$, $OD15_{305}$ and $OD06_{354}$. Figure 5.9 shows how the r^2 and F-ratio values of the regression changed with the addition of each successive polymorphism to the model. While the r^2 values changed in accordance with the expected trend (an initial rapid increase followed by the levelling out of the curve, Figure 4.1, p. 62), the F values deviated considerably from

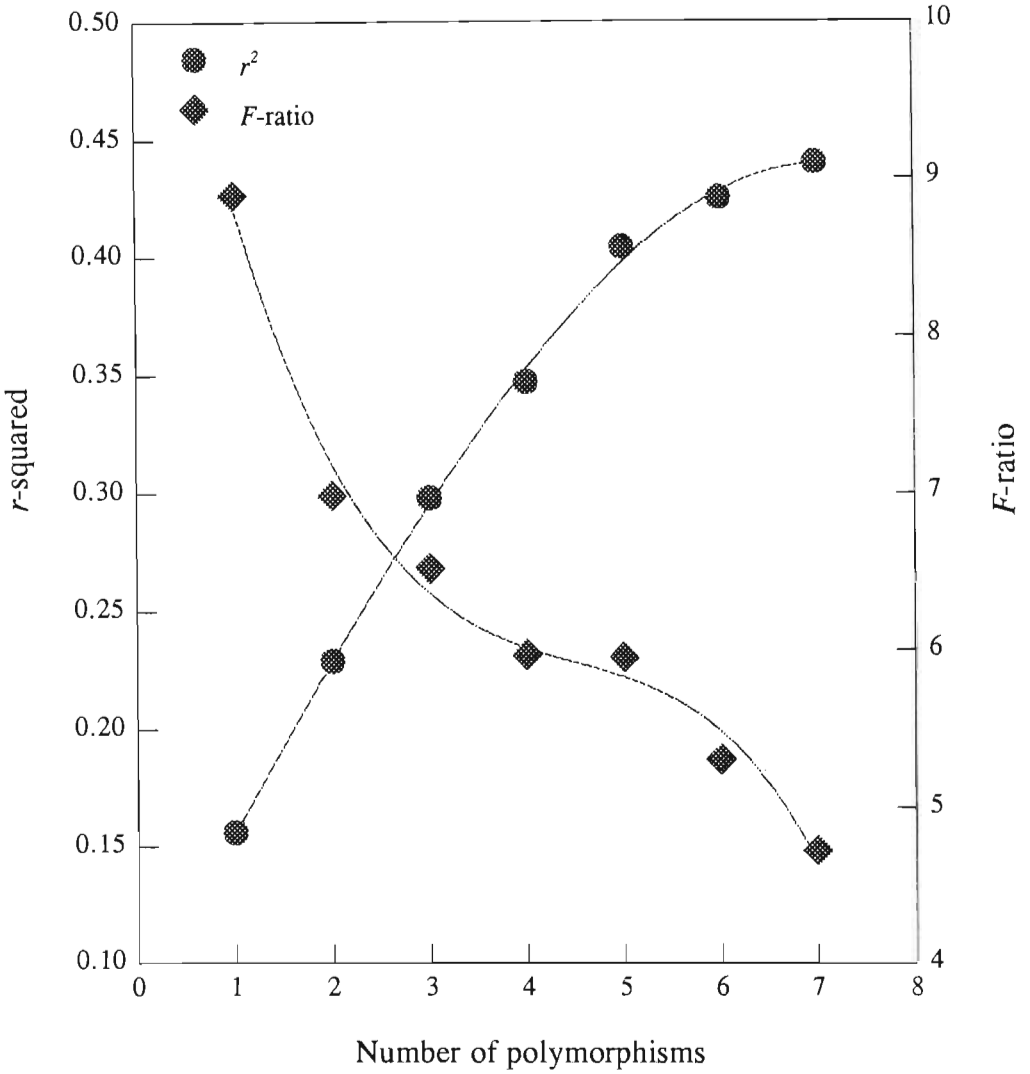


Figure 5.9 Relationship between r^2 and F-ratio values of regression models calibrated for eldana resistance, using successively higher numbers of polymorphisms

the expected trend (a bell shaped curve which gradually levels out as the variables in the model increase, Figure 4.1, p. 62). The reason for this lies in the unexpectedly high *F* value for the single polymorphism most highly correlated with eldana resistance ($OF14_{801}$, with an *F* ratio of 8.896). This value indicates a significant goodness of fit of the regression equation to the data. The subsequent drop in the *F*-ratio indicates that no other polymorphisms could be found that were able to improve the goodness of fit of the regression model to the eldana resistance data.

According to Figure 5.9, the most significant model for predicting eldana resistance would be the single-polymorphism model ($OF14_{801}$), since this has the highest *F*-ratio. However, since this model has a relatively low r^2 (0.156), it does not account for much of the variation in the eldana resistance ratings. It was therefore decided to test the predictive potential of the four-polymorphism model (Table 5.5). The r^2 of this model is 0.347, and the *F*-ratio is 5.972, which is still statistically significant (2.3 times the value required for significance at the 5% probability level).

Table 5.5 Predictive model for eldana resistance, determined from multiple regression analysis of the data from the 50 varieties[a]

Polymorphisms		r^2 **of the calibration model**
$OF14_{801}$	(x)[b]	0.347
$OA08_{1249}$	(x)	***F*-ratio**
$OD03_{984}$	(s)	5.972
$OD14_{1316}$	(x)	

[a] This model was calculated using the eldana resistance ratings that were available in 1995.
[b] Origin of the polymorphic fragment, from Table 5.2 (p. 76). (s): *S. spontaneum*, (x): neither *S. officinarum* nor *S. spontaneum*.

The four-polymorphism model for eldana resistance was used in a "leave out one at a time" approach on the 50 varieties. This involved the application of multiple regression analysis to 49 of the 50 varieties at a time, with the prediction of the fiftieth. The predicted ratings for all 50 varieties are listed in Table 5.6. This procedure allowed for the fitting of 50 separate

Table 5.6 The actual and predicted eldana resistance ratings for the 50 varieties, obtained from the "leave out one at a time" multiple regression analysis

Variety	Predicted rating	Actual rating ('95)	Actual rating ('96)	Correlation coefficient (r)*
74M659	3.604	3	4	0.5813
75E247	5.328	3	3	0.5928
76H333	6.304	9	6	0.5809
76M1101	3.749	1	3	0.5758
77L1720	4.937	3	3	0.5891
77W1241	5.962	9	4	0.5857
78F909	4.667	6	5	0.5919
79F1043	6.304	9	5	0.5809
79H181	4.807	7	9	0.5942
79L181	6.331	6	6	0.5882
79L1294	4.308	3	4	0.5842
80L432	4.807	7	8	0.5942
80W1459	4.467	2	3	0.5844
81L1308	6.274	7	6	0.5849
81W133	2.249	4	5	0.5925
81W447	8.198	4	4	0.6254
82F2907	2.441	5	5	0.6018
83F448	4.396	9	9	0.6142
J59/3	7.747	9	8	0.5679
N8	3.749	1	1	0.5758
N11	7.747	9	9	0.5679
N13	6.208	7	7	0.5851
N14	4.832	8	7	0.5990
N16	7.837	8	7	0.5759
N17	4.847	4	5	0.5880
N18	6.361	6	7	0.5882
N19	6.448	5	5	0.5935
N20	3.677	2	1	0.5776
N21	5.118	1	2	0.5981
N22	6.448	5	5	0.5935
N23	5.806	5	4	0.5901
N24	5.130	5	8	0.5888
N25	3.531	4	7	0.5868
NCo293	4.486	8	5	0.6045
NCo376	8.108	5	5	0.6103
77F790	1.796	6	7	0.6227
77L1143	6.100	9	6	0.5845
79M955	3.604	3	6	0.5813
CP57/614	5.378	3	3	0.5929
CB3822	3.604	3	2	0.5813
N52/219	6.100	9	5	0.5845
75E1293	5.130	5	5	0.5888
76M1566	3.749	1	3	0.5758
N12	8.198	4	3	0.6254
84F2753	4.733	9	8	0.6070
77F637	4.467	2	3	0.5844
75L1157	3.120	2	4	0.5734
NM214	6.187	8	5	0.5836
80F2147	7.927	7	7	0.5857
75L1463	3.677	2	3	0.5776
				0.5896 = Average r **0.3477** = Average r-squared

* The correlation coefficient has been calculated from multiple regression analysis between the actual ratings from 1995 and the predicted ratings

models for eldana resistance. These models will be very slightly different from each other, since a different variety was left out in each case. However, the average correlation coefficient of the models was determined, and hence the average r^2 calculated (Table 5.6). Since this value was derived from the 50 "leave out one at a time" models, it was almost identical to the r^2 of the models calibrated from all 50 varieties (compare the r^2 values in Table 5.5 and Table 5.6). The "leave out one at a time approach" can therefore be considered as the equivalent of using the four chosen polymorphisms in a regression model calibrated from the 50 varieties to predict eldana resistance in 50 unknown varieties, thereby allowing the predictive potential of the polymorphisms to be evaluated.

The 50 actual and predicted values (Table 5.6) were subjected to regression analysis, the output of which is shown in Table 5.7. The r^2 of the regression between the actual and predicted values can be interpreted as a measure of the performance of the selected polymorphisms in predicting eldana resistance in unknown varieties.

Table 5.7 Output of the regression between the 50 actual and predicted eldana resistance ratings

Constant	3.7547
Standard error of Y estimate[a]	1.4339
r^2 of the predictions	0.2137
Number of observations	50
Degrees of freedom	48
X Coefficient	0.2796
Standard error of coefficient	0.0774

[a] Standard error of the predictions

The predicted eldana resistance ratings were plotted against the actual ratings, and a line of best fit was drawn through the data points using the "constant" value of the regression output as the y-axis intercept and the "x coefficient" value as the gradient (Figure 5.10). This provided a visual indication of the association between the ratings predicted with the selected polymorphisms, and the actual resistance ratings.

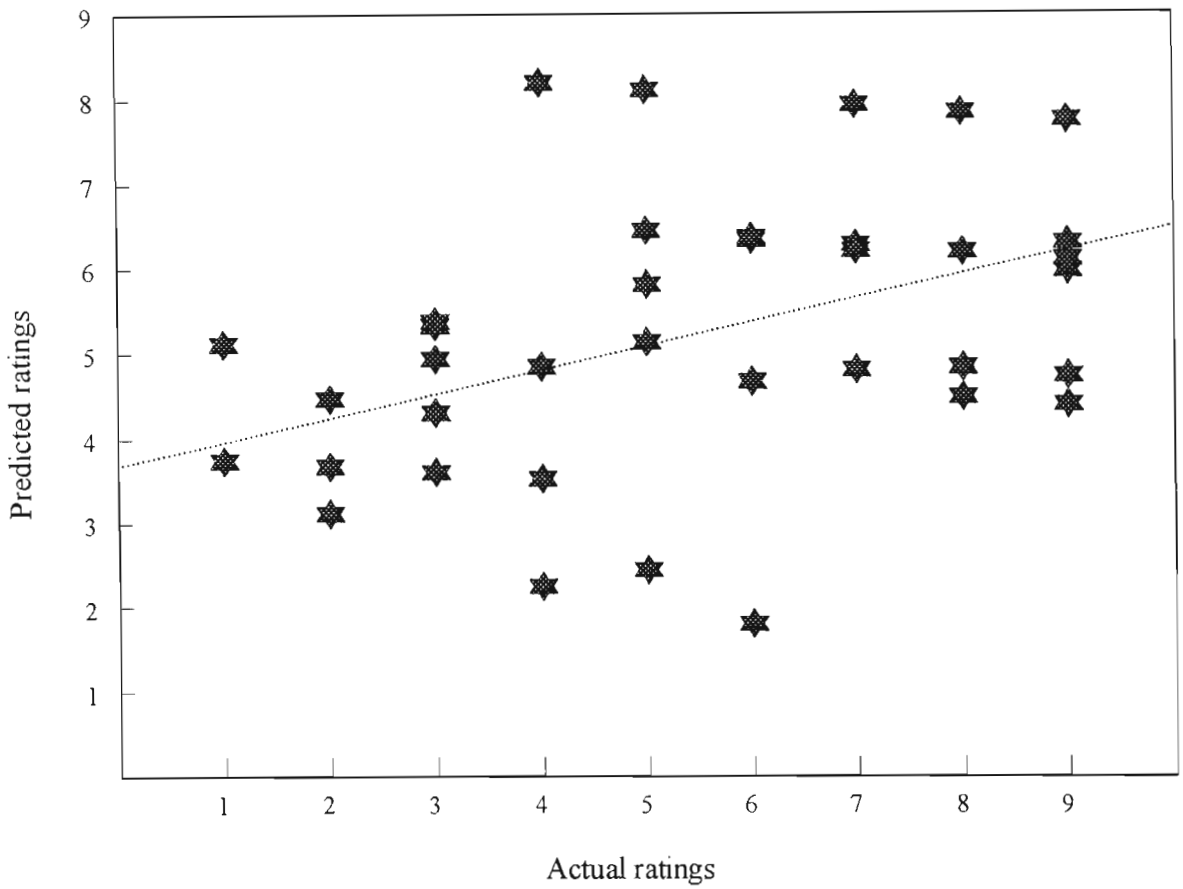


Figure 5.10 Association between the actual ratings for eldana resistance and the predicted ratings, calculated for each variety in turn based on the performance of the remaining 49

It must be emphasized that the actual ratings listed in Table 5.6 and depicted in Figure 5.10 are not absolute. The environmental conditions under which sugarcane varieties are grown can have a considerable effect on the phenotype. As a consequence, the ratings were assigned to the varieties after considering the results of numerous trials. The "actual" ratings must therefore be considered as the average performance of the variety and it should be borne in mind that these ratings may vary slightly depending on the conditions under which the variety is grown.

The regression analyses carried out above were all based on the eldana resistance ratings that were available in 1995. Subsequent trials resulted in the modification of those ratings, and these became available in 1996. The updated ratings are listed alongside those from 1995 in Table 5.6, and this provides some indication of the variability in phenotype that can occur with sugarcane. To determine whether the same polymorphisms were selected for the new ratings, multiple regression analysis was performed as before using the polymorphism data from the 50 varieties and the updated resistance ratings. For comparative purposes, only the four-polymorphism model was calculated, and this is shown in Table 5.8. A comparison of this model with that calculated from the 1995 ratings (Table 5.5) shows that only two of the polymorphisms ($OA08_{1249}$ and $OD14_{1316}$) have remained in the model. Two new polymorphisms were selected to account for the new variation in the 1996 ratings. It should be noted that neither of these were selected when up to the seven-polymorphism model for eldana resistance was determined using the 1995 ratings.

Table 5.8 Predictive model for eldana resistance, calculated from the updated resistance ratings (1996)

Polymorphisms		r^2 **of the calibration model**
$OD10_{1192}$	(o)[a]	0.314
$OF12_{1081}$	(o/s)	***F*-ratio**
$OA08_{1249}$	(x)	5.149
$OD14_{1316}$	(x)	

[a] Origin of the polymorphic fragment, from Table 5.2 (p. 76). (o): *S. officinarum*, (o/s): *S. officinarum* and *S. spontaneum*, (x): neither *S. officinarum* nor *S. spontaneum*.

The r^2 of the updated model (0.314) is slightly lower than the r^2 of the original model (0.347) (Table 5.5). Although the updated model was not used to predict the ratings of the 50 varieties using the "leave out one at a time" approach, it would be expected that the predictions would vary from those listed in Table 5.6. Since the r^2 of the new model is slightly lower, the r^2 of the new predictions is likely to be lower. This would indicate that the variation in the updated ratings cannot be accounted for to the same extent as the variation in the original ratings, despite the selection of new polymorphisms.

From the point of view of their applicability in a sugarcane selection programme, the resistance ratings (which range from 1 to 9) could be divided into various groups. For example, three groups could be used whereby ratings of 1, 2 and 3 signify resistance; ratings of 4, 5 and 6 signify an intermediate level of resistance; and ratings of 7, 8 and 9 signify susceptibility. An alternative possibility could be to omit the intermediate group and have ratings of 1 to 4 constituting the resistant category, and ratings of 5 to 9 constituting the susceptible category. Should markers be found that are able to predict the resistance ratings with an acceptable level of accuracy, the selection programme would then be able to determine which clones to discard and which to carry through to successive stages of selection. For example, only resistant clones could be kept, or resistant and intermediate (if applicable) clones could be kept. In order to determine how accurately the selected polymorphisms would able to predict eldana resistance ratings from the point of view of a sugarcane selection programme, the two approaches to grouping the ratings described above were considered.

The first approach treated ratings of 1 to 3 as resistant, 4 to 6 as intermediate, and 7 to 9 as susceptible. However, since the multiple regression does not predict the ratings as whole numbers, these three categories were adjusted to 1 to 3.4, 3.5 to 6.4, and 6.5 to 9, respectively. The actual (1995) and predicted ratings (Table 5.6) were compared, and the results listed in Table 5.9.

Table 5.9 Accuracy of the predicted eldana resistance ratings, assuming three categories of resistance[a]

Varieties predicted in correct groups	30%
Varieties predicted in adjacent groups[b]	70%
Varieties predicted in opposite groups[c]	0%

[a] Resistant (1-3), intermediate (4-6), susceptible (7-9)
[b] Resistant/intermediate, intermediate/susceptible
[c] Resistant/susceptible

The second approach treated ratings of 1 to 4 as resistant, and ratings of 5 to 9 as susceptible. To accommodate the multiple regression predictions, these were adjusted to 1 to 4.4, and 4.5 to 9, respectively. The results of the comparison between the actual (1995) and predicted ratings are listed in Table 5.10.

Table 5.10 Accuracy of the predicted eldana resistance ratings, assuming two categories of resistance[a]

Varieties predicted in correct groups	76%
Varieties predicted in opposite groups	24%

[a] Resistant (1-4), susceptible (5-9)

5.6.2.2 Prediction of SCMV resistance

In an attempt to obtain a significant fit of the polymorphism data from the 50 varieties to SCMV resistance, stepwise regression was performed on the data until the seven-polymorphism model was obtained. The following polymorphisms were selected in sequence: $OD04_{806}$, $OD11_{1085}$, $OF15_{1112}$, $OD17_{1109}$, $OD15_{767}$, $OD10_{400}$ and $0F15_{654}$. Figure 5.11 shows how the r^2 and F-ratio values of the regression changed with the addition of each successive polymorphism to the model. Both the r^2 and F-ratio changed in accordance with the expected trends as shown in Figure 4.1 (p. 62).

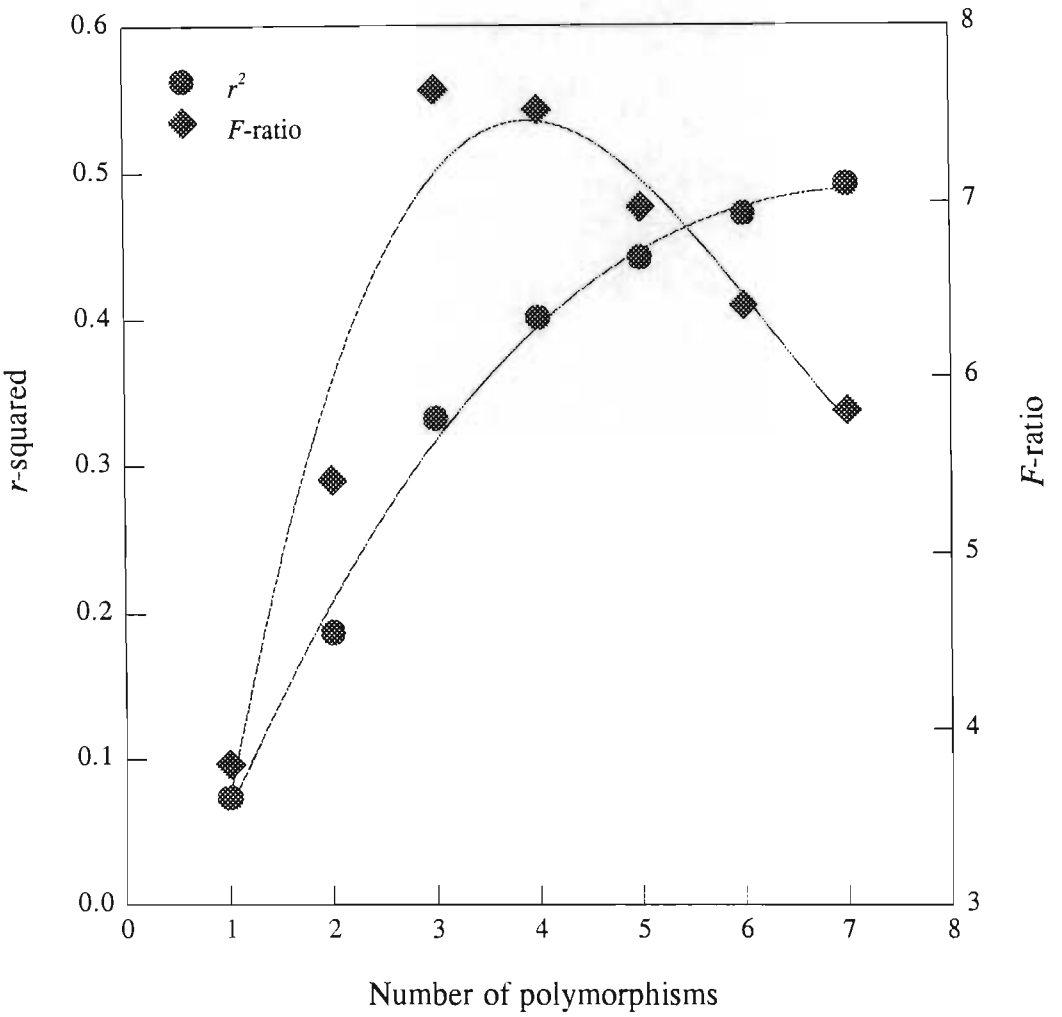


Figure 5.11 Relationship between r^2 and F-ratio values for regression models calibrated for SCMV resistance, using successively higher numbers of polymorphisms

The four-polymorphism model (Table 5.11) was chosen as being the most significant predictive model for SCMV resistance, since the *F*-ratio of the regression appears to be maximized at this point (Figure 5.11). This model has an r^2 of 0.401 and an *F*-ratio of 7.525 (2.9 times the value required for significance at the 5% probability level).

Table 5.11 Predictive model for SCMV resistance, determined from multiple regression analysis of the data from the 50 varieties

Polymorphisms		**r^2 of the calibration model**
$OD04_{806}$	(s)[a]	0.401
$OD11_{1085}$	(o)	***F*-ratio**
$OF15_{1112}$	(x)	7.525
$OD17_{1109}$	(o/s)	

[a] Origin of the polymorphic fragment, from Table 5.2 (p. 76). (o): *S. officinarum*, (s): *S. spontaneum*, (o/s): *S. officinarum* and *S. spontaneum*, (x): neither *S. officinarum* nor *S. spontaneum*.

The four-polymorphism model for SCMV resistance was used in a "leave out one at a time" approach on the 50 varieties, where multiple regression analysis was applied to 49 of the 50 varieties at a time with the prediction of the fiftieth. The predicted ratings for SCMV resistance thus obtained are listed in Table 5.12. As described for the prediction of eldana resistance, the average r^2 of the 50 individual models was calculated (Table 5.12) and this proved to be almost identical to the r^2 of the model calibrated from all 50 varieties (Table 5.11). As for the prediction of eldana resistance, the "leave out one at a time" approach can therefore be considered as the equivalent of using the four chosen polymorphisms in a regression model calibrated from the 50 varieties to predict SCMV resistance in 50 unknown varieties.

Table 5.12 The actual and predicted SCMV resistance ratings for the 50 varieties, obtained from the "leave out one at a time" multiple regression analysis

Variety	Predicted rating	Actual rating	Correlation coefficient (r)
74M659	6.283	7	0.6222
75E247	2.059	9	0.7033
76H333	4.022	2	0.6357
76M1101	6.545	7	0.6216
77L1720	4.087	1	0.6401
77W1241	6.034	9	0.6164
78F909	3.699	7	0.6447
79F1043	5.490	3	0.6434
79H181	2.479	2	0.6274
79L181	3.957	3	0.6334
79L1294	2.479	2	0.6274
80L432	4.087	1	0.6401
80W1459	2.539	1	0.6246
81L1308	5.077	3	0.6403
81W133	7.544	8	0.6113
81W447	6.545	7	0.6216
82F2907	0.369	2	0.6315
83F448	2.539	1	0.6246
J59/3	4.022	2	0.6357
N8	3.928	2	0.6335
N11	5.246	1	0.6574
N13	3.828	5	0.6349
N14	4.908	5	0.6320
N16	5.715	3	0.6431
N17	4.908	5	0.6320
N18	2.299	5	0.6473
N19	7.544	8	0.6113
N20	6.545	7	0.6216
N21	0.635	2	0.6300
N22	2.479	2	0.6274
N23	2.479	2	0.6274
N24	4.087	1	0.6401
N25	5.077	3	0.6403
NCo293	3.634	8	0.6529
NCo376	4.570	9	0.6413
77F790	2.479	2	0.6274
77L1143	2.479	2	0.6274
79M955	0.369	2	0.6315
CP57/614	4.087	1	0.6401
CB3822	5.162	2	0.6477
N52/219	2.479	2	0.6274
75E1293	4.729	8	0.6331
76M1566	6.159	8	0.6182
N12	4.908	5	0.6320
84F2753	2.479	2	0.6274
77F637	3.828	5	0.6349
75L1157	6.159	8	0.6182
NM214	2.419	3	0.6321
80F2147	5.185	5	0.6320
75L1463	3.828	5	0.6349
			0.6336 = Average r **0.4015** = Average r-squared

To evaluate the predictive potential of the selected polymorphisms, the 50 actual and predicted values for SCMV resistance (Table 5.12) were subjected to regression analysis. The output of this analysis is shown in Table 5.13. As before, the r^2 of the regression can be interpreted as a measure of the performance of the polymorphisms in predicting SCMV resistance in unknown varieties.

Table 5.13 Output of the regression between the 50 actual and predicted SCMV resistance ratings

Constant	2.6468
Standard error of Y estimate[a]	1.4738
r^2 of the predictions	0.2952
Number of observations	50
Degrees of freedom	48
X Coefficient	0.3519
Standard error of coefficient	0.0785

[a] Standard error of the predictions

To provide a visual indication of the association between the ratings predicted with the selected polymorphisms and the actual resistance ratings, the predicted ratings were plotted against the actual ratings (Figure 5.12). The line of best fit was drawn using the "constant" value of the regression as the y-axis intercept and the "x coefficient" value as the gradient.

Once again, it must be emphasized that the actual SCMV ratings referred to in this section are not fixed values. These values have been assigned to the varieties after several trials, and must be considered as the average performance of each variety, subject to slight variation depending on the prevailing environmental conditions.

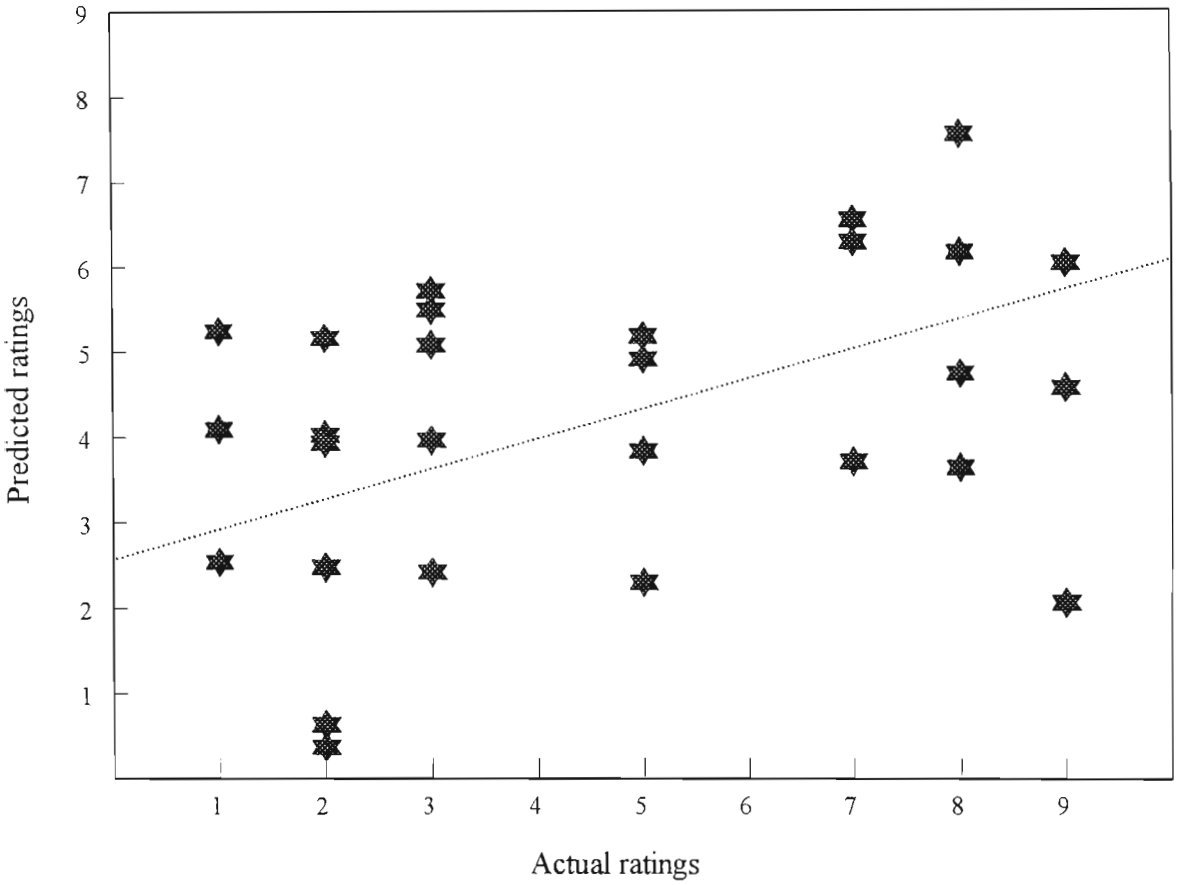


Figure 5.12 Association between the actual ratings for SCMV resistance and the predicted ratings, calculated for each variety in turn based on the performance of the remaining 49

The accuracy of the four selected polymorphisms in predicting SCMV resistance in terms of various categories of ratings was then considered. Two approaches to grouping the ratings were assessed. The first involved three groups: resistant (1-3), intermediate (4-6), and susceptible (7-9). Since the multiple regression does not predict the ratings as whole numbers, these categories were adjusted to 1 to 3.4, 3.5 to 6.4, and 6.5 to 9 respectively. The actual and predicted ratings (Table 5.12) were compared and the results listed in Table 5.14.

Table 5.14 Accuracy of the predicted SCMV resistance ratings, assuming three categories of resistance[a]

Varieties predicted in correct groups	52%
Varieties predicted in adjacent groups[b]	46%
Varieties predicted in opposite groups[c]	2%

[a] Resistant (1-3), intermediate (4-6), susceptible (7-9)
[b] Resistant/intermediate, intermediate/susceptible
[c] Resistant/susceptible

The second approach treated ratings of 1 to 4 as resistant, and ratings of 5 to 9 as susceptible. Once again, to accommodate the multiple regression predictions, these were adjusted to 1 to 4.4, and 4.5 to 9, respectively. The results of the comparison between the actual and predicted ratings are listed in Table 5.15.

Table 5.15 Accuracy of the predicted SCMV resistance ratings, assuming two categories of resistance[a]

Varieties predicted in correct groups	74%
Varieties predicted in opposite groups	26%

[a] Resistant (1-4), susceptible (5-9)

5.6.2.3 Prediction of smut resistance

Using stepwise regression and starting with the best single polymorphism model, models with progressively higher numbers of polymorphisms were assessed in an attempt to obtain a significant fit of the polymorphism data from the 50 varieties to smut resistance. Stepwise regression can result in previously selected variables being replaced with other variables that improve the overall fit of the model. While this did not occur for either eldana resistance or SCMV resistance, the selection of polymorphisms for smut resistance was influenced by this effect. However, all the polymorphisms that were initially removed from the model were reselected by the time the seven-polymorphism model had been calculated. This model included the following polymorphisms: $OD14_{926}$, $OF01_{653}$, $OD17_{863}$, $OD20_{933}$, $OA08_{967}$, $OD15_{767}$ and $OF06_{618}$. $OF01_{653}$ was the single polymorphism most highly correlated with smut resistance. As shown in Figure 5.13, the r^2 and F-ratio values of the regression followed the expected trends as successive polymorphisms were added to the model (refer to Figure 4.1, p. 62).

According to Figure 5.13, the F-ratio of the regression is maximized in the three-polymorphism model, indicating the most significant fit of the regression equation to the smut resistance data. The three-polymorphism model was therefore chosen as the predictive model for smut resistance (Table 5.16). The r^2 of this model is 0.316, and the F-ratio is 7.094 (2.5 times the value required for statistical significance at the 5% probability level).

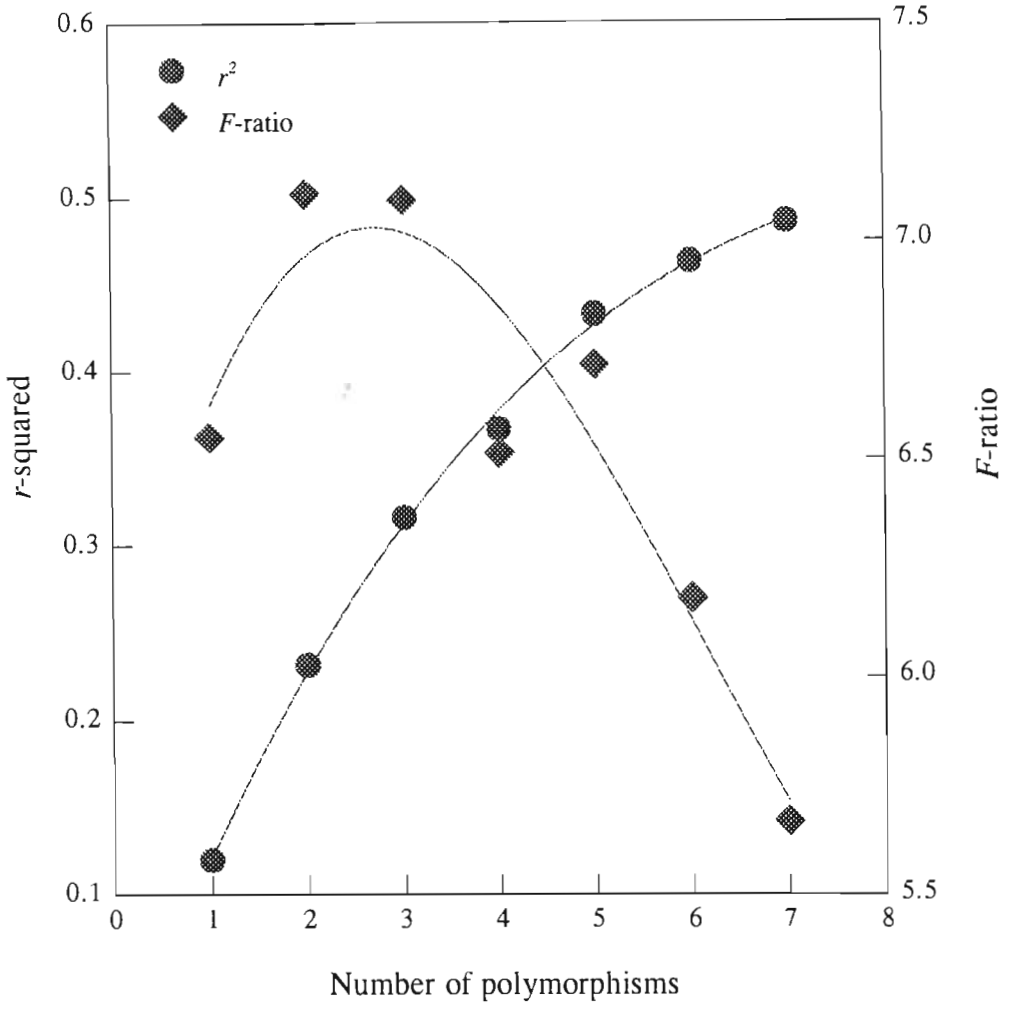


Figure 5.13 Relationship between r^2 and F-ratio values for regression models calibrated for smut resistance, using successively higher numbers of polymorphisms

Table 5.16 Predictive model for smut resistance, determined from multiple regression analysis of the data from the 50 varieties

Polymorphisms		r^2 **of the calibration model**
$OD14_{926}$	(x)[a]	0.316
$OF06_{618}$	(s)	***F*-ratio**
$OF01_{653}$	(x)	7.094

[a] Origin of the polymorphic fragment, from Table 5.2 (p. 76). (s): *S. spontaneum*, (x): neither *S. officinarum* nor *S. spontaneum*.

To test the predictive potential of the three-polymorphism model, it was used in a "leave out one at a time" approach on the 50 varieties. As described previously, the application of multiple regression analysis to 49 of the 50 varieties at a time with the prediction of the fiftieth can be considered as the equivalent of using the three chosen polymorphisms in a regression model calibrated from the 50 varieties to predict smut resistance in 50 unknown varieties. The predicted ratings for smut resistance are listed in Table 5.17. The average r^2 of the 50 individual models was calculated (Table 5.17), and this value was almost identical to the r^2 of the model calibrated from all 50 varieties (Table 5.16).

The 50 actual and predicted values of smut resistance (Table 5.17) were subjected to regression analysis, the output of which is shown in Table 5.18. The r^2 of the regression between the actual and predicted values can be interpreted as a measure of the performance of the selected polymorphisms in predicting smut resistance in unknown varieties.

Table 5.17 The actual and predicted smut resistance ratings for the 50 varieties, obtained from the "leave out one at a time" multiple regression analysis

Variety	Predicted rating	Actual rating	Correlation coefficient (r)
74M659	6.340	7	0.5492
75E247	3.907	2	0.5613
76H333	3.907	2	0.5613
76M1101	5.299	5	0.5618
77L1720	5.160	8	0.5576
77W1241	3.580	7	0.5794
78F909	2.059	1	0.5425
79F1043	3.517	3	0.5596
79H181	3.907	2	0.5613
79L181	8.660	7	0.5544
79L1294	3.580	7	0.5794
80L432	5.437	2	0.5832
80W1459	5.207	7	0.5571
81L1308	5.391	3	0.5741
81W133	5.391	3	0.5741
81W447	3.842	3	0.5608
82F2907	3.628	2	0.5583
83F448	5.452	3	0.5724
J59/3	1.968	2	0.5515
N8	5.114	9	0.5603
N11	5.437	2	0.5832
N13	5.207	7	0.5571
N14	3.517	3	0.5596
N16	5.207	7	0.5571
N17	3.907	2	0.5613
N18	3.296	5	0.5688
N19	5.437	2	0.5832
N20	5.207	7	0.5571
N21	3.907	2	0.5613
N22	3.973	1	0.5639
N23	2.059	1	0.5425
N24	1.968	2	0.5515
N25	3.296	5	0.5688
NCo293	8.095	9	0.5172
NCo376	5.160	8	0.5576
77F790	5.391	3	0.5741
77L1143	5.114	9	0.5603
79M955	4.496	7	0.5628
CP57/614	4.974	5	0.5616
CB3822	1.968	2	0.5515
N52/219	3.973	1	0.5639
75E1293	3.711	5	0.5659
76M1566	3.711	5	0.5659
N12	5.345	4	0.5670
84F2753	3.739	1	0.5594
77F637	5.299	5	0.5618
75L1157	5.253	6	0.5585
NM214	1.876	3	0.5621
80F2147	1.785	4	0.5746
75L1463	3.449	9	0.6022
			0.5628 = Average r
			0.3168 = Average r-squared

Table 5.18 Output of the regression between the 50 actual and predicted smut resistance ratings

Constant	3.1684
Standard error of Y estimate[a]	1.2887
r^2 of the predictions	0.2265
Number of observations	50
Degrees of freedom	48
X Coefficient	0.2704
Standard error of coefficient	0.0721

[a] Standard error of the predictions

To provide a visual indication of the association between the ratings predicted with the selected polymorphisms and the actual resistance ratings, the predicted ratings were plotted against the actual ratings (Figure 5.14). The line of best fit was drawn through the data points using the "constant" value of the regression as the y-intercept and the "x coefficient" value as the gradient.

As for eldana and SCMV, the smut resistance ratings are not fixed values. Since the environmental conditions under which the sugarcane varieties are grown can affect their phenotype, the actual smut ratings must be considered as the average performance of each variety, subject to slight variation.

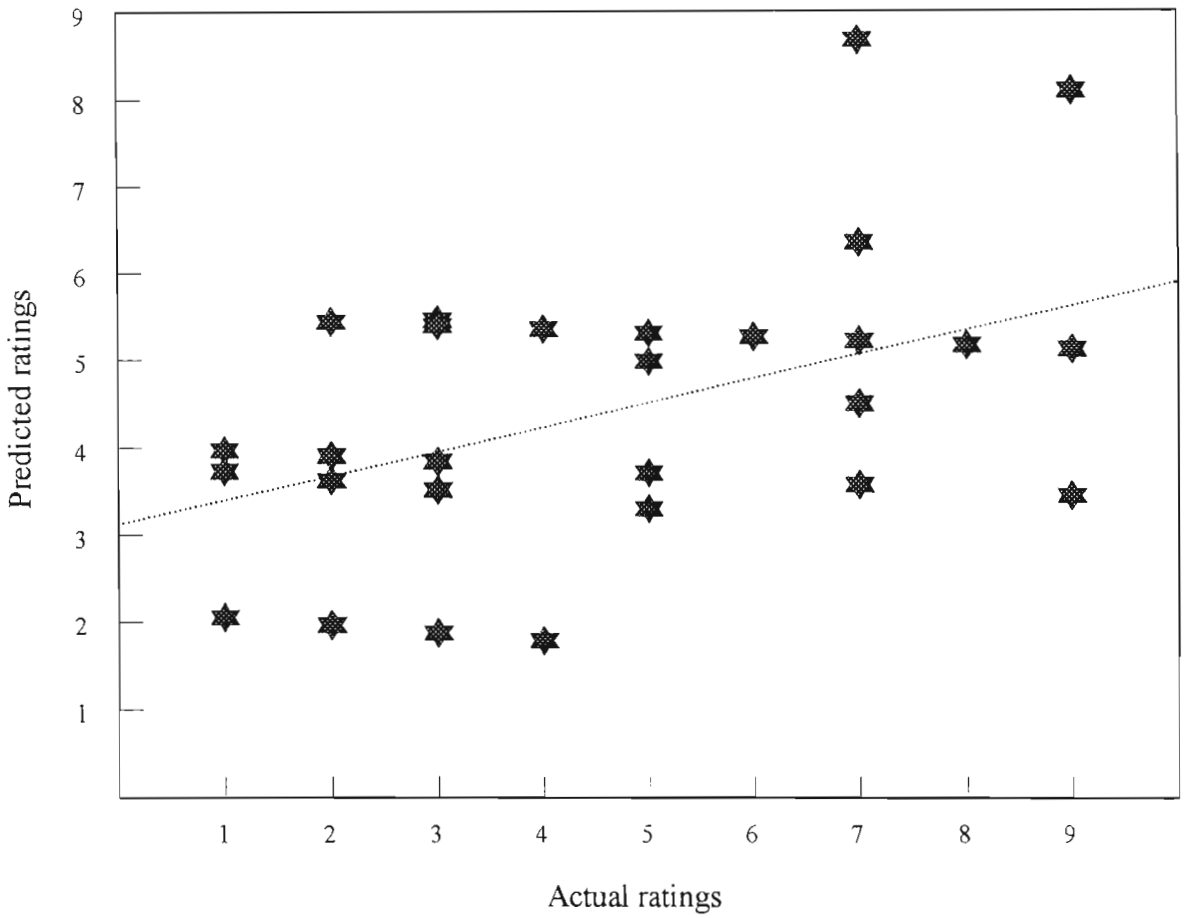


Figure 5.14 Association between the actual ratings for smut resistance and the predicted ratings, calculated for each variety in turn based on the performance of the remaining 49

The accuracy of the three selected polymorphisms in predicting smut resistance in terms of various categories of ratings was then considered. Two approaches to grouping the ratings were assessed. The first involved three groups: resistant (1-3), intermediate (4-6), and susceptible (7-9). Since the multiple regression does not predict the ratings as whole numbers, these categories were adjusted to 1 to 3.4, 3.5 to 6.4, and 6.5 to 9 respectively. The actual and predicted ratings (Table 5.17) were compared and the results listed in Table 5.19.

Table 5.19 Accuracy of the predicted smut resistance ratings, assuming three categories of resistance[a]

Varieties predicted in correct groups	30%
Varieties predicted in adjacent groups[b]	68%
Varieties predicted in opposite groups[c]	2%

[a] Resistant (1-3), intermediate (4-6), susceptible (7-9)
[b] Resistant/intermediate, intermediate/susceptible
[c] Resistant/susceptible

The second approach treated ratings of 1 to 4 as resistant, and ratings of 5 to 9 as susceptible. Once again, to accommodate the multiple regression predictions, these were adjusted to 1 to 4.4, and 4.5 to 9, respectively. The results of the comparison between the actual and predicted ratings are listed in Table 5.20.

Table 5.20 Accuracy of the predicted smut resistance ratings, assuming two categories of resistance[a]

Varieties predicted in correct groups	70%
Varieties predicted in opposite groups	30%

[a] Resistant (1-4), susceptible (5-9)

CHAPTER 6

DISCUSSION

6.1 Choice of Phenotypes

There is a wide variety of phenotypes, or traits, that can be used in a study aimed at finding associations between genetic markers and traits. With sugarcane, the obvious traits to choose would be those with agronomic and economic advantages, such as fibre content, cane tonnage, sucrose percentages, and resistance to the various pathogens and pests that infect and damage the crop. As in most crops, the characters of economic importance in sugarcane (such as cane yield) are mainly quantitative, and are controlled by multiple genes (Stevenson, 1965; Skinner *et al.*, 1987). Disease resistance in sugarcane appears to be generally inherited as a complex of a few major genes, rather than as a simple Mendelian character or as a typical polygene system (Stevenson, 1965).

The ideal traits for which to find markers would be those that are monogenic, however it appears as though simple Mendelian inheritance has limited applicability in sugarcane (Hogarth, 1987). Of the traits mentioned above, disease resistance is probably a more suitable candidate for finding markers since, if not simply inherited, then fewer genes are likely to be involved in the expression of disease resistance than with complex traits such as yield or sucrose production.

For this study, it was decided to choose levels of resistance to three pests or pathogens for the initial survey, since the ratings for these traits are readily available for the chosen varieties. Due to the nature of the analysis, other traits can easily be added to the survey at a later stage in order to identify useful markers linked to additional traits. Resistance to the sugarcane borer *Eldana saccharina*, sugarcane mosaic virus (SCMV) and the fungal pathogen smut (*Ustilago scitaminea*) were chosen. Although these three traits all involve resistance to pests or diseases, the actual mechanisms and genetic bases of resistance are likely to be quite varied and this

could have an effect on the relative ease with which potential markers might be found.

Resistance to pests or pathogens can show a continuous variation, ranging from complete resistance to extreme susceptibility. Such variation is usually caused by the effects of multiple genes as well as a range of nongenetic factors. It is worthwhile to note, however, that continuous variation is not exclusively limited to polygenic inheritance, since nongenetic factors may blur small differences between distinct genotypic values (Geiger & Heun, 1989).

6.1.1 Resistance to Smut (*Ustilago scitaminea*)

Many cases of resistance to fungal diseases are under the control of a number of genes, although the actual number may well be smaller than for more typical polygenic characters such as growth, weight, etc. (Geiger & Heun, 1989). A recent study by Daugrois *et al.* (1996) showed that resistance to common rust of sugarcane (*Puccinia melanocephala*) segregated in a 3 (resistant) : 1 (susceptible) fashion, indicative of a probable dominant resistant gene. However, this is the first documented report of monogenic inheritance of disease resistance in sugarcane, and is contrary to the general opinion that rust resistance is a quantitatively inherited trait in sugarcane (Hogarth *et al.*, 1993). The genetic basis of resistance to smut in sugarcane is unknown, although if the above mentioned factors are taken into consideration, it could well be a polygenic character.

6.1.2 Resistance to Sugarcane Mosaic Virus

In contrast to fungal disease resistance, many cases of resistance to plant viruses are under very simple genetic control while fewer cases appear to be controlled by a few or several genes (Fraser, 1990). Vector transmission of viruses can be an important factor in the development of disease, and the apparent resistance of the plant can be altered by host genes that affect plant preference and feeding behaviour of the vectors, and by viral genes that affect transmissibility (Fraser, 1990). SCMV is transmitted to sugarcane by aphids, although the disease is also spread through the use of infected seedcane. Studies have yet to be done to establish the genetic basis of the observed resistance of sugarcane to SCMV, although it is possible that aphid resistance and virus resistance both contribute towards this trait.

6.1.3 Resistance to *Eldana saccharina*

Resistance of sugarcane to the stalk borer eldana is probably a fairly complex trait with a number of mechanisms working together to determine the level of resistance. Studies using near-infrared spectrophotometry have suggested that stalk surface wax components and flavonoid components of nodal budscales of sugarcane contribute towards resistance to eldana (Rutherford *et al*, 1995; Rutherford & van Staden, 1996). The fibre content of sugarcane is also believed to play a role in eldana resistance (Nuss *et al.*, 1986). If all these factors, and possibly others, contribute to the observed levels of resistance, this particular trait could well be more complex than the two traits mentioned above, with a greater number of genetic loci affecting the resistance levels.

6.2 RAPD Amplification of the Sugarcane DNA

The RAPD protocol used in this study was previously optimized for the reliable amplification of sugarcane genomic DNA (Sobral & Honeycutt, 1993). This protocol makes use of the Stoffel fragment (Perkin-Elmer), which is a truncated form of *Taq* DNA polymerase. The 289 N-terminal amino acids have been deleted from this modified polymerase, with the resultant loss of 5' to 3' polymerization-dependent exonuclease activity. In the optimization of the RAPD amplification of sugarcane DNA, Sobral and Honeycutt (1993) compared the relative efficiencies of various thermostable polymerases: the Stoffel fragment; *AmpliTaq* (Perkin-Elmer), which is *Taq* DNA polymerase with 5' to 3' exonuclease activity; and *Pfu* polymerase, which is a polymerase from *Pyrococcus furiosis* that has 5' to 3' DNA polymerase activity as well as 3' to 5' exonuclease activity. They found that the use of the Stoffel fragment resulted in the increase of the average number of loci amplified per primer, as well as the average number of polymorphisms amplified per primer relative to *AmpliTaq* and *Pfu*. In addition, the number of primers that failed to give any products decreased with the use of the Stoffel fragment.

The RAPD amplification products, or profiles, that are obtained with different DNA polymerases vary considerably and require different amplification parameters (Caetano-Anollés & Bassam, 1993; Sobral & Honeycutt, 1993). The output of the RAPD reactions obtained in

this study was therefore compared to the results obtained by other authors who used the same RAPD protocol for the amplification of sugarcane genomic DNA. On average, 11.6 loci were amplified per primer, and this compared favourably with the results of Al-Janabi *et al.* (1993), who obtained an average of 16.5 loci per primer; Huckett and Botha (1995), who obtained an average of 11.6 loci per primer; and Harvey and Botha (1996), who obtained an average of 10.6 loci per primer. From the primers that produced scorable polymorphisms, this study elucidated an average of 2.1 polymorphisms per primer, Al-Janabi *et al.* (1993) obtained an average of 2.3, Huckett and Botha (1995) obtained an average of 1.7, and Harvey and Botha (1996) obtained an average of 6.7. The apparent discrepancy in the average number of polymorphisms per primer obtained by Harvey and Botha (1996) can be attributable to the approach used to score the polymorphisms. Whereas only the clear, easily scorable polymorphisms were considered in both the present study and that of Al-Janabi *et al.* (1993), all polymorphic fragments were scored by Harvey and Botha (1996). It is also important to note that Harvey and Botha (1996) selected primers that detected at least a few polymorphisms, as opposed to at least one which was the criterion used in the other studies, including this one. The comparatively lower average number of polymorphisms per primer obtained by Huckett and Botha (1995) is in keeping with the particularly close relationship of the two varieties used in that study.

6.3 Origin of the Polymorphisms With Respect to *S. officinarum* and *S. spontaneum*

The sugarcane varieties that are currently cultivated are derived mainly from the interspecific hybridization of *S. officinarum* and *S. spontaneum* that was carried out in the late 1800's. *S. barberi* and *S. sinense* were also used in the breeding of sugarcane varieties, however the contribution of these species was very limited (Berding & Roach, 1987), and they are themselves the product of natural interspecific hybridization between *S. officinarum* and *S. spontaneum* (Daniels & Roach, 1987). It was this narrow genetic background that prompted the suggestion that modern sugarcane varieties could essentially be considered as a large progeny derived from the parents, *S. officinarum* and *S. spontaneum* (Lu *et al.*, 1994a), which formed the rationale for this study.

The DNA from *S. officinarum* and *S. spontaneum* was therefore amplified in addition to the DNA from the 50 varieties used in this study, in order to allow the detected polymorphisms to be traced back to the two "parent" species. Assessment of the genetic variability in *S. officinarum* and *S. spontaneum* with the use of RFLPs (Lu *et al.*, 1994b) has shown that diversity is the highest in *S. spontaneum*, and by comparison is considerably lower in *S. officinarum*. Three clones of *S. officinarum* and six clones of *S. spontaneum* were therefore chosen to represent the two "parent" species. The respective DNA extracts from these clones were pooled to form two bulks representative of the two species, and these were then amplified alongside the DNA from the 50 varieties. As a result, 37.0% of the polymorphisms were found to be *S. officinarum*-specific, 14.8% were *S. spontaneum*-specific, 5.6% were present in both *S. officinarum* and *S. spontaneum*, while 42.6% were absent in both species (Table 5.2., p. 76).

Due to the three or four generations of nobilization that occurred in the early stages of sugarcane breeding, modern varieties have chromosome numbers ranging from 100 to 130, of which only 5 to 10% come from *S. spontaneum*. This was recently demonstrated in a study using fluorescent *in situ* hybridization with different coloured probes for *S. officinarum* and *S. spontaneum* DNA, respectively (D'Hont *et al.*, 1996). It is therefore not surprising that only 14.8% of the detected polymorphisms appeared to originate from *S. spontaneum* while 37% came from *S. officinarum*, assuming that loci that are randomly distributed throughout the whole genome are targeted by the RAPD technique.

According to a study by Lu *et al.* (1994b) in which RFLPs were used to study the relationships among ancestral species of sugarcane, the average genetic similarity between *S. spontaneum* and *S. officinarum* is 20%. This divergence explains why so few of the polymorphisms (5.6%) were present in both the *S. spontaneum* and *S. officinarum* samples. However, the two DNA bulks did not necessarily contain all the clones that were used in the ancestry of the 50 varieties. In addition, it must be remembered that only the origin of clear, easily scorable polymorphisms were investigated. Consequently, the total number of amplified fragments that were, in reality, in common between the two species may have been closer to the 20% similarity reported by Lu *et al.* (1994b).

The majority of the polymorphic fragments detected in this study (42.6%) were not amplified in either the *S. officinarum* or the *S. spontaneum* samples. This may have been attributable to the fact that not all the clones used in the ancestry of the varieties were used to prepare the two DNA bulks. Another possible reason could be genetic recombination. Büscher *et al.* (1994), working with grapevine, and Huckett and Botha (1995), working with sugarcane, both reported on the presence of RAPD fragments in a particular variety that was not observed in either of the parents, and explained these inconsistencies as being a result of either recombination events at the primer binding sites, or questionable paternal identity. The 50 sugarcane varieties used in this study do not represent the true progeny of a cross between *S. officinarum* and *S. spontaneum*, and have been separated from the original *S. officinarum* x *S. spontaneum* crosses by six or seven generations. It is therefore possible that a relatively large proportion of the polymorphisms that were not amplified in the *S. officinarum* and *S. spontaneum* samples may have been due to recombination events. In the event of recombination at the primer binding site, the target sequence that is amplified as a polymorphism in the varieties will be present in *S. officinarum* or *S. spontaneum*, but will not be amplified due to the absence of the primer binding site.

A third reason for the occurrence of RAPD fragments in the varieties that were not amplified in either *S. officinarum* or *S. spontaneum* could be the formation of heteroduplexes. Heteroduplexes are "hybrid" double-stranded DNA complexes that are formed following the PCR amplification of two DNA fragments that have a high degree of homology to one another, yet differ in their nucleotide sequences due to base pair deletions, additions or substitutions (Davis *et al.*, 1995; Novy & Vorsa, 1996). The annealing of single strands from the two fragments results in the formation of heteroduplexes. During electrophoresis, heteroduplexes generally migrate at different rates than the two original fragments due to conformational changes. The formation of heteroduplexes can be a consequence of heterozygosity, and can result in the presence of non-parental RAPDs in the progeny of parents homozygous for alternate alleles at a particular locus.

6.4 Associations Between RAPD Polymorphisms and Phenotypes

6.4.1 Correlation Analysis Between Polymorphisms and Phenotypes

The first method that was investigated for its potential to identify associations between RAPD polymorphisms and specific traits was a simple correlation approach. This approach merely attempted to determine whether the presence of any polymorphisms in a sugarcane variety implied the existence of a particular phenotypic state. Initially, polymorphisms were selected that were able to group together greater than half of the resistant varieties with less than half of the susceptible varieties for each trait. Following this, additional "high stringency" polymorphisms were identified that, when added to those from the first step, would increase the percentage of resistant varieties that were "selected", while maintaining the percentage of susceptible varieties as low as possible. Polymorphisms from the first and second steps were used in pairs such that the second polymorphism "selected" as many resistant varieties as possible that had not already been selected by the first polymorphism, and as few susceptible varieties as possible.

A variety was considered to be selected if it possessed one or both of the polymorphisms in the pair. This criterion was chosen rather than requiring the varieties to possess both polymorphisms in the pair. Assuming that the polymorphisms are in some way linked to the trait, different polymorphisms may correspond to different loci of a quantitative trait, and while the presence of both loci may result in a higher level of resistance, the presence of just one of the loci may be enough to prevent the variety from being susceptible. In the event of monogenic resistance, a single polymorphism may be found that is able to group most of the resistant varieties together, bearing in mind the effect environmental conditions may have on the apparent resistance. Since the resistance ratings for the two diseases as well as for eldana range from 1 to 9, it seems unlikely that these traits would be monogenic, but rather polygenic or perhaps oligogenic.

For eldana resistance, the only polymorphisms that were found were those that were able to group more than half of the resistant varieties with less than half of the susceptible varieties; no "high stringency" polymorphisms were identified that were able to improve those groups. However, the single polymorphisms, alone, were able to group a relatively high proportion

of the resistant varieties with a relatively low proportion of the susceptible varieties (refer to Table 5.3, p. 77). The groups of varieties selected by the individual polymorphisms for eldana resistance compared favourably with the groups of varieties selected by the pairs of polymorphisms for both SCMV and smut resistance (compare eldana resistance, Table 5.3, p. 77, with SCMV resistance and smut resistance, Table 5.4, p. 79). This may indicate that a major gene is responsible for conferring the observed resistance of sugarcane varieties to eldana, and that the effects of this locus are modified by a few or several minor genes. However, linkage of the polymorphisms to the trait would have to be established before any definite conclusions can be drawn. As an initial test, a new group of varieties could be screened with the three individual polymorphisms that seem to be associated with eldana resistance. This would determine whether the same correlations exist within a much larger group of varieties.

For both SCMV and smut resistance, the combination of three or more polymorphisms invariably resulted in the selection of too many susceptible varieties. This raises the question of whether the polymorphisms are actually associated with the traits. If polymorphisms are found that are actually linked to the traits, and assuming that the traits are quantitative, three or more polymorphisms could well be used together without the selection of a large proportion of susceptible varieties. In this study, however, only a relatively small number of polymorphisms were tested (54). The larger the number of polymorphisms that can be tested in such an analysis, the greater the chances of finding markers that are linked to the traits in question.

Although no conclusions can be drawn regarding the actual linkage of the polymorphisms with the three traits, it is interesting to note the features of the polymorphisms that appear to be associated with the traits from this initial analysis. The random nature of the RAPD polymorphisms is demonstrated by the apparent association of different polymorphisms amplified from the same primer with different traits. For example, $OA08_{1249}$ and $OA08_{1379}$ appear to be associated with SCMV resistance and smut resistance, respectively (Table 5.3, p. 77). Alternatively, a single polymorphism ($OD17_{1109}$) appeared to be associated with both SCMV and smut resistance (Table 5.3, p. 77). However, with no direct means of proving

linkage of the polymorphisms to the traits, the apparent association of $OD17_{1109}$ with the two traits may be purely coincidental. In terms of the origins of the polymorphisms, there does not appear to be any immediate consensus among the polymorphisms for each trait (Table 5.3, p. 77; Table 5.4, p. 79). Each of the traits have one polymorphism that appeared in both *S. officinarum* and *S. spontaneum* samples, leading to the possibility that all three traits originated in both species. However, in the absence of any linkage information, no deductions can actually be drawn regarding the contribution of traits by either *S. officinarum* or *S. spontaneum* to the germplasm of modern sugarcane varieties.

6.4.2 Prediction of Resistance With Multiple Regression Analysis

Multiple regression is a statistical method of predicting one variable from one or more other variables. This approach was therefore used to determine whether polymorphisms could be used to predict the resistance ratings of the sugarcane varieties. A similar analysis was carried out by Virk and co-workers (1996) on rice. Rather than using the progeny from a planned cross, this group used 48 accessions, or varieties, of rice in order to identify associations between RAPD markers and quantitative traits. Using the multiple regression approach, they were able to obtain significant associations between RAPD markers and all six of the traits that they investigated. The predictive power of this approach was demonstrated in two ways. Firstly, multiple regression analysis was carried out on 40 of the accessions, and the regression parameters thus estimated were used to predict the quantitative variation in the remaining eight accessions. Secondly, the "leave out one at a time" method was used, whereby each of the 48 accessions in turn was predicted based on the regression analysis of the remaining 47. In both of the above-mentioned analyses, the authors obtained good predictions of the performance of the accessions. In the case of the "leave out one at a time" analysis, only seven predictions differed significantly from the observed performances. These failures constitute 2.4% of the 288 accession/trait combinations to which models were fitted, and in no case were more than two failures noted for any of the six traits. According to these authors, up to 5% of the predictions can be expected to differ from the observed values. They therefore claim that their predictions are highly reliable.

Virk *et al.* (1996) used the SAS statistics package (SAS, 1990) to carry out the multiple

regression analysis. The maximum r^2 improvement option of this package was used to determine the most appropriate regression model for each trait. Initially, one variable (i.e. marker) models were assessed and the marker with the highest r^2 was identified. Then the second variable was added and the best two-variable model was selected using the usual criterion of the largest r^2. The model fitting was continued until all significant variation in Y (the value for the trait) was exhausted. A further condition was imposed on the model, whereby a new variable was retained in the model only if it was significant at the 5% level. A total of 63 polymorphic RAPD bands were used in the analysis, and using the maximum r^2 improvement option, the number of bands that showed significant association with the six different traits varied from 12 to 32. Furthermore, the multiple regression accounted for all the significant variation in all the traits (the r^2 varied from 0.89 to 0.99).

It may be a particularly desirable objective to have 89% to 99% of the variation in a trait accounted for by various markers. However, the use of 12 to 32 markers to select for specific traits in a selection programme is economically and practically unrealistic, since selection programmes typically deal with thousands of plants and they would each have to be screened with the appropriate markers. For this study it was therefore decided to use the most significant regression models based on the F-ratio values, rather than continue the model fitting until all the significant variation in the traits was accounted for. This approach resulted in only three or four polymorphisms being selected for each of the three traits, with an associated drop in the r^2 values (the r^2 values for the three traits in this study ranged from 0.314 to 0.401). The use of only three or four markers in a selection programme is a far more feasible option than having to use up to 32 markers to screen for a trait. Additional analyses may provide markers with even higher associations with the traits, thereby allowing some of the markers in the models to be replaced and resulting in a more suitable r^2. However, this will be discussed in more detail later.

In their study on rice, Virk and co-workers (1996) demonstrated the predictive power of the multiple regression analysis in two ways. They first calibrated the regression models on 40 of the accessions, and then predicted the trait performances of the remaining eight accessions. The second approach they used was the "leave out one at a time" method whereby each

accession in turn was predicted based on the regression analysis of the remaining 47 accessions. This resulted in the fitting of 48 models, each with 47 accessions. By considering the 48 models as a whole, this approach can be seen as the equivalent of using a model calibrated from all 48 accessions to predict 48 "new" accessions. Of the two approaches, this will provide a more accurate picture of the actual performance of the markers than predicting eight accessions based on only 40. For this study on sugarcane varieties, it was therefore decided to use only the "leave out one at a time" approach to determine the predictive potential of the various polymorphisms.

It is interesting to note that some of the polymorphisms that were selected according to the correlation analysis discussed in the preceding section were also considered to be associated with the traits according to the multiple regression analysis. $OF14_{801}$ appeared in the four-polymorphism model for eldana resistance (Table 5.5, p. 81), and was the single polymorphism most highly correlated with eldana resistance (r^2 of 0.156). According to the correlation analysis, 100% of the eldana resistant varieties and 30% of the eldana susceptible varieties were positive for this polymorphism (Table 5.3, p. 77). $OD17_{1109}$ and $OD11_{1085}$ both appeared in the four-polymorphism model for SCMV resistance (Table 5.11, p. 89). The former was also selected by the correlation analysis as being associated with SCMV resistance (Table 5.3, p. 77), while the latter was selected as a "high stringency" polymorphism which, when used together with another polymorphism, appeared to be correlated with SCMV resistance (Table 5.4, p. 79). $OF01_{653}$ appeared in the three-polymorphism model for smut resistance (Table 5.16, p. 96), and was the single polymorphism most highly correlated with smut resistance according to the stepwise regression (r^2 of 0.120). $OF06_{618}$ was also in the three-polymorphism model for smut resistance, and both of these polymorphisms were considered, as a pair, to be associated with smut resistance in the correlation analysis (Table 5.4, p. 79).

In the regression models that were selected for each of the traits, there were no instances of one polymorphism being associated with more than one trait, as was found in the correlation analysis. This tends to be a more credible situation, since with three such diverse traits as resistance to an insect pest, resistance to a fungal pathogen, and resistance to a viral pathogen,

it is unlikely that any of the traits will have genes in common. However, if subsequent analysis revealed that a polymorphism was indeed found to be linked to two traits, an explanation could be that genes for each trait are very tightly linked to each other as well as to the particular polymorphism. Resistance genes in several plant species are often clustered in the genome, either as multi-allelic series at a locus or as multiple tightly linked loci (Michelmore, 1995). While some of these clusters involve different resistance genes for one particular pathogen, some genes within a cluster can bestow resistance to a variety of diverse pathogens or parasites. For example, the largest cluster of resistance genes in lettuce (*Lactuca sativa*) contains at least eight downy mildew (*Bremia lactucae*) resistance genes plus a gene for root-aphid resistance; while the second largest cluster contains at least two downy mildew resistance genes, a gene providing resistance against turnip mosaic virus, and a gene conferring resistance to root downy mildew (*Plasmopara lactucae-radicis*) (Witsenboer *et al.*, 1995).

As with the correlation analysis, the random nature of the RAPD polymorphisms is demonstrated by the apparent association of different polymorphisms amplified from the same primer with different traits. For example, $OD14_{1316}$ and $OD14_{926}$ appear in the predictive models for eldana resistance and smut resistance, respectively. If these two polymorphisms are, indeed, linked to the respective traits, an indication is provided how the RAPD technique amplifies regions of DNA without any regard for the actual sequence (except for the primer binding sites) and, therefore, purpose of the DNA.

The origins of some of the polymorphisms in the predictive models for the three traits can be traced to either *S. spontaneum* or *S. officinarum*, or to both species. For both eldana resistance (using the 1995 ratings) and smut resistance, one polymorphism in each of the models was *S. spontaneum*-specific while the remaining polymorphisms in the respective models were not detected in either *S. spontaneum* or *S. officinarum* (Table 5.5, p. 81; Table 5.16, p. 96). This could indicate that both of these traits were contributed to sugarcane varieties by *S. spontaneum*. It must be remembered, however, that the clones used to make up the *S. spontaneum* and *S. officinarum* DNA bulks for RAPD screening did not necessarily include all the clones used in the ancestry of the 50 sugarcane varieties. A trait that appears to be *S. spontaneum*-specific in this analysis may, therefore, be present in *S. officinarum* as well. This

circumstance is illustrated by the predictive model for eldana resistance that was obtained from the updated resistance ratings (Table 5.8, p. 85). According to this model, one polymorphism was *S. officinarum*-specific, one was present in both *S. officinarum* and *S. spontaneum*, while the remaining two were absent in both species. This disparity of the two models could be attributable to the probable polygenic nature of eldana resistance. If a number of factors (for example, fibre content, wax and flavonoid components) contribute to the observed levels of eldana resistance in sugarcane varieties, it is possible that some of the factors originate from *S. spontaneum* while others originate from *S. officinarum*. If this is the case, the different polymorphisms in the two eldana resistance models may be associated with the different resistance factors.

For SCMV resistance, one polymorphism in the predictive model was *S. spontaneum*-specific, one was *S. officinarum*-specific, one was present in both species, while one was absent in both species (Table 5.11, p. 89). This could indicate that SCMV resistance originated from both of the main progenitors of sugarcane varieties. It should be noted, however, that with r^2 values for the three traits ranging from 0.314 to 0.401, the polymorphisms do not explain enough of the variation in the traits for any definite deductions to be drawn regarding the contribution of traits by either *S. officinarum* or *S. spontaneum* to the germplasm of modern sugarcane varieties.

A general idea of the accuracy of the predictions by the regression models for each of the traits can be obtained by comparing the predicted ratings with the actual resistance ratings, as determined by the "leave out one at a time" approach (Table 5.6, p. 82; Table 5.12, p. 90; Table 5.17, p. 97). However, emphasis must be placed on the variation in the observed resistance that can occur due to the environmental conditions under which the sugarcane varieties are grown. The actual ratings that were assigned to the varieties for the three traits were obtained after the assessment of the performance of the varieties in numerous field trials. The actual ratings must therefore be considered as the average performance of the varieties, and while a variety may have a rating of, for example, 7, the resistance that will be observed in the field under various growing conditions may vary slightly on either side of this value. Consequently, a predicted rating of 5 for the variety in the above example should not

necessarily be regarded as inaccurate.

An illustration of the variation that can occur with the apparent resistance of the varieties is provided by the eldana resistance ratings. The predicted ratings for the 50 varieties were obtained by multiple regression analysis using the actual ratings that were currently available (listed in Table 5.6 as "Actual ratings (1995)", p. 82). These ratings were subsequently updated due to the results of further trials (listed in Table 5.6 as "Actual ratings (1996)"). Many of these ratings varied slightly from the original list, thus the four-polymorphism predictive model was calculated for these updated ratings as well. In comparison to the original predictive model, the updated model selected two new polymorphisms to account for the variation in the ratings (Table 5.5, p. 81; Table 5.8, p. 85). This provides an indication that a predictive model will only be as good as the values from which it has been calibrated. With the variability that occurs for the phenotype of sugarcane, the most accurate phenotypic values will be those that have been determined from replicated field trials representing the range of growing conditions that can occur. This indicates the usefulness of finding markers that are associated with specific traits in sugarcane varieties, rather than a progeny derived from a particular cross. If a progeny was to be used for the identification of markers, phenotypic data would have to be obtained from one or two trials using single stools. This would not necessarily be an accurate reflection of the overall performance of the individuals in the progeny.

To determine the accuracy of the predicted ratings in terms of their applicability in a sugarcane selection programme, various groups of resistance ratings were considered. The first approach involved three groups: resistant (ratings of 1 to 3), intermediate (ratings of 4 to 6) and susceptible (ratings of 7 to 9). The second approach involved two groups: resistant (ratings of 1 to 4) and susceptible (ratings of 5 to 9). In order for a predictive model to be used in a selection programme, the classification of unknown varieties into the various groups should be as accurate as possible. Both the r^2 of the model (the proportion of variability explained by the model) and the number of groups used to classify the individuals will affect the accuracy of the classification by a predictive model (Schenk & Westerhaus, 1993). The general relationship between r^2 and the percentage of times an individual is correctly classified

into a group is shown in Table 6.1 and Figure 6.1.

Table 6.1 Relationship between r^2 and percentage of times an individual is correctly classified by a regression model (from Schenk & Westerhaus, 1993)

r^2	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
correctly identified in 2 groups (gps)[a] (%)	61.4	64.8	68.3	71.8	75.0	78.2	81.7	85.2	89.8
correctly identified in 3 gps (%)	43.2	47.1	51.0	54.9	59.0	63.1	68.4	73.6	81.4
correctly identified in exact or adjacent gp (of 3 gps) (%)	86.0	89.0	92.0	94.0	96.0	97.0	99.0	100	100

[a] Groups = categories used to classify the individuals (e.g. resistant and susceptible)

For example, if individuals being analysed were to be divided into resistant and susceptible groups using a model with an r^2 of 0.20, from Table 6.1 it can be seen that 64.8% of the individuals would be correctly classified. If individuals were to be split into three groups using a model with an r^2 of 0.60, 63.1% of the individuals would be correctly classified. In general, classification will be more accurate when the r^2 is higher and the number of groups is fewer (Schenk & Westerhaus, 1993).

When the resistance ratings were divided into three groups (resistant, intermediate and susceptible), the predictive model for eldana resistance identified 30% of the varieties in the correct groups while the remaining 70% of the varieties were classified in the adjacent group (Table 5.9, p. 87). This model had an r^2 of 0.347. Thus, by interpolation of the data in Table 6.1, 53% of the varieties should have been correctly classified, 40% of the varieties should have been classified in the adjacent group, while 7% of the varieties should have been classified in the opposite group (i.e predicted as resistant when they were susceptible, and vice versa). However, the predictions were only based on 50 varieties. If a much larger group had been used, it is possible that values closer to those suggested by Table 6.1 would have been obtained.

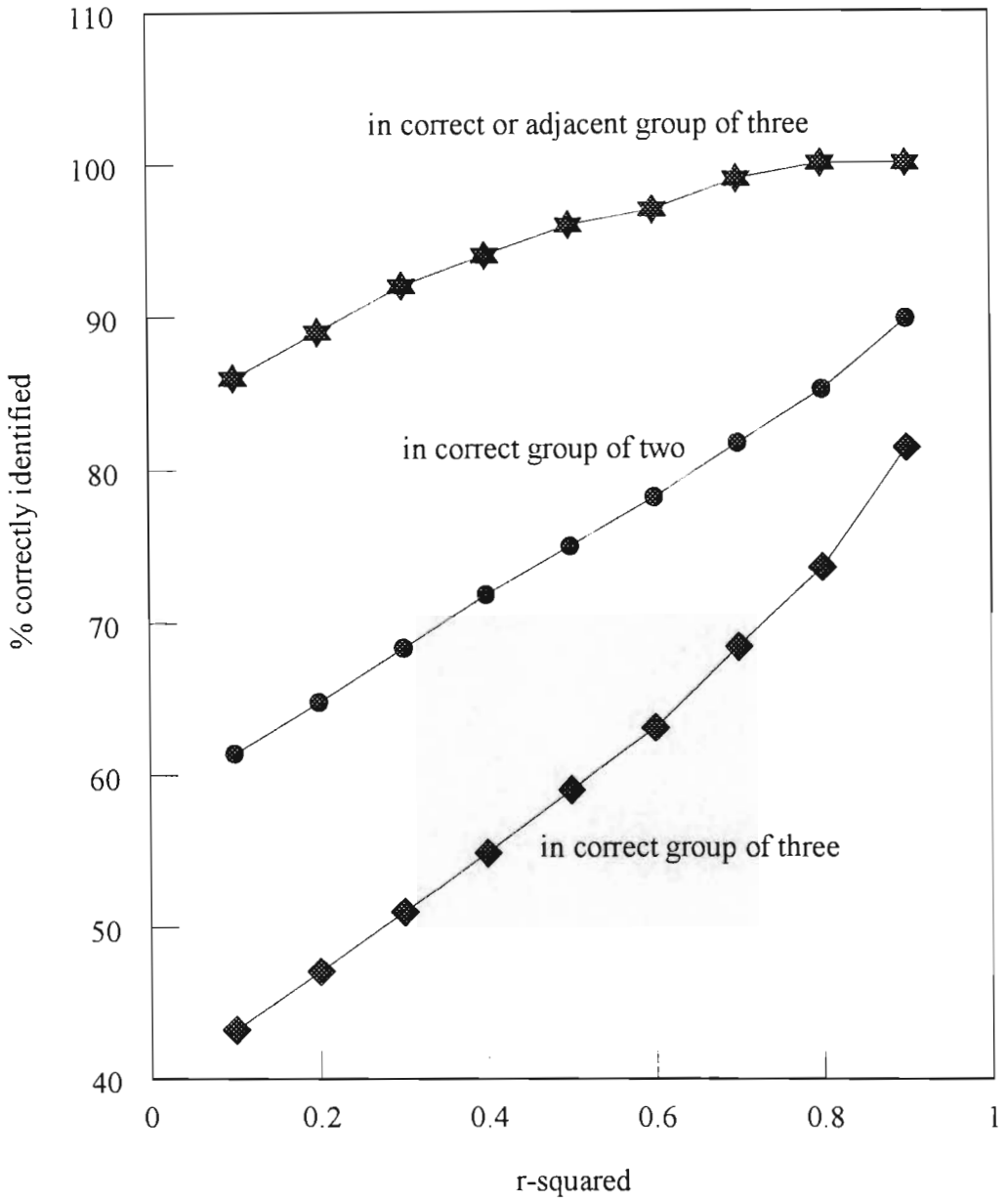


Figure 6.1 Relationship between r^2 and percentage of times an individual is correctly classified by a regression model (from data presented in Schenk & Westerhaus, 1993).

When the resistance ratings were divided into two groups (resistant and intermediate), the predictive model for eldana resistance identified 76% of the varieties in the correct group, while 24% of the varieties were identified in the opposite group (Table 5.10, p. 87). These values are relatively close to the values suggested by Table 6.1 for a model with an r^2 of 0.347 (70% in the correct group and 30% in the opposite group).

The predictive model for SCMV resistance had an r^2 of 0.401. According to Table 6.1, if three groups of resistance ratings were used, 54.9% of the varieties should have been classified in the correct group, 39.1% of the varieties should have been classified in the adjacent group, while 6% of the varieties should have been classified in the opposite group. The model for SCMV resistance correctly identified 52% of the varieties, while 46% of the varieties were identified in the adjacent group and only 2% of the varieties were identified in the opposite group (Table 5.14, p. 93). These values are relatively close to those suggested by Table 6.1, considering that only 50 varieties were used for the predictions.

When two groups of resistance ratings were used, the model for SCMV resistance predicted 74% of the varieties in the correct group and 26% of the varieties in the opposite group (Table 5.15, p. 93). These values are very similar to those suggested by Table 6.1 for a model with an r^2 of 0.401 (71.8% and 28.2%, respectively).

The r^2 of the predictive model for smut resistance was equal to 0.316. Thus, according to Table 6.1, 51.8%, 40.6% and 7.6% of the varieties should have been classified in the correct, adjacent, and opposite groups respectively. The model for smut resistance correctly identified 30% of the varieties, while 68% of the varieties were predicted in the adjacent group and 2% of the varieties were predicted in the opposite group (Table 5.19, p. 100). Once again, if a greater number of varieties had been used for the predictions, the values may have been closer to those suggested by Table 6.1.

When two groups of resistance ratings were used, the model for smut resistance predicted 70% of the varieties in the correct group and 30% of the varieties in the opposite group (Table 5.20, p. 100). These values are almost identical to those suggested by Table 6.1 for a model with

an r^2 of 0.316 (69% and 31%, respectively).

From the above interpretation of the results along with the data presented in Table 6.1, it becomes apparent that for markers to be useful in a selection programme, the r^2 values of the predictive models need not be as high as 0.80 or 0.90, as suggested by Virk *et al.* (1996). For example, if the resistant ratings were to be divided into three groups (resistant, intermediate and susceptible) and a predictive model for eldana resistance could be obtained with an r^2 of 0.50, 96% of the varieties would be predicted in either the correct or adjacent group (Table 6.1). Thus, only 4% of the varieties that are resistant would be predicted as susceptible, or vice versa. If only the varieties that get predicted as susceptible were to be discarded from the selection programme, 4% or less of the resistant varieties would be erroneously discarded. If marker-assisted selection was to be carried out in the early stages of the selection programme, it can then be argued that the few susceptible varieties that get carried through to successive stages, due to prediction as intermediate or resistant, could then get discarded using the conventional techniques of artificial inoculation and field trials.

6.4.3 Comparison of the Correlation Analysis With the Multiple Regression Analysis

These two methods were investigated in order to determine their relative efficiencies at identifying associations between the RAPD polymorphisms and the traits, and to compare the nature of the results obtained in each case. The correlation analysis was a very simple approach, based only on the presence and absence of both polymorphisms and the traits. The aim of this approach was to determine whether the presence of any polymorphisms implied the existence of a particular phenotypic state in the varieties. The multiple regression analysis, on the other hand, involved the statistical analysis of the polymorphisms and the traits in order to determine whether particular polymorphisms could be identified which would be able to predict the resistance ratings of the varieties. It must be remembered that only 54 polymorphisms were used for the analyses. It was therefore not expected to obtain particularly significant associations between the polymorphisms and traits, but rather to determine whether associations could, indeed, be found using sugarcane varieties, and to assess the potential of the different approaches at identifying genetic markers.

Both approaches appeared to identify associations between the polymorphisms and the three traits. However, the main disadvantage of the correlation analysis is that there is no direct means of determining the importance or strength of the different associations. Additional varieties could be screened with the potential markers, however this will merely support or reject the possibility that an association exists. Unless the results obtained from the correlation analysis are subsequently subjected to statistical analysis, a significance value cannot be placed on the apparent association. However, if the results from the correlation analysis need to be analysed statistically, it can be argued that a statistical approach may as well be used from the outset.

Unlike the correlation analysis, the multiple regression analysis identified those polymorphisms that showed associations with the traits, and assigned statistical values to the associations in the form of the r^2 and F-ratio values. The most significant predictive models could therefore be selected, and the predictive potential of the models determined using the "leave out one at a time" approach.

Another difference between the two methods is that the multiple regression analysis actually results in the prediction of the ratings of unknown varieties. In contrast, the correlation analysis will only establish whether a variety is likely to be "resistant" or "susceptible". Since the observed levels of resistance can range anywhere between the two extremes, the information content of the results from the correlation analysis is considerably lower than that of the results from the multiple regression.

From the above comparisons of each approach, the multiple regression analysis is far more informative than the correlation analysis. Some of the polymorphisms that were selected according to the correlation analysis were also considered to be associated with the traits when multiple regression was used. This indicates that the correlation approach is capable of identifying significant associations, however without any direct means of determining which of the associations are significant and which are not, the usefulness of this approach is relatively narrow in comparison to multiple regression analysis. The results that were obtained in this study, therefore, suggest that RAPD polymorphisms can be identified that appear to be

associated with specific traits in sugarcane varieties, and that multiple regression analysis has particular potential for identifying these associations.

6.5 Comments on the Use of Varieties for Marker Identification

The success of marker-assisted selection for the production of improved genotypes depends exclusively on the extent of genetic linkage between the markers and the relevant loci. Studies on genetic linkage are invariably based on populations derived from planned crosses. However, a recent study by Virk *et al.* (1996) demonstrated that diverse rice germplasm (i.e. accessions or varieties) could be used to identify associations between various quantitative traits and RAPD molecular markers. They argue that there is no reason why the principles that are applied to a segregating population cannot be applied to a collection of landraces or varieties, assuming that similar associations are observed between marker loci and the various allelomorphic forms of QTLs, and that the basis of these is genetic linkage.

Causes other than linkage may exist for the association between markers and phenotypes. These include linkage disequilibrium involving chance associations resulting from correlated allele frequencies in small samples (Virk *et al.*, 1996). However, Virk *et al.* (1996) claim that linkage between markers and QTLs is most likely to be responsible for the results that they obtained, although this can only be demonstrated unambiguously by analysing populations derived from single crosses between two parents. If linkage is responsible for the observed associations in the rice varieties, it would appear that linkage between alleles at QTLs and at marker loci has been conserved throughout the period of diversification of rice germplasm. This deduction can be extended to the sugarcane varieties used in this study. Indeed, given the narrow genetic background of sugarcane varieties, the low number of meioses since the first interspecific crosses (six or seven generations), and the apparent low rate of recombination in sugarcane (D'Hont *et al.*, 1994), most of the chromosomes initially contributed by the progenitors may have conserved the initial arrangements between markers and useful genes (Lu *et al.*, 1994a). If this is the case, the necessity to test markers derived from one population on other populations segregating for the same trait to determine the effects of genetic background is eliminated.

Another advantage of using a collection of varieties for identifying markers rather than a progeny derived from a specific cross, is that any trait can be studied in the same investigation. In this particular study, only three traits were investigated. However, additional traits, such as factors contributing to sucrose yield, could easily be included by performing multiple regression analysis on the new trait data with the same 54 polymorphisms that were used for the three resistance traits. Conventionally, separate populations segregating for each of the traits would have to be analysed independently, since it is unlikely that a single cross will yield a population segregating for each of the required traits.

6.6 Future Prospects for Marker Identification in Sugarcane Varieties

This study dealt with only 50 sugarcane varieties. Although this number may have been sufficient for an initial analysis to determine the potential of using varieties to identify markers, the list of varieties should be expanded if this approach is to be used for the identification of markers for marker-assisted selection. By increasing the number of varieties used, the extent of genetic variation that can occur for the traits will increase in the calibration set used to obtain a regression model. Thus the likelihood of coming across variation that cannot be accounted for in unknown varieties that are subjected to MAS will be decreased. In order to obtain a more accurate representation of the origins of the polymorphisms, the study should also be extended to include those clones of *S. officinarum* and *S. spontaneum* that were used in the ancestry of all the selected varieties.

The next step in marker identification would be to obtain more RAPD polymorphisms. In this study, a total of 41 RAPD primers were screened and from these, 54 reproducible polymorphisms were obtained. Although the four-polymorphism regression model was used for the prediction of eldana resistance, only one of the polymorphisms had a particularly significant fit to the eldana resistance data. This polymorphism, therefore, is the only one worth keeping in terms of eldana resistance. Similarly, for SCMV resistance and smut resistance, it would only be worth keeping the three and four polymorphisms that were identified in the respective predictive models for these two traits. It would therefore be strongly recommended to screen as many new RAPD primers as possible to obtain new

polymorphisms. Multiple regression analysis of those polymorphisms could then be carried out to identify new polymorphisms that could supplement, or possibly even replace, the potential markers identified in this study. This is likely to result in a much stronger association between the polymorphisms and the traits, indicated by higher r^2 values than those obtained in this study.

In order to efficiently screen as many new primers as possible, it would only be necessary for a subset of the varieties to be screened. However, if that subset of varieties were to be genetically similar, it is possible that some potentially important polymorphisms may appear monomorphic in that particular subset. It would thus be advisable to assemble together varieties that represent the range of diversity found within the whole group of varieties. The data on the various traits under analysis, obtained from all the varieties in the group, could be subjected to cluster analysis using a technique such as the unweighted pair group method using arithmetic averages (UPGMA), as described by Virk *et al.* (1996). This would result in the generation of a dendrogram, which could then be used for the selection of a subset of diverse varieties. The screening of a smaller number of varieties (for example, 25) with new primers would be far easier than screening the entire set. Those primers that produce clear, easily scorable polymorphisms could then be used for the amplification of the DNA from all the varieties, followed by the scoring of the polymorphisms.

An alternative to the above approach would be to use bulked segregant analysis to reduce the number of samples that need to be screened with each primer. BSA can be used to identify markers associated with quantitative traits (Michelmore *et al.*, 1991), in which case the individuals at the extremes of the distribution of the variation would be used to form the DNA bulks. For the traits analysed in this study, those varieties with resistance ratings of 1, 2 and 3 could form the resistant bulk, while varieties with ratings of 7, 8 and 9 could form the susceptible bulk. Alternatively, only those varieties with ratings of 1 and 2, and 8 and 9 could be used for the respective bulks, resulting in a more stringent analysis. BSA would thus result in the screening of only two DNA samples for each trait. Those primers that produce polymorphisms between the two bulks could then be screened against all the varieties to allow the polymorphisms to be scored for multiple regression analysis.

Due to the interaction which occurs between the genotype of sugarcane and the environmental conditions under which it was grown, extensive field trials over several years are required before the traits of sugarcane clones can be accurately evaluated. Consequently, there is a general lack of well characterized populations. An exception to this at SASEX is the AA40 population, consisting of 150 clones derived from a cross between the varieties N18 and CP57/614. This cross was made in 1990, and the clones are currently undergoing trials for trait evaluation. The AA40 population could provide a valuable means to evaluate any potential markers that are identified from sugarcane varieties. One approach would be to treat the clones in AA40 as unknowns, and allow any potential markers to predict the traits of the clones. The predicted trait values could then be compared with the actual trait values to determine the predictive potential of the markers. A second approach would be to conduct conventional linkage analysis of the markers in the population to determine the extent of genetic linkage of the markers to the traits.

Once polymorphisms have been found that can be used as markers for the selection of improved genotypes, these can be converted into SCARs (Paran & Michelmore, 1993) or ASAPs (Gu *et al.*, 1995). These PCR-based markers are more reliable than RAPDs due to the use of longer primers and more stringent annealing conditions than the RAPD assay. The advantage of ASAPs is that they are allele-specific, whereas SCAR primers sometimes amplify fragments that are linked to both alleles at the locus of interest. Since ASAP primers amplify only one fragment, the need to separate amplified DNA fragments by electrophoresis is eliminated. Samples possessing the appropriate allele can be identified in a quick plus/minus assay by direct staining of the DNA in the reaction tubes or microtitre plate with ethidium bromide (Gu *et al.*, 1995). This makes the technique applicable to the large-scale screening of samples that is necessary for MAS.

6.7 Marker-Assisted Selection in Sugarcane

The amount of effort expended in a selection programme to overcome susceptibility to a pest or disease should depend on the economic importance of that pest or disease. Testing for susceptibility to an important disease should be started as early in the selection programme as

possible in order to eliminate the more susceptible clones early in the testing schedule (Walker, 1987), thereby diminishing the numbers of clones that need to be taken into the field. However, in commercial practice sugarcane is a clonally propagated crop, and in order for a new variety to be able to occupy many thousands of hectares, years of vegetative multiplication are necessary following the initial sexual crossing of breeding material. Consequently, screening for pest and disease resistance, apart from the discarding of obviously susceptible material, is delayed until sufficient clonal material is available for replicated field trials.

The inability to reliably test clones at an early selection stage constitutes one of the factors that contribute towards the need for alternative means of screening for pest and disease resistance. Other factors include the variability in aggressiveness or availability of a pest or pathogen under natural conditions, the requirement to conduct some screens at particular locations or times of year, and the impracticality of the simultaneous or sequential screening of clones with several pathogens when artificial inoculation trials are used. For example, smut is mainly a problem in the northern irrigated areas of the South African sugarcane belt, and as a consequence, smut resistance trials are restricted to these areas. Another example involves eldana, where low levels of the pest are generally observed during years of high rainfall, thus hampering field (naturally infested) trials.

From the above considerations it becomes apparent that MAS will have particular utility in a sugarcane selection programme. The detection of resistance genes on the basis of their linkage to markers can be carried out at the seedling stage, as soon as the plants are large enough to yield sufficient DNA. The screen is independent of any environmental conditions, can be carried out in the absence of the pathogen or pest, and resistance to multiple pathogens can be carried out simultaneously.

Although this discussion of MAS in sugarcane has concentrated on its applicability for selecting disease or pest resistance, it should be noted that it is not limited to resistance traits. Any phenotypic characteristics that are considered during the selection programme, from sucrose content to ratooning ability, can be subjected to MAS, provided that appropriate markers are found for each trait. MAS should not replace traditional selection methods, but

should rather supplement them in the development of new and better varieties. While markers can be used to select plants with desirable genes from among the many seedlings in the initial stages of selection, the performance of these plants should be evaluated over a wide range of environmental conditions at a later stage in the field to ensure that the appropriate genes are still present and effective.

The extent to which MAS will be used will depend on the relative cost and expected return of this approach in comparison to conventional selection, and it will probably be the most useful for traits that are either expensive or difficult to evaluate by conventional methods. Such traits can be decided upon in collaboration with those involved in sugarcane breeding and selection. At this stage, it can be decided what proportion of the variability of the trait needs to be accounted for by the markers in order for MAS to be successful; if multiple regression is to be used, this would be determined by the r^2 value. The aim would then be to identify as few markers as possible to attain the necessary level of association with the trait. Once markers have been identified, the cost of using them in the sugarcane selection programme can be determined. Only if the expected return of using MAS instead of conventional selection exceeds the relative cost of the new technology, will MAS be a feasible option.

CURRICULUM VITAE

Julie Megan Barnes was born on 30 May 1972 in Stellenbosch.

1990: Matriculated at Eshowe High School with an "A" aggregate.

1993: Obtained a Bachelor of Science degree at Rhodes University, Grahamstown, with majors in Botany and Microbiology.

1994: Obtained a Bachelor of Science (Honours) degree in Microbiology, with distinction, at Rhodes University, Grahamstown.

1995: Enrolled for a Masters of Science degree in Plant Molecular Biology at the University of Natal, Pietermaritzburg.

She has presented one paper and one poster at national conferences.

REFERENCES

Ajmone Marsan P, Egidy G, Monfredini G, Di Silvestro S & Motto M (1993) RAPD markers in maize genetic analysis. *Maydica* **38**: 259-264.

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA & Cregan PB (1995) Integration of simple sequence repeat DNA markers into a soybean linkage map. *Crop Science* **35**: 1439-1445.

Al-Janabi SM, Honeycutt RJ, McClelland M & Sobral BWS (1993) A genetic linkage map of *Saccharum spontaneum* (L.) 'SES 208'. *Genetics* **134**: 1249-1260.

Al-Janabi SM, Honeycutt RJ & Sobral BWS (1994) Chromosome assortment in *Saccharum*. *Theoretical and Applied Genetics* **89**: 959-963.

Bai D, Reeleder R & Brandle JE (1995) Identification of two RAPD markers tightly linked with the *Nicotiana debneyi* gene for resistance to black root rot of tobacco. *Theoretical and Applied Genetics* **91**: 1184-1189.

Beckmann JS & Soller M (1986) Restriction fragment length polymorphisms and genetic improvement of agricultural species. *Euphytica* **35**: 111-124.

Berding N & Roach BT (1987) Germplasm Collection, Maintenance, and Use. In: DJ Heinz (ed.) *Developments in Crop Science 11: Sugarcane Improvement Through Breeding*. Elsevier, Amsterdam, p. 143-210.

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

Bonierbale MW, Plaisted RL & Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* **120**: 1095-1103.

Botstein D, White RL, Skolnik M & Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**: 314-331.

Bremer G (1961) Problems in breeding and cytology of sugar cane. I. A short history of sugar cane breeding - The original forms of *Saccharum*. *Euphytica* **10**: 59-78.

Bubeck DM, Goodman MM, Beavis WD & Grant D (1993) Quantitative trait loci controlling resistance to gray leaf spot in maize. *Crop Science* **33**: 838-847.

Büscher N, Zyprian E, Bachmann O & Blaich R (1994) On the origin of the grapevine variety Müller-Thurgau as investigated by the inheritance of random amplified polymorphic DNA (RAPD). *Vitis* **33**: 15-17.

Caetano-Anollés G & Bassam BJ (1993) DNA amplification fingerprinting using arbitrary oligonucleotide primers. *Applied Biochemistry and Biotechnology* **42**: 189-200.

Caetano-Anollés G, Bassam BJ & Gresshoff PM (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/Technology* **9**: 553-557.

Caetano-Anollés G, Bassam BJ & Gresshoff PM (1993) Enhanced detection of polymorphic DNA by multiple arbitrary amplicon profiling of endonuclease-digested DNA: identification of markers tightly linked to the supernodulation locus in soybean. *Molecular and General Genetics* **241**: 57-64.

Chang C, Bowman JL, DeJohn AW, Lander ES & Meyerowitz EM (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences USA* **85**: 6856-6860.

Cregan PB, Akkaya MS, Bhagwat AA, Lavi U & Rongwen J (1994) Length Polymorphisms of Simple Sequence Repeat (SSR) DNA as Molecular Markers in Plants. In: PM Gresshoff (ed.) *Plant Genome Analysis*. CRC Press, Boca Raton, FL, p. 47-56.

Daniels J & Roach BT (1987) Taxonomy and Evolution. In: DJ Heinz (ed.) *Developments in Crop Science 11: Sugarcane Improvement Through Breeding*. Elsevier, Amsterdam, p. 7-84.

Da Silva JAG, Honeycutt RJ, Burnquist W, Al-Janabi SM, Sorrells ME, Tanksley SD & Sobral BWS (1995) *Saccharum spontaneum* L. 'SES 208' genetic linkage map combining RFLP- and PCR-based markers. *Molecular Breeding* **1**: 165-179.

Da Silva JAG & Sobral BWS (1996) Genetics of Polyploids. In: BWS Sobral (ed.) *The Impact of Plant Molecular Genetics*. Birkhäuser, Boston, p. 3-37.

Da Silva JAG, Sorrells ME, Burnquist WL & Tanksley SD (1993) RFLP linkage map and genome analysis of *Saccharum spontaneum*. *Genome* **36**: 782-791.

Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, Glaszmann JC & D'Hont A (1996) A putative major gene for rust resistance linked with an RFLP marker in sugarcane cultivar R570. *Theoretical and Applied Genetics* **92**: 1059-1064.

Davis TM, Yu H, Haigis KM & McGowan PJ (1995) Template mixing: a method of enhancing detection and interpretation of codominant RAPD markers. *Theoretical and Applied Genetics* **91**: 582-588.

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N & Glaszmann JC (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics* **250**: 405-413.

D'Hont A, Lu Y-H, González de Léon D, Grivet L, Feldmann P, Lanaud C & Glaszmann JC (1994) A molecular approach to unravelling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. *Genome* **37**: 222-230.

Diers BW, Mansur L, Imsande J & Shoemaker RC (1992) Mapping phytophthora resistance loci in soybean with restriction fragment length polymorphism markers. *Crop Science* **32**: 377-383.

Dion Y, Gugel RK, Rakow GFW, Séguin-Swartz G & Landry BS (1995) RFLP mapping of resistance to the blackleg disease [causal agent, *Leptosphaeria maculans* (Desm.) Ces. et de Not.] in canola (*Brassica napus* L.) *Theoretical and Applied Genetics* **91**: 1190-1194.

Dudley JW (1993) Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Science* **33**: 660-668.

Foolad MR, Jones RA & Rodriguez RL (1993) RAPD markers for constructing intraspecific tomato genetic maps. *Plant Cell Reports* **12**: 293-297.

Fraser RSS (1990) The genetics of resistance to plant viruses. *Annual Reviews of Phytopathology* **28**: 179-200.

Ganal MW, Simon R, Brommonschenkel S, Arndt M, Phillips MS, Tanksley SD & Kumar A (1995) Genetic mapping of a wide spectrum nematode resistance gene (*Hero*) against *Globodera rostochiensis* in tomato. *Molecular Plant-Microbe Interactions* **8**: 886-891.

Gebhardt CE, Ritter E, Debener T, Schachtschabel U, Walkemeier B, Urig H & Salamini F (1989) RFLP-analysis and linkage mapping in *Solanum tuberosum*. *Theoretical and Applied Genetics* **78**: 65-75.

Geiger HH & Heun M (1989) Genetics of quantitative resistance to fungal diseases. *Annual Reviews of Phytopathology* **27**: 317-341.

Glaszmann JC, Fautret A, Noyer JL, Feldmann P & Lannaud C (1989) Biochemical genetic markers in sugarcane. *Theoretical and Applied Genetics* **78**: 537-543.

Goodman MM, Stuber CW, Newton K & Weissinger HH (1980) Linkage relationships of 19 enzyme loci in maize. *Genetics* **96**: 697-710.

Grivet L, D'Hont A, Dufour P, Hamon P, Roques D & Glaszmann JC (1994) Comparative genome mapping of sugar cane with other species within the Andropogoneae tribe. *Heredity* **73**: 500-508.

Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C & Glaszmann JC (1996) RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* **142**: 987-1000.

Gu WK, Weeden NF, Yu J & Wallace DH (1995) Large-scale, cost-effective screening of PCR products in marker-assisted selection applications. *Theoretical and Applied Genetics* **91**: 465-470.

Haley SD, Afanador LK, Miklas PN, Stavely JR & Kelly JD (1994a) Heterogeneous inbred populations are useful as sources of near-isogenic lines for RAPD marker localization. *Theoretical and Applied Genetics* **88**: 337-342.

Haley SD, Miklas PN, Afanador L & Kelly JD (1994b) Random amplified polymorphic DNA (RAPD) marker variability between and within gene pools of common bean. *Journal of the American Society for Horticultural Science* **119**: 122-125.

Harvey M & Botha FC (1996) Use of PCR-based methodologies for the determination of DNA diversity between *Saccharum* varieties. *Euphytica* **89**: 257-265.

Harvey M, Huckett BI & Botha FC (1994) Use of the polymerase chain reaction (PCR) and random amplification of polymorphic DNAs (RAPDs) for the determination of genetic distances between 21 sugarcane varieties. *Proceedings of The South African Sugar Technologists' Association* **68**: 36-40.

Hogarth DM (1987) Genetics of Sugarcane. In: DJ Heinz (ed.) *Developments in Crop Science 11: Sugarcane Improvement Through Breeding*. Elsevier, Amsterdam, p. 255-271.

Hogarth DM, Ryan CC & Taylor PWJ (1993) Quantitative inheritance of rust resistance in sugarcane. *Field Crops Research* **34**: 187-193.

Honeycutt RJ, Sobral BWS, Keim P & Irvine JE (1992) A rapid extraction method for sugarcane and its relatives. *Plant Molecular Biology Reporter* **10**: 66-72.

Horvath DP, Dahleen LS, Stebbing J & Penner G (1995) A co-dominant PCR-based marker for assisted selection of durable stem rust resistance in barley. *Crop Science* **35**: 1445-1450.

Huckett BI & Botha FC (1995) Stability and potential use of RAPD markers in a sugarcane genealogy. *Euphytica* **86**: 117-125.

Hulbert SH, Richter TE, Axtell JD & Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proceedings of the National Academy of Sciences USA* **87**: 4251-4255.

Jin Y, Statler GD, Franckowiak JD & Steffenson BJ (1993) Linkage between leaf rust resistance genes and morphological markers in barley. *Phytopathology* **83**: 230-233.

Johnston CF, Fielding WJ, Talbot, M & Bennett-Easy M (1995) The variability of varieties of sugar cane (*Saccharum* spp., comm. hybrids). *Tropical Agriculture (Trinidad)* **72**: 1-6.

Jourdren C, Barret P, Horvais R, Foisset N, Delourme R & Renard M (1996) Identification of RAPD markers linked to the loci controlling erucic acid level in rapeseed. *Molecular Breeding* **2**: 61-71.

Kaya Z & Neale DB (1995) Utility of random amplified polymorphic DNA (RAPD) markers for linkage mapping in Turkish red pine (*Pinus brutia* Ten.) *Silvae Genetica* **44**: 110-116.

Kelly JD (1995) Use of random amplified polymorphic DNA markers in breeding for major gene resistance to plant pathogens. *HortScience* **30**: 461-465.

Kishimoto N, Foolad MR, Shimosaka E, Matsuura S & Saito A (1993) Alignment of molecular and classical linkage maps of rice, *Oryza sativa*. *Plant Cell Reports* **12**: 457-461.

Kjær B, Jensen J & Giese H (1995) Quantitative trait loci for heading date and straw characters in barley. *Genome* **38**: 1098-1104.

Klein-Lankhorst RM, Vermunt A, Weide R, Liharska T & Zabel P (1991) Isolation of molecular markers for tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD). *Theoretical and Applied Genetics* **83**: 108-114.

Kutcher HR, Bailey KL, Rossnagel BG & Legge WG (1996) Identification of RAPD markers for common root rot and spot blotch (*Cochliobolus sativus*) resistance in barley. *Genome* **39**: 206-215.

Lander ES & Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185-199.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE & Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174-181.

Landry BS, Kesseli RV, Farrara B & Michelmore RW (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. *Genetics* **116**: 331-337.

Lawson WR, Goulter KC, Henry RJ, Kong GA & Kochman JK (1996) RAPD markers for a sunflower rust resistance gene. *Australian Journal of Agricultural Research* **47**: 395-401.

Lebreton C, Lazić-Jančić V, Steed A, Pekić S & Quarrie SA (1995) Identification of QTL for drought responses in maize and their use in testing causal relationships between traits. *Journal of Experimental Botany* **46**: 853-865.

Leon AJ, Lee M, Rufener GK, Berry ST & Mowers RP (1995) Use of RFLP markers for genetic linkage analysis of oil percentage in sunflower seed. *Crop Science* **35**: 558-564.

Li J & Niwa M (1996) Microsatellite DNA markers linked to a gene controlling days to flowering in soybean (*Glycine max*) under shortday conditions. *Breeding Science* **46**: 81-84.

Lin J-Z & Ritland K (1996) Construction of a genetic linkage map in the wild plant *Mimulus* using RAPD and isozyme markers. *Genome* **39**: 63-70.

Lodhi MA, Daly MJ, Ye G-N, Weeden NF & Reisch BI (1995) A molecular marker based linkage map of *Vitis*. *Genome* **38**: 786-794.

Lu YH, D'Hont A, Paulet F, Grivet L, Arnaud M & Glaszmann JC (1994a) Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* **78**: 217-226.

Lu YH, D'Hont A, Walker DIT, Rao PS, Feldmann P & Glaszmann JC (1994b) Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. *Euphytica* **78**: 7-18.

Martin GB, Williams JGK & Tanksley SD (1991) Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proceedings of the National Academy of Sciences USA* **88**: 2336-2340.

Mehlenbacher SA (1995) Classical and molecular approaches to breeding fruit and nut crops for disease resistance. *HortScience* **30**: 466-477.

Michelmore R (1995) Molecular approaches to manipulation of disease resistance genes. *Annual Reviews of Phytopathology* **15**: 393-427.

Michelmore RW, Paran I & Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences USA* **88**: 9828-9832.

Msomi N & Botha FC (1994) Identification of molecular markers linked to fibre using bulk segregant analysis. *Proceedings of The South African Sugar Technologists' Association* **68**: 41-45.

Nair S, Bentur JS, Prasada Rao U & Mohan M (1995) DNA markers tightly linked to a gall midge resistance gene (*Gm2*) are potentially useful for marker-aided selection in rice breeding. *Theoretical and Applied Genetics* **91**: 68-73.

Novy RG & Vorsa N (1996) Evidence for RAPD heteroduplex formation in cranberry: implications for pedigree and genetic-relatedness studies and a source of co-dominant RAPD markers. *Theoretical and Applied Genetics* **92**: 840-849.

Nuss KJ, Bond RS & Atkinson PR (1986) Susceptibility of sugarcane to the borer *Eldana saccharina* Walker and selection for resistance. *Proceedings of The South African Sugar Technologists' Association* **60**:153-155.

O'Brien SJ (ed.) (1990) *Genetic Maps: Locus Maps of Complex Genomes*, fifth edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbour, New York.

Ohmori T, Murata M & Motoyoshi F (1995) RAPD markers linked to the tomato mosaic virus resistance gene, *Tm-1*, in tomato. *Japanese Journal of Genetics* **70**: 179-184.

Ohmori T, Murata M & Motoyoshi F (1996) Molecular characterization of RAPD and SCAR markers linked to the *Tm-1* locus in tomato. *Theoretical and Applied Genetics* **92**: 151-156.

Ordon F, Bauer E, Friedt W & Graner A (1995) Marker-based selection for the *ym4* BaMMV-resistance gene in barley using RAPDs. *Agronomie* **15**: 481-485.

Orozco-Castillo C, Chalmers KJ, Waugh R & Powell W (1994) Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theoretical and Applied Genetics* **87**: 934-940.

Paran I & Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics* **85**: 985-993.

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES & Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics* **127**:181-197.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE & Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**: 721-726.

Plomion C, Bahrman N, Durel C-E & O'Malley DM (1995) Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. *Heredity* **74**: 661-668.

Poulsen DME, Henry RJ, Johnston RP, Irwin JAG & Rees RG (1995) The use of bulk segregant analysis to identify a RAPD marker linked to leaf rust resistance in barley. *Theoretical and Applied Genetics* **91**: 270-273.

Price S (1967) Interspecific hybridization in sugarcane breeding. *Proceedings of the International Society of Sugar Cane Technologists* **12**: 1021-1026.

Qi L, Cao M, Chen P, Li W & Liu D (1996) Identification, mapping, and application of polymorphic DNA associated with resistance gene *Pm21* of wheat. *Genome* **39**: 191-197.

Rafalski JA, Hanafey MK, Tingey SV & Williams JGK (1994) Technology for Molecular Breeding: RAPD Markers, Microsatellites and Machines. In: PM Gresshoff (ed.) *Plant Genome Analysis*. CRC Press, Boca Raton, FL, p.19-27.

Rajapakse S, Belthoff LE, He G, Estager AE, Scorza R, Verde I, Ballard RE, Baird WV, Callahan A, Monet R & Abbott AG (1995) Genetic linkage mapping in peach using morphological, RFLP and RAPD markers. *Theoretical and Applied Genetics* **90**: 503-510.

Redoña ED & Mackill DJ (1996a) Mapping quantitative trait loci for seedling vigor in rice using RFLPs. *Theoretical and Applied Genetics* **92**: 395-402.

Redoña ED & Mackill DJ (1996b) Molecular mapping of quantitative trait loci in *japonica* rice. *Genome* **39**: 395-403.

Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV & Scolnik PA (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proceedings of the National Academy of Sciences USA* **89**: 1477-1481.

Rutherford RS, Meyer JH & van Staden J. (1995) NIR prediction of resistance to the sugarcane borer *Eldana saccharina* using stalk surface wax and budscale extracts. In: GD Batten, PC Flinn, LA Welsh & AB Blakeney (eds), *Leaping ahead with near-infrared spectroscopy*. NIR Spectroscopy Group, North Melbourne, p. 222-226.

Rutherford RS & van Staden J (1996) Towards a rapid near-infrared technique for prediction of resistance to sugarcane borer *Eldana saccharina* Walker (Lepidoptera: Pyralidae) using stalk surface wax. *Journal of Chemical Ecology* **22**: 681-694.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB & Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA & Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350-1354.

Saito A, Yano M, Kishimoto N, Nakagahra M, Yoshimura A, Saito K, Kuhara S, Ukai Y, Kawase M, Nagamine T, Yoshimura S, Ideta O, Ohsawa R, Hayano Y, Iwata N & Sugiura M (1991) Linkage map of restriction fragment length polymorphism loci in rice. *Japanese Journal of Breeding* **41**: 665-670.

Sambrook J, Fritsch EF & Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbour, New York.

SAS (1990) *SAS/STAT User's Guide*, Version 6, fourth edition, volume 1. SAS Institute Inc., Cary, NC.

Schenk JS & Westerhaus MO (1993) *Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy*. Monograph, Infrasoft International, Port Matilda, PA.

Schön CS, Lee M, Melchinger AE, Guthrie WD & Woodman WL (1993) Mapping and characterization of quantitative trait loci affecting resistance against second-generation European corn borer in maize with the aid of RFLPs. *Heredity* **70**: 648-659.

Shoemaker RC & Specht JE (1995) Integration of the soybean molecular and classical genetic linkage groups. *Crop Science* **35**: 436-446.

Sills GR, Bridges W, Al-Janabi SM & Sobral BWS (1995) Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes & Jesw. ex Grassl). *Molecular Breeding* **1**: 355-368.

Skinner JC, Hogarth DM & Wu KK (1987) Selection Methods, Criteria, and Indices. In: DJ Heinz (ed.) *Developments in Crop Science 11: Sugarcane Improvement Through Breeding*. Elsevier, Amsterdam, p. 409-453.

Sobral BWS & Honeycutt RJ (1993) High output genetic mapping of polyploids using PCR-generated markers. *Theoretical and Applied Genetics* **86**: 105-112.

Sreenivasan TV, Ahloowalia BS & Heinz DJ (1987) Cytogenetics. In: DJ Heinz (ed.) *Developments in Crop Science 11: Sugarcane Improvement Through Breeding*. Elsevier, Amsterdam, p. 211-253.

Stevenson GC (1965) *Genetics and Breeding of Sugar Cane*. Longmans, London.

Tanksley SD, Bernatzky R, Lapitan NL & Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. *Proceedings of the National Academy of Sciences USA* **85**: 6419-6423.

Tanksley SD, Ganal MW, Prince JP, de Vincente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W & Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141-1160.

Tanksley SD & Orton TJ (eds) (1983) *Isozymes in Plant Genetics and Breeding, Part A*. Elsevier, Amsterdam.

Tanksley SD & Rick CM (1980) Isozyme linkage map of the tomato: applications in genetics and breeding. *Theoretical and Applied Genetics* **57**: 161-170.

Tanksley SD, Young ND, Paterson AH & Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. *Bio/Technology* **7**: 257-264.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 4127-4138.

Tingey SV & del Tufo JP (1993) Genetic analysis with random amplified polymorphic DNA markers. *Plant Physiology* **101**: 349-352.

Uphoff H & Wricke G (1995) A genetic map of sugar beet (*Beta vulgaris*) based on RAPD markers. *Plant Breeding* **114**: 355-357.

Vierling RA & Nguyen HT (1992) Use of RAPD markers to determine the genetic diversity of diploid, wheat genotypes. *Theoretical and Applied Genetics* **84**: 835-838.

Virk PS, Ford-Lloyd BV, Jackson MT, Pooni HS, Clemeno TP & Newbury HJ (1996) Predicting quantitative variation within rice germplasm using molecular markers. *Heredity* **76**: 296-304.

Walker DIT (1987) Breeding for Disease Resistance. In: DJ Heinz (ed.) *Developments in Crop Science 11: Sugarcane Improvement Through Breeding*. Elsevier, Amsterdam, p. 455-502.

Wang G, Castiglione S, Zhang J, Fu R, Ma J, Li W, Sun Y & Sala F (1994) Hybrid rice (*Oryza sativa* L.): identification and parentage determination by RAPD fingerprinting. *Plant Cell Reports* **14**: 112-115.

Waugh R & Powell W (1992) Using RAPD markers for crop improvement. *Trends in Biotechnology* **10**: 186-191.

Welsh J & McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**: 7213-7218.

Williams JGK, Hanafey MK, Rafalski JA & Tingey SV (1993) Genetic analysis using random amplified polymorphic DNA markers. *Methods in Enzymology* **218**: 704-740.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA & Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531-6535.

Winter P & Kahl G (1995) Molecular marker technologies for plant improvement. *World Journal of Microbiology & Biotechnology* **11**: 438-448.

Witsenboer H, Kesseli RV, Fortin MG, Stanghellini M & Michelmore RW (1995) Sources and genetic structure of a cluster of genes for resistance to three pathogens in lettuce. *Theoretical and Applied Genetics* **91**: 178-188.

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH and Tanksley SD (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theoretical and Applied Genetics* **83**: 294-300.

Yang H & Krüger J (1994) Identification of a RAPD marker linked to the *Vf* gene for scab resistance in apples. *Euphytica* **77**: 83-87.

Yang X & Quiros CF (1995) Construction of a genetic linkage map in celery using DNA-based markers. *Genome* **38**: 36-44.

Young RA & Kelly JD (1996) RAPD markers flanking the *Are* gene for anthracnose resistance in common bean. *Journal of the American Society for Horticultural Science* **121**: 37-41.

Young ND, Zamir D, Ganal MW & Tanksley SD (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato. *Genetics* **120**: 579-585.

Yu J, Gu WK, Provvidenti R & Weeden NF (1995) Identifying and mapping two DNA markers linked to the gene conferring resistance to pea enation mosaic virus. *Journal of the American Society for Horticultural Science* **120**: 730-733.