The potential of bulk segregant analysis and RAPD technology for identification of molecular markers linked to traits in sugarcane

by

Nhlanhla Sobantu Msomi

Submitted in fulfilment of the academic

requirements for the degree of

Doctor of Philosophy

in the

Department of Biology,

University of Natal

Durban

1998

PREFACE

The experimental work described in this thesis was carried out in the Biotechnology Department of the South African Sugar Association Experiment Station (SASEX), Mount Edgecombe under the supervision of Professor F.C. Botha.

These studies are the original work of 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 thanks to God, our Heavenly Father, from whom all things are derived.

My heartfelt gratitude to my supervisor, Prof. Frikkie Botha, for his advice, support, and patience throughout the duration of the study.

Thanks to the Foundation for Research Development (FRD), and the South African Sugar Association Experiment Station for their financial support.

I am indebted to the following people for their assistance and support:

The staff of the Plant Breeding Department (SASEX), especially Karl Nuss (Head) and Arnold van der Merwe, for assistance in the sampling of sugarcane clones and phenotypic assessment.

Murt Murdoch, Kerry Redshaw and Mike Butterfield for their assistance with statistical analyses, and the discussions regarding sugarcane quantitative genetics.

Sarita Groenewald and Julie Barnes for use of the $SA11_{640}$ RAPD fragment and DNA samples of different sugarcane varieties, respectively.

Mornay Harvey and Bernadette Goodman for technical assistance, proofreading of parts of the thesis and friendship.

All the members of the Biotechnology Department, past and present, for their generosity of spirit and friendship.

Mama for her love and wholehearted support at all times.

Louise for her strength, love, and keeping it real.

Sibusiso, Buyani, Thabani, Gala, Zamambo, Stefan, Joy, Anti, Vusi, Nozipho, Themba, Lungi, Jacqui, Thami, Moss, Fiona, Stuart, Tazz, Natalie, Ronnie, Barbara, Xola, Lee, Lulu, Sgele, Mathupha, Gugu, my whole extended family and all my friends for their encouragement and support.

Dedicated to my Mother, and the memory of my Father.

Abstract

The objective of the present study was to investigate the potential use of bulk segregant analysis (Michelmore *et al.*, 1991) as a method to rapidly identify genetic markers linked to traits in sugarcane. Four bulked DNA samples were prepared from progeny of a sugarcane cross, AA157, based on segregation for the fibre trait. The bulks comprised five and ten individuals on either side of the fibre phenotypic extreme. The random amplified polymorphic DNA (RAPD) technique (Williams *et al.*, 1990) was used to screen for differences between the low and high fibre bulks. A total of 749 fragments were amplified in the bulks, eight of which were polymorphic. The segregation of the bulk specific polymorphism was analysed in 80 progeny of the AA157 cross; and seven were found to reproducibly segregate on a 1:1 basis. This indicates that they are single dose fragments. A total of 79 polymorphisms were detected between the parents of the cross, indicating 10.5% variation in the genomic region sampled. Twenty two of the parental polymorphisms segregated as single dose fragments in the progeny of the cross AA157.

Analyses of variance (ANOVAs), and multiple regression analyses, were used to ascertain linkage of the putative RAPD markers to fibre, and if linked, to determine the fibre variation ascribed respectively. Three RAPD fragments were found linked to the fibre trait. Fragments $OPA17_{438}$ and $OPC16_{889}$ (at the 99% significance level), and $OPB11_{464}$ (at the 95% significance level). These putative markers ascribed a total of 28.6% fibre variation in the 1993 season. The association of the RAPD markers with fibre in the different seasons (1992, 1993 and 1994) was investigated. Three RAPD markers were found linked to the fibre trait in each season, with a total of 5.5% and 31,4% fibre variation ascribed in the 1992 and 1994

seasons respectively. Marker $OPA17_{438}$ was found to be linked to the fibre trait in all three seasons investigated, and marker $OPC16_{889}$, was found linked to the fibre trait in the 1992 and 1993 seasons. Cross validation of the linkages of the RAPD markers to the fibre trait was carried out by a modified form of 'jacknifing' where the sample size was reduced to N-10, and RAPD marker-fibre trait associations investigated as before. RAPD markers $OPA17_{438}$ and $OPC16_{889}$ were still consistent across the seasons, however marker $OPA17_{438}$ was no longer linked to the fibre trait in the 1992 season.

To investigate the genetic behaviour of RAPD based markers in sugarcane and the potential for their application in marker-assisted selection (MAS), two putative RAPD markers were converted to sequence characterised amplified regions (SCARs) (Paran and Michelmore, 1993). The RAPD fragments $OPA17_{438}$, $OPB11_{464}$, and $OPC16_{889}$ were excised from agarose gels, re-amplified and cloned into the pCR-Script SK (+) phagemid for sequencing. RAPD markers $OPA17_{438}$ and $OPB11_{464}$ were converted to SCARs by using their sequences to design longer specific primers. A third SCAR marker, $SA11_{640}$, originally derived from sugarcane cDNA as a potential stem preferential expressed sequence tag, was included in the analysis to increase the sample size.

All three SCAR markers segregated in a monomorphic fashion in the parents and progeny of the cross AA157. In addition, monomorphic length variants for markers, $OPA17_{438}$ and $OPB11_{464}$ were detected with the SCAR amplification. All three SCARs segregated in a monomorphic fashion in different commercial varieties and bulks of *S. officinarum* and *S. spontaneum*, the progenitors of modern commercial varieties. The segregation analyses of the

SCAR markers indicate that the RAPD polymorphism of marker $SA11_{640}$ was probably due to a point mutation or mismatch in the priming site. The segregation analyses of SCARs for the markers $OPA17_{438}$ and $OPB11_{464}$ indicate that their segregation in the RAPD analyses was due to an insertion mutation in the genetic locus. The combined results of the SCAR and RAPD segregation of markers $OPA17_{438}$ and $OPB11_{464}$ are indicative of preferential pairing in the cross AA157.

Finally, to investigate the extent of linkage disequilibrium in a modern commercial variety, twenty two single dose RAPD fragments were investigated for their association with four traits in 53 progeny of cross AA157. The four traits investigated were fibre %cane, brix %cane, pol %cane and ers %cane over three seasons (1992, 1993 and 1994), at different ages of harvest (12, 8, and 9 months respectively). Seventeen linkages of RAPD markers to the four traits, over the three seasons, were detected. The phenotypic variation ascribed by the RAPD markers ranged from 7.6% fibre %cane variation explained by one marker in 1992, 29.6% fibre %cane (three markers) in the 1993 season to 10% (three markers) in 1994. A total of 14.1% brix %cane variation was ascribed by two markers in 1992, 9.6% (one marker) in 1993 and 16.3% (two markers) in the 1994 season. A total of 13.5% estimated recoverable sucrose %cane was ascribed by one marker in 1992, 12% (two markers) in 1993 and 15.3% (two markers) in the 1994 season. Two markers explained 17.2% pol %cane variation in 1992 and 25.4% in the 1994 season. Only four markers were detected across different environments, three of which were linked to fibre. These were $OPA17_{438}$, $OPB16_{618}$ and $OPC16_{889}$, each linked to fibre in two seasons. RAPD marker $OPB11_{464}$ was linked to estimated recoverable sucrose %cane in two seasons. Two markers were found associated with different traits in a single season.

RAPD marker $OPB11_{464}$ was found associated with brix %cane and estimated recoverable sucrose %cane in the 1993 season, and RAPD marker $OPA17_{438}$ was found associated with all four traits in the 1994 season.

CONTENTS

List of Tables

List of Figures

List of Abbreviations

AFLP	amplified fragment length polymorphism
Anon	anonymous
ANOVA	analysis of variance
AP-PCR	arbitrarily primed PCR
ASA	allele-specific amplification
BC	backcross
BNL	Brookhaven National Laboratory
bp	base pair
BSA	bulk segregant analysis
CAPs	cleaved amplified polymorphims
cDNA	complementary DNA
cM	centiMorgans
cpDNA	chloroplast DNA
CTAB	cetyltrimethylammonium bromide
cv%	coefficient of variation
DAF	DNA amplification fingerprinting
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DH	double haploid

DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphates
dTTP	deoxythymidine triphosphate
dUTP	deoxyuridine triphosphate
EBN	endosperm balance number
EDTA	ethylenediaminetetracetic acid
ERS	estimated recoverable sucrose
F prob	the variance ratio
$F_{1, 2,...x}$	first filial generation, second filial generation, (with subsequent numbers indicating generational stage after initial cross)
g	gravity
G.Mean	grand mean
GISH	genomic in situ hybridisation
H	haploid
ISSR	inter-simple sequence repeats
kb	kilobase
KCl	potassium chloride
Lg	linkage group
LOD	logarithm of odds
m/m	mass by mass
m/v	mass by volume
MAS	marker assisted selection
Max Val	maximum value

Mb	megabase
MG	megagametophytes
$MgCl_2$	magnesium chloride
Min Val	minimum value
mitDNA	mitochondrial DNA
n	haploid number of chromosomes
NaCl	sodium chloride
NILs	near isogenic lines
nm	nanometres
NPA,B, C	N (author's first initial), P plasmid designation, A, B, and C indicate Operon primer lot series from which it was derived.
°C	degrees Celsius
OD	optical density
OPA (B, C)	Operon primer with A, B, and C indicating lot series
PCR	polymerase chain reaction
PEG	polyethylene glycol
pers comm	personal communication
PVE	phenotypic variation explained
QTLs	quantitative trait loci
R^2	regression coefficient
RAMP	random amplified microsatellite polymorphism
RAPD	random amplification of polymorphic DNA
rDNA	ribosomal DNA

RFLP	restriction fragment length polymorphisms
RI	recombinant inbred
RILs	recombinant inbred lines
S	*Saccharum*
SD	standard deviation
SASEX	South African Sugar Association Experiment Station
SCAR	sequence characterised amplified regions
SPA/SPB	SCAR primer (with A and B indicating Operon primer lot series from which it was derived)
SSC	sodium chloride/sodium citrate (buffer)
STS	sequence tagged sites
T	threshold
TBE	Tris/borate (buffer)
TE	Tris/ethylenediaminetetracetic acid
Tris-HCl	Tris hydrochloric
UMC	University of Missouri Columbia
UV	ultra violet
V/cm	voltage per centimetre
v/v	volume by volume
x	base number of chromosomes

Chapter 1

Introduction

The genus *Saccharum* is a polyploid grass belonging to the family Gramineae in the Saccharine subtribe of the Andropogoneae (Jeswiet, 1925). Members of the genera *Saccharum*, *Erianthus* (= section Ripidium), *Miscanthus* (= section Diandra), *Sclerostachya* and *Narenga* are thought to have arisen as a result of polyploidization and hybridisation events (Daniels and Roach, 1987). These genera constitute a closely related interbreeding group thought to have been involved in the origin of sugarcane and are thus referred to as the *Saccharum* complex (Mukherjee, 1954). Modern sugarcane cultivars are hybrids belonging to the genus *Saccharum*, a complex genus comprising six species, namely *Saccharum officinarum*, *Saccharum spontaneum*, *Saccharum sinense*, *Saccharum barberi*, *Saccharum robustum* and *Saccharum edule* (Daniels and Roach, 1987).

However this classification has been disputed as several workers have concluded that *S. barberi* may actually be a natural hybrid of *S. spontaneum* and *S. officinarum* (Daniels and Roach, 1987). Sugarcane (*S. officinarum* L) is a domesticated crop that has been cultivated for millenia because of its high sucrose content (Stevenson, 1965). *S. robustum* was identified by Brandes in a 1928 expedition to Papua New Guinea, and was proposed to be the wild progenitor of cultivated sugarcane (Brandes, 1929).

Until the end of the last century the principal cultivated canes were *S. officinarum* (noble canes), *S. barberi* and *S. sinense* (Stevenson, 1965). Breeding was primarily restricted to their collection and examination from native gardens. Propagation was only done with stem cuttings. It was only late in the last century (around 1860) that sexual fertility of sugarcane was discovered in Java and Barbados (Stevenson, 1965). This discovery, together with pressures

brought about by the advent of the Sereh disease, prompted the breeders in Java to start interspecific hybridisation programs to facilitate the introgression of disease resistance from the wild *S. spontaneum* (Stevenson, 1965). The initial interspecific hyridisations between *S. officinarum* and *S. spontaneum* were followed by repeated backcrossing with *S. officinarum* to minimise the negative effect of the wild parent, a process referred to as nobilisation (Bremer, 1961). The resultant interspecific hybrids, POJ varieties, were to become the progenitors of the current major commercial varieties (Moore and Wu, 1991). Sugarcane breeding at present involves the intercrossing of certain nobilised hybrids derived from a very small number of *S. spontaneum* parents.

Sugarcane has lagged behind most major crop species in terms of development of genetic maps because of its high ploidy level and complex genetics. Some of these complexities are the aneuploid nature of sugarcane (Sreenivasan *et al*, 1987), and the possible multiple origins of its genome (Price, 1963). Furthermore, no suitable diploid relatives are available to develop genetic maps that could be extrapolated for use in sugarcane genetics. These factors meant that the usual linkage theory of Mendelian inheritance had to be modified to cater for the peculiar genetics of sugarcane.

In 1992 Wu *et al.* described a framework that could be used to genetically map sugarcane, and other complex polyploids. The theory of single dose restriction fragment analysis in polyploids enabled the mapping of polymorphic markers that segregated in a 1:1 ratio. Since then several studies have been carried out to investigate the genetics of sugarcane through map construction (Burnquist, 1991, Da Silva, 1993, Da Silva *et al.*, 1993, Al-Janabi *et al.*, 1993, Da Silva *et al.*, 1995, Grivet *et al.*, 1996). In addition, the abundance of DNA markers has facilitated an increase in the number of evolutionary and phylogenetic analyses (D'Hont *et al.*, 1993, Harvey *et al.*, 1994, Al-Janabi *et al.*, 1994a, Sobral *et al.*, 1994, Lu *et al.*, 1994a, b, Huckett and Botha,

1995, D'Hont *et al.*, 1996, Harvey and Botha, 1996, Besse *et al.*, 1997, Oropeza and de Garcia, 1997), molecular marker-phenotypic trait analyses (Da Silva, 1993, Msomi and Botha, 1994, Sills *et al.*, 1995, Daugrois *et al.*, 1996), comparative genetic analyses (Grivet *et al.*, 1994, Dufour *et al.*, 1997) and evaluation of true crosses (D'Hont *et al.*, 1995, Harvey *et al.*, 1997).

The final product of genetic analysis is the application of results in the improvement of plant crops through plant breeding, either in conventional or marker-assisted selection (MAS). The development of genetic maps is one of the prerequisites for the eventual tagging of genes, or genetic loci controlling traits of agronomic interest, with molecular markers (Michelmore *et al.*, 1992). Such markers can then be used in map-based gene cloning (Arondel *et al.*, 1992, Leyser *et al.*, 1993), chromosome walking (Bender *et al.*, 1983), resolution of quantitative trait loci (QTLs) (Lander and Botstein, 1989) or directly in MAS if tightly linked to the genetic locus of interest (Gimelfarb and Lande, 1994). The development of genetic maps is an elaborate process, and populations used for initial mapping are usually chosen for their marker segregation rather than linkage disequilibrium (Paterson and Wing, 1993).

Michelmore *et al.* (1991) described a method, termed bulk segregant analysis (BSA), to rapidly identify markers linked to phenotypic traits that did not require the prior construction of genetic maps. The main condition for BSA is a well-defined population segregating for the phenotypic trait of interest, with the proviso that the parents of such a population are genetically different at the locus of interest. The method involves bulking the DNA of individuals at the phenotypic extremes of such a population, and screening with suitable marker technology. The differences between the opposing bulks, detected as polymorphic markers, are likely to represent linkage of those markers to the genetic region controlling the phenotypic trait of interest.

In the present study the potential use of the BSA as a method to rapidly identify genetic markers linked to traits in a complex polyploid like sugarcane was investigated (Chapter 3). The BSA was carried out in a population of 150 F_1 individuals from a near commercial sugarcane cross, AA157, segregating for the fibre trait. To maximise the throughput of DNA markers the random amplification of polymorphic DNA (RAPD) technique was employed (Williams *et al.*, 1990). The RAPD technique is based on the amplification of anonymous regions of the genome, and does not require prior knowledge of the DNA sequence.

The RAPD technique has been associated with several limitations, which may prevent the utilisation of RAPD markers in further genetic analyses, and subsequent application in MAS. Some of these problems are reproducibility among different laboratories (Penner *et al.*, 1993a, Jones *et al.*, 1997). RAPD markers are also dominant by nature i.e. unable to distinguish zygosity and resolution of dosage (Tingey and del Tufo, 1993). In addition they are susceptible to point mutations and mismatches resulting in cross-specificity (Miklas *et al.*, 1993), which may restrict their applicability in MAS. Designing longer specific primers based on the sequences of the RAPD polymorphic fragments can circumvent some of these problems. This requires cloning and sequencing of the RAPD fragments, essentially converting them to sequence characterised amplified regions (SCARs) (Paran and Michelmore, 1993, Ohmori *et al.*, 1996).

Two RAPD markers found to be putatively linked to the fibre trait by BSA were converted to SCARs, and together with a third SCAR marker derived as a potential stem preferential expressed sequence tag (S. Groenewald[1], personal communication), their behaviour in sugarcane was investigated (Chapter 4). The potential of SCARs to resolve dosage or zygosity in a high polyploid like sugarcane, where the basic chromosome number of modern varieties

1 S. Groenewald, South African Sugar Association Experiment Station, Private Bag X02, Mt Edgecombe, 4300.

is estimated to range between 8x and 12x (Sreenivasan *et al.*, 1987, D'Hont *et al.*, 1995, D'Hont *et al.*, 1996), was also investigated. A non-cross specific RAPD marker, also putatively linked to fibre, was included in the analysis for comparative purposes.

The observation of limited recombination between *S. officinarum* and *S. spontaneum* chromosomes in a modern sugarcane cultivar, R570 (D'Hont *et al.*, 1994, D'Hont *et al.*, 1996, Grivet *et al.*, 1996) may imply that there is strong linkage disequilibrium with respect to traits from the two progenitors. The high levels of inbreeding, due to the backcrossing involving *S. officinarum*, and the remaining *S. spontaneum* chromosomes may provide alternate alleles for some of the contrasting traits presumably derived from either base species.

The availability of extensive phenotypic data for several traits in the cross, AA157, coupled with single dose RAPD markers generated in the BSA study presented an opportunity to investigate the extent of linkage disequilibrium with respect to some of these traits in a modern sugarcane variety (Chapter 5). The analysis of linkage of the RAPD markers to traits presumably derived from either *S. officinarum* or *S. spontaneum* enabled a preliminary analysis of the genetic nature of traits that are amenable to QTL mapping in sugarcane.

At present the South African sugarcane breeding programme takes an average of 14 years before a new commercial variety can be released. This is due to the over reliance on morphological markers, which are strongly influenced by the environment thereby masking the true genetic potential of the plants (Moore and Irvine, 1991). The unravelling of the complex genetics of sugarcane through molecular markers will greatly speed up the process of genetic analyses. Hopefully, these will eventually lead to marker-assisted selection in the plant breeding programme.

Chapter 2

Molecular markers, mapping strategies and the genetic improvement of crops

2.1 General Literature Review

2.1.1 Historical perspectives of agricultural improvement

Mankind has been involved in the cultivation of crop species, such as grasses, for thousands of years. The initial domestication of many wild species of food crops was followed by years of deliberate manipulation to improve their agronomic properties. This improvement involved selecting and breeding individuals that were higher yielding, more resistant to pests and other environmental hazards, or better looking and longer lasting depending on the initial reasons for domestication. Such improvement programs involved the manipulation of the genetic resources of the plants. Accumulation of knowledge over many centuries laid the basis for a formal approach to genetic manipulation of major food crop species namely plant breeding.

2.1.2 Genetic analysis of crop species

Genetic analysis is a necessary prerequisite for the continued improvement of crop species. This may include characterisation of germplasm to establish the genetic relationships between clones whether they are wild or domesticated, which is crucial in breeding programs for evaluation of potential clones for use as parents (Lee, 1995). It also enables the inference of phylogenetic and evolutionary relationships among species (Menkir *et al.*, 1997).

Initially, genetic analysis of plants was based solely on the characterisation of the genetic basis of naturally occurring variation in phenotype. For this purpose, morphological markers either

pleiotropically linked to the desired trait, or characteristic of the trait itself, were exploited. Recently both biochemical and DNA-based molecular markers have found increasing usage in genetic analysis.

One of the main objectives of genetic analysis has been the development of genetic maps for major crop species. Genetic maps in many plant crops have been made possible by the recent advances in the generation of molecular markers. Their ease of generation, and independence from ontogenic processes has meant that appropriate populations can be developed whose marker segregation can be assayed for linkage analysis (Moore and Irvine, 1991).

2.1.3 Morphological markers

Traditionally morphological markers have been used extensively for taxonomic classification of crop species. For example in sorghum (*Sorghum bicolor*), Harlan and de Wet (1972) used morphological markers such as grain shape, glumes, and panicle shape to partition genetic variability in cultivated sorghums into five races and 10 hybrids.

By definition most morphological markers developed thus far are for simply inherited traits i.e. they are based on macro-mutations of alleles with highly qualitative traits, while most agronomically important traits are thought to be under the control of quantitative trait loci (Mather and Jinks, 1974). Furthermore, these markers usually have secondary pleiotropic effects on economic characters such as yield and therefore, have very limited usage in crop improvement (Beckmann and Soller, 1983).

Their utility is also hampered by their scarcity, for example in various cultivars of *Brassica rapa* L. (syn. campestris), seventy morphological markers of simple inheritance have been identified

(Chyi *et al.*, 1992). Only one linkage group with more than two loci has been established (Stringham, 1977). Another example of morphological markers used in linkage analysis was in papaya (*Carica papaya* L.), where Hofmeyr (1939) identified a loose linkage between sex inheritance and flower and petiole colour. The map of papaya described by Hofmeyr (1939) comprised three markers and only covered 41 cM, further illustrating the scarcity of morphological markers.

2.1.4 Biochemical markers

Isozymes, the most common type of biochemical marker, are differently charged protein molecules that can be separated using electrophoretic procedures, usually starch gels (Markert and Moller, 1959) or more recently polyacrylamide gels (Feldmann, 1984).

Isozymes have also been used for studies of genetic variability e.g. sorghum (Morden *et al.*, 1989, Morden *et al.*, 1990, Aldrich *et al.*, 1992). In sugarcane, earlier studies were restricted to the use of leaf flavonoids as chemotaxonomic markers (Waldron and Glasziou, 1972, Daniels *et al.*, 1980). Within the genus *Saccharum* seven esterase bands were found to be common to all *Saccharum* species (Waldron and Glasziou, 1972). A flavonoid marker designated as F4 was present in all *S. spontaneum* clones, but absent in all *S. officinarum* and was thus used as a means of distinguishing the two *Saccharum* species (Daniels *et al.*, 1980). Another flavonoid marker, F13, was found to be specific to *Erianthus* (section=Ripidium) (Daniels *et al.*, 1980). In a more recent study to ascertain the potential utility of isozyme markers in sugarcane genetics and breeding, Glaszmann *et al.* (1989) found that they could use isozymes, together with multivariate analysis, to isolate an *Erianthus* clone and also separate *S. spontaneum* clones from *S. robustum* and *S. officinarum* clones.

The major restriction to widespread application of isozyme markers to genetic analysis of plants is their paucity, as many plant cultivars have been found to be remarkably uniform with respect to isozyme markers (Rick and Forbes, 1975, Nevo *et al.*, 1979). Such results have been confirmed in sorghum (Aldrich *et al.*, 1992). In addition, most isozyme markers have the disadvantage of being developmentally regulated i.e. they are phenotypically expressed only at specific developmental stages and/or in certain specific tissues or organs (Beckmann and Soller, 1983). The above-mentioned studies on leaf flavonoids in *Saccharum* species attest to their tissue specificity; while the *Idh*-2 isozyme in the young leaves of apple (*Malus*) linked to fruit colour is an example of developmental stage limitation (Weeden *et al.*, 1994).

Another limitation of isozyme analysis can be observed in the study of evolutionary processes implicated in the formation of polyploids, where they may confound the estimation of multiple polyploid events (Soltis and Soltis, 1993). The differences in the fixed patterns of isozymes among different polyploid populations may be due to initial segregation and subsequent fixation by breeding, drift, or selection (or combination thereof) in the progeny of highly heterozygous progenitors, not different origins (Soltis and Soltis, 1993).

2.1.5 DNA-based markers

DNA markers offer many advantages over morphological markers or biochemical markers in that they segregate in many plants, and scoring one does not interfere with scoring another (Moore and Irvine, 1991). This means that large numbers of DNA markers can be analysed in a population derived from a single cross, making it possible to construct a unified linkage map composed of many genetic markers. DNA markers exploit the differences in sequence between interfertile plants to define the genotype of each individual (Young, 1993).

There are two ways that differences in the size of DNA fragments are characterised for genetic analysis. The older and the most widely utilised method has been that of restriction fragment length polymorphisms (RFLPs) analysis (Botstein *et al.*, 1980). The second and more recent method is based on the polymerase chain reaction (PCR) technique (Saiki *et al.*, 1985). Some of the PCR-based methods involve microsatellites, random amplified polymorphic DNA (RAPD), arbitrarily primed PCR (AP-PCR), and DNA amplification fingerprinting (DAF) analyses (Tautz, 1989, Williams *et al.*, 1990, Welsh and McClelland, 1990, Caetano-Anolles *et al.*, 1991). Recent advances have seen methods that use a combination of both techniques such as cleaved amplified polymorphic sequences (CAPs) (Jarvis *et al.*, 1994) and amplified fragment length polymorphisms (AFLPs) (Vos *et al.*, 1995).

2.1.5.1 Restriction Fragment Length Polymorphisms (RFLPs)

RFLPs, the most widely used DNA markers, are based on restriction endonucleases which recognise and cut DNA at or adjacent to short, specific sequences of nucleotides (Zabeau and Roberts, 1979). They have genetic characteristics such as the ability to detect multiple allelic forms and co-dominance, which means they can distinguish between a heterozygote and a homozygote (Beckmann and Soller, 1983).

RFLP polymorphisms are detected as a result of changes in the locations of restriction sites between individuals, either due to mutations in DNA creating or abolishing a restriction site, insertions, deletions or inversions between two restriction sites; thus providing alternate alleles that can be used in segregation analysis and genetic mapping (Botstein *et al.*, 1980). Restriction enzyme digestion produces DNA fragments between a few hundred and a few thousand base pairs

in length. This distinctive set of DNA fragments defines an organisms RFLP genotype (Young, 1993).

A restriction fragment pattern is analysed by using cloned pieces of DNA, usually derived from the plant's DNA. The cloned DNA probes are used to characterise corresponding restriction fragments from the plants' DNA by specifically binding to those fragments complimentary to the cloned sequence (Beckmann and Soller, 1983). These cloned unique DNA sequences are labelled, with radioactive ^{32}P or ^{33}P or non-radioactive digoxigenin, so that their hybridisation to complementary DNA fragments can be visualised through autoradiography (Southern, 1975).

2.1.5.2 Random Amplified Polymorphic DNA (RAPDs)

The second method for finding DNA markers utilises PCR, which was first described by Saiki *et al.* (1985). It employs a thermocycling block, the plant DNA being analysed, and a cocktail of DNA primer sequences, enzyme (usually Taq polymerase), and nucleotides. By repeatedly alternating between a high temperature to separate the two strands of the double stranded DNA, and lower temperatures for the primers to anneal to the separate strands, allowing synthesis of new second strands, a rapid exponential increase occurs in the number of copies of that particular stretch of DNA. The high product output enables direct visualisation by electrophoresis through agarose gels and staining with ethidium bromide.

Initially, PCR was described for DNA segments whose sequences had been characterised, and primers could thus be designed. One of the earliest uses of the technique was for medical diagnostics such as testing for anaemia (Saiki *et al.*, 1985). Polymorphisms were scored according to the absence or presence of the expected PCR product due to the variation in the priming site

caused by mutations such as base pair substitutions, insertions or deletions. Alternatively, different sized fragments may be diagnostic of an insertion or deletion mutation in the target locus. The requirement for prior knowledge of the DNA sequence to be amplified limited widespread application of the technique.

Welsh and McClelland (1990) and Williams *et al.* (1990) independently described a PCR-based technique that did not require prior knowledge of the DNA/target sequence to amplify distinct segments of the DNA. This method made use of arbitrary primers to amplify a specific set of randomly distributed loci in any genome. These oligonucleotide primers are synthetically produced random DNA sequences usually ten nucleotides in length. This method, RAPD-PCR or AP-PCR has contributed to an increased output of genetic markers, the majority of which are inherited in a Mendelian fashion. Some of the examples of its application include population studies (Hu and Quiros, 1991, Welsh and McClelland, 1992), biosystematics (Stiles *et al.*, 1993), gene tagging (Giovannoni *et al.*, 1991, Martin *et al.*, 1991, Michelmore *et al.*, 1991) and genetic mapping of crop plants e.g. sugarcane (Al-Janabi *et al.*, 1993) and papaya (Sondur *et al.*, 1996). Recently, RAPD markers, together with isozymes, have been used to genetically map a wild plant, *Mimulus* (Lin and Ritland, 1996a).

The major advantage of this technique is its simplicity and the fact that very small amounts of template DNA are necessary for successful amplification. In sugarcane, nanogram amounts of template DNA are used (Al-Janabi *et al.*, 1993, Harvey *et al.*, 1994). As RAPDs involve no Southern hybridisation, this eliminates the need for radioactive labelling. This also means the turnover time for the generation of genetic markers is greatly reduced. In addition, the throughput as measured by the number of amplification products per reaction is generally high.

In sugarcane an average of 11 fragments are amplified per reaction (Harvey *et al.*, 1994, Huckett and Botha, 1995), 8.5 bands per primer in lettuce (*Lactuca sativa*) (Kesseli *et al.*, 1992) and 6 fragments in wheat (*Triticum aestivum* L.) (Qi *et al.*, 1996). An added advantage is that polymorphisms can be detected in fragments containing highly repetitive sequences (Williams *et al.*, 1990). This provides markers in regions previously inaccessible to RFLP analysis, which is useful in germplasm where genetic variability is limited (Foolad *et al.*, 1993). It also means that RAPD markers can be used to target specific regions or intervals in RFLP-based genetic maps for further saturation e.g. lettuce (Giovannoni *et al.*, 1991), or chromosome walking (Bender *et al.*, 1983).

2.1.5.3 Comparison of RFLP and RAPD markers

The key difference between RAPD and RFLP markers lies in the method by which changes in the DNA sequences are detected. With RAPDs, variations in sequence are observed due to differences in the ability to bind short oligonucleotide primers for the polymerase chain reaction (PCR), whereas in RFLPs the variation is detected due to differences in the restriction site. This difference has important implications for each type of marker's utility in genetic analysis.

The level of information provided by any pair of markers to evaluate the recombination fraction depends on the type of marker i.e. dominance versus co-dominance, and their phase i.e. repulsion versus coupling (Allard, 1956). This means that in the construction of genetic maps the level of information provided by RAPD markers (dominant) is lower than that of RFLPs (co-dominant) if F_2 populations are used. One way of increasing the information content of RAPD markers is to use the type of populations whose loci are homozygous like doubled haploids.

In species where the level of polymorphisms is low, either due to inbreeding or recombination suppression, RAPD markers may be the most suitable marker system for map construction because of their sensitivity to point mutations and mismatches (Foolad *et al.*, 1993).

2.1.5.4 Microsatellites

Microsatellites, which are also known as short tandem repeats or simple sequence length polymorphisms, consist of tandemly arrayed di-, tri-, and tetra-nucleotide repeats, which are hypervariable and ubiquitously distributed throughout eukaryotic genomes (Lagercrantz *et al.*, 1993). PCR-based microsatellite markers have been developed by using the unique sequences that flank the microsatellite repeats to design primers for PCR amplification (Tautz, 1989, Weber and May, 1989).

DNA markers derived from the amplification of microsatellite loci have been shown to be more variable than RFLPs such as in soybean (*Glycine max.*) (Akkaya *et al.*, 1992), *Brassica* (Lagercrantz *et al.*, 1993), maize (*Zea mays*) (Senior and Heun, 1993), rice (*Oryza sativa*) (Wu and Tanksley, 1993, Akagi *et al.*, 1996), barley (*Hordeum vulgare*) (Becker and Heun, 1995) and wheat (Röder *et al.*, 1995). They can also be used as co-dominant markers linked to important traits, as has been shown in rice (Akagi *et al.*, 1996). High-density molecular maps based on microsatellites have also been constructed for the mouse (Dib *et al.*, 1996) and human genomes (Dietrich *et al.*, 1996). They have also been used in the detection of genetic diversity between closely related cultivars of bread wheat (Plaschke *et al.*, 1995).

Other types of PCR-based markers that exploit the ubiquity of repeated sequences in plant genomes include random amplified microsatellites polymorphisms (RAMPs) and inter-simple sequence

repeats (Wu *et al.*, 1994, Zietkiewcz *et al.*, 1994). Both these techniques involve PCR profiling of anonymous DNA sequences in the genome. The RAMP technique makes use of a RAPD primer on one end, and a radioactively labelled, anchored microsatellite primer on the other end. The sequence of the anchored microsatellite primer is based on an arbitrary set of bases with a repeat sequence motif. The advantage of RAMP markers is that they are co-dominant, while their disadvantage is that they require the use of radioactivity.

In the inter-simple sequence repeat (ISSR) technique, a single anchored primer specific to a particular simple sequence repeat is used to amplify the DNA of the same type of repeat sequence. Polymorphisms are revealed as a result of deletion mutations in one of the priming sites, or deletion/insertion within the locus. Like RAPDs, the ISSR markers are dominant. One of the disadvantages of the ISSR technique is the number of faint bands in the PCR profiles that are caused by replication slippage (Perry *et al.*, 1994), which may lead to scoring errors.

2.1.5.5 Amplified fragment length polymorphisms (AFLPs)

The AFLP technology (Vos *et al.*, 1995) generates one of the newest types of DNA marker that is likely to find widespread application. AFLPs are based on selective amplification of restriction fragments from a total digest of genomic DNA. They are typically inherited in a Mendelian fashion and can therefore be used for genetic fingertyping, identification of molecular markers, and mapping of genetic loci.

Briefly, three steps are followed to obtain AFLP markers: (i) digestion of genomic DNA with two restriction enzymes, such as *Mse*I and *EcoRI* (frequent-cutter enzyme); (ii) ligation of adapter oligonucleotides to the restricted ends and (iii) selection of fragments by two successive PCR-based

amplification steps, using primers complimentary to the adapter nucleotides with one to three additional selective nucleotides on the 3' ends of the PCR primers (Anon., 1996a, Cevera *et al.*, 1996). The inclusion of selective nucleotides, extending into the restriction fragments, ensures that only the restriction fragments in which the nucleotides flanking the restriction site match the selective nucleotides will be amplified. The subset of amplified fragments are then analysed by denaturing polyacrylamide gel electrophoresis to generate a fingerprint.

The fact that the AFLP technology involves restriction analysis means that the DNA preparations used are required to be much cleaner than in RAPDs, which promotes reproducibility (Jones *et al.*, 1997). Another advantage is that no prior knowledge of sequence information is necessary for successful PCR amplification, however, annealing of primers to the complimentary adapter oligonucleotides is highly specific thus ensuring reproducibility. The drawback is that restriction analysis requires large amounts of DNA in comparison with RAPDs, and can be time consuming.

The output of the AFLP technology, in terms of genetic loci that can be analysed per experiment, is substantially higher than in RAPDs, approximately 10-fold the number of informative markers in the genus *Populus* (Cevera *et al.*, 1996). In addition, AFLP markers are co-dominant, which means they can differentiate between the homozygous and heterozygous states thus providing more information per marker than RAPDs.

2.1.5.6 Sequence specific PCR markers

PCR-based markers that sample anonymous portions of the genome, like RAPDs, are prone to problems due to variations in reaction conditions (Hallden *et al.*, 1996, Staub *et al.*, 1996a, Talbert *et al.*, 1996). This may prevent reproducibility of RAPD-PCR results across different laboratories

(Penner *et al.*, 1993a). In addition, they are dominant by nature i.e. an inability to distinguish zygosity (Tingey and del Tufo, 1993), due to the fact that in the multi-locus RAPD-PCR profile related alleles cannot be readily distinguished. Using sequence specific primers, designed for high stringency reaction conditions, which tag specific genomic regions and can therefore distinguish zygosity, can alleviate some of these problems. Sequence characterised amplified regions (SCAR) are one type of markers that are designed from RAPD loci, and have been utilised for distinguishing zygosity in gene tagging studies (Paran and Michelmore, 1993, Adam-Blondon *et al.*, 1994, Ohmori *et al.*, 1996). Primers designed from the unique regions flanking microsatellites, also produce co-dominant markers. Another type of sequence specific markers is the allele-specific amplification (ASA) technique, which makes use of primers whose sequences are designed to partially overlap the allelic sequence differences and a dual-PCR system (Wu *et al.*, 1989, Williams *et al.*, 1996). ASA can detect single base differences between two alleles (Williams *et al.*, 1996).

2.2 Utilisation of molecular markers

2.2.1 Assessment of genetic diversity in crop plants

Analyses of the extent and distribution of genetic diversity in crop plants is essential for understanding the evolutionary relationships between accessions, which is a useful tool in optimising sampling of genetic resources and breeding strategies (Menkir *et al.*, 1997). In the past plant breeders have selected among closely related strains on the basis of morphological markers such as size, and colour, that are readily observable and inherited with the desired trait.

In genetic diversity studies the cost per data unit, and the informativeness of the marker system is different from that measured in genetic mapping. The main criteria of informativeness is heterozygosity, number of markers or probes used, and genome coverage (Nei, 1973, Messmer *et*

al., 1991, Leberg, 1992). A major disadvantage of RFLPs is that they are usually single locus specific and detect polymorphisms directly related to the DNA sequence used as a probe (Lu *et al.*, 1996). On the other hand, PCR-based markers such as RAPDs, AFLPs and microsatellites, have a high throughput. A major technical advantage of PCR-based markers is that they can be applied to a large number of samples, and they are open to automation. Furthermore, the RAPDs have low developmental costs. It is for this reason that most of the recent studies in genetic diversity have made use of PCR-based markers (Table 2.1).

Table 2.1 A summary of some published genetic diversity studies in crop plants, and species.

Crop Plant/Species	**Marker system**	**Reference**
Maize (*Zea mays*)	RFLP,Isozyme, Proteins	Burstin *et al.*, 1994
Rye (*Secale cereale*)	RAPD	Iqbal and Rayburn, 1994
Bread wheat (*Triticum aestivum* L.)	Microsatellites	Plaschke *et al.*, 1995
Conifers	RFLP (cpDNA)	Tsumura *et al.*, 1995
Triticeae	RAPD	Wei and Wang, 1995
Lilium	RAPD	Yamagishi, 1995
Sunflower (*Helianthus anuus*)	RFLP	Zhang *et al.*, 1995
Tomato(*Lycopersicon esculentum*)*	RAPD	Bogani *et al.*, 1996
Hops (*Humulus lupulus*)	RAPD, microsatellites	Brady *et al.*, 1996
Rapeseed (*Brassica napus* L.)	Microsatellites	Charters *et al.*, 1996
Sorghum (*Sorghum bicolor*)	RFLP, RAPD, microsatellites	De Oliveira *et al.*, 1996
Eastern white pine (*Pinus strobus* L.)	Microsatellites	Echt *et al.*, 1996
Barley (*Hordeum vulgare*)	RFLP, RAPD	Linc *et al.*, 1996
Pea (*Pisum sativum* L.)	RFLP,RAPD,AFLP,microsatellites	Lu *et al.*, 1996
Soybean (*Glycine max*)	AFLP	Maughan *et al.*, 1996
Beta vulgaris L.	Microsatellites	Morchen *et al.*, 1996
Barley (*Hordeum vulgare*)	RAPD, microsatellites	Sanchez de la Hoz *et al.*, 1996
Brassica species	Microsatellites	Szewc-McFadden *et al.*, 1996
Aspen (*Populus tremuloides* Michx.)	RAPD	Tuskan *et al.*, 1996
Rice (*Oryza sativa*)	STS (glutelin genes)	Wu *et al.*, 1996
Chinese sorghums (*Hordeum vulgare*)	RFLP, RAPD, microsatellites	Yang *et al.*, 1996
Plum cultivars (*Prunus domestica*)	RAPD	Ortiz *et al.*, 1997

* indicates protoplast-derived regenerants

One of the disadvantages of PCR-based marker system is that similar mobilities of unrelated bands may be misinterpreted as representing the same locus (Thormann *et al.*, 1992, Quiros *et al.*, 1995). This is likely to be more pronounced in RAPDs, while the pre-digestion step and polyacrylamide gel resolution in AFLPs will reduce this type of error. In addition, the tolerance of mismatches in RAPDs close to the 5'end of the primer but not in the 3'end may lead to a situation where two individuals sharing the band are less related to each other than some individuals not having the band because of a mutation due to a single mismatch at the 3' end (Jones *et al.*, 1997). The Nei and Li (1979) algorithm for estimating genetic distance,

$$1 - F = 1 - [\, 2n_{xy} / (n_x + n_y)]$$

where F is the ratio of shared bands between x and y, $2n_{xy}$ is the number of shared bands, and n_x and n_y are the number of bands observed in individuals x and y respectively, is the most commonly used algorithm to assess genetic variability. The two types of scoring errors can either lead to an over/underestimation of the $2n_{xy}$ factor in this equation.

Errors due to technical shortcomings of RAPDs may be ameliorated by the fact that they sample the whole genome and have a high throughput such that if more markers are used, precision should improve (Moser and Lee, 1994, Tivang *et al.*, 1994). Moreover, the PCR analyses should be replicated. With regard to PCR-based microsatellite analysis, the problem may be genome coverage since some families of microsatellite loci may be restricted to specific regions of the chromosome (Lu *et al.*, 1996). Sampling of the genome targeting as many families as possible should circumvent this problem.

The most effective approach to study genetic diversity should involve a combination of the different marker systems to maximise the number of polymorphisms, and thus informativeness. However,

since practical considerations may necessitate that one or two marker systems is used due to cost, any of the PCR-based methods should suffice provided the number of polymorphisms is high to reduce the error rates. The observations that the estimates of genetic relationships as detected by PCR-based markers (RAPD, AFLP, and microsatellite-derived) are in general agreement with those of RFLP and isozyme analyses (dos Santos *et al.*, 1994, Heun *et al.*, 1994, de Oliveira *et al.*, 1996, Lu *et al.*, 1996, Powell *et al.*, 1996, Yang *et al.*, 1996), coupled with low developmental costs and high throughput, mean that more studies on genetic diversity will make use of the PCR-based methodologies especially in crop plants where RFLP probes are in short supply.

2.2.2 Construction of genetic maps in plants

The abundance of DNA markers has opened up the area of genetic mapping. Many crop species previously inaccessible to genetic analysis because of the paucity of morphological and isozyme markers have been or are in the process of being genetically mapped (O'brien, 1990, Table 2.2). The process of map construction is mediated by the selection of an appropriate population segregating for many markers, that are in turn used for linkage analysis.

2.2.2.1 Linkage analysis

Linkage analysis in map construction involves the calculation of pairwise recombination frequencies between markers, establishing linkage and determining the type of linkage (coupling or repulsion) (Staub *et al.*, 1996b). Once this has been achieved, map distances and map orders can be ascertained. Even though the algorithms for inferring pairwise linkages and calculating recombination frequencies between markers are fairly simple, the computation of the number of possible linkages is a daunting task. This means that all of the mapping efforts currently underway have been increasingly making use of computer software to simplify the task. Some of the most

widely used computer software packages include Mapmaker (Lander *et al.*, 1987), MapManager, (Manly and Elliot, 1991), DrawMap (van Ooijen, 1994) and JoinMap (Stam, 1993).

Linkage between any two points (markers) is a function of recombination frequency i.e. the further two points are away from each other, the more likely they are to undergo recombination (Suzuki *et al.*, 1989). The test of linkage between markers therefore involves quantification of recombinants in the progeny i.e. the percentage of recombinants, with the upper limit being 50%. The recombination frequency enables the calculation of genetic map distance. The two most commonly used mapping functions for assigning map distances are the Haldane (1919) and Kosambi (1944) functions, expressed as centiMorgans (cM). The difference between the two mapping functions is that the Haldane function assumes no interference, while the Kosambi function assumes positive interference.

Interference is an important parameter as it pertains to the rate of recombination and the factors that influence it (Suzuki *et al.*, 1989). Assuming non-interference implies that recombination (as defined by cross overs) is independent. Translated to map distance, this means that in the calculation of these values an assumption is made that linkage between any two points is independent of any other linkages, which may have an effect on the estimation of map size and the relationship between genetic map distance and physical distance (Suzuki *et al.*, 1989). However, the detection of double recombinants and presence of recombination hot spots indicates that cross over events are not independent (Suzuki *et al.*, 1989). Recombination heterogeneity has been reported in several species such as hexaploid bread wheat (Dvorak and Chen, 1984, Chao *et al.*, 1989, Gill and Gill, 1994) and bean (*Phaseolus vulgaris*) (Vallejos *et al.*, 1992). In addition, the more physically separate two points are (within the 50% limit), the more likely is the possibility of

double crossovers. This implies that in addition to the effect of the number of marker loci, recombination rate is also an important determinant in the estimation of map size.

In interspecific and other wide crosses there is recombination repression with result that the length of the genetic maps based on such crosses may be underestimated. An illustration of this factor is the rice map described by Kurata *et al.* (1994), which is 355 cM longer than the map of Tanksley *et al.* (1993). The earlier map was defined by 214 positions (Tanksley *et al.*, 1993) while the map published in 1996 (Kurata *et al.*, 1994) had 927 marker positions. One reason for the discrepancy in the size of the maps is that the map of Tanksley *et al.* (1993) was based on an interspecific cross, while the map of Kurata *et al.* (1994) was intraspecific with a higher chance of recombination.

Another consequence of variable recombination rate is that the ratio between genetic map distance and physical distance will vary according to the region of the genome. It also means that map distances cannot be directly extrapolated between, and within species. For example, in the tomato genome, 1cM equals 900 kilobases (kb) on average. However, in the *Tm2a* region of chromosome 9, 1cM is equivalent to 4-16Mb (Ganal *et al.*, 1989). In wheat the kb: cM ratio is 3.5Mb on average in the genome, while 1cM is approximately equal to 1Mb in the α-amylase gene on chromosome 6 (Cheung *et al.*, 1991). The relationship between the genetic distance in maps, and actual physical distance is crucial to the success of map-based gene cloning (Arondel *et al.*, 1992, Leyser *et al.*, 1993), through chromosome walking (Bender *et al.*, 1983) or chromosome jumping (Tanksley *et al.*, 1995).

Once two point linkages have been established, the next step is ordering of markers into linkage groups. The linkage groups can then be coalesced into homologous groups or chromosome groups

on the basis of common probes (Da Silva *et al.*, 1993), or linkage phase. Coupling phase linkages, i.e. markers linked on the same homolog (cis) are useful for determining homologous groups. Repulsion phase linkages, i.e. linked markers residing in different homologues of the same chromosome (*trans*), are used to identify a chromosome map. The total map distance and order can then be inferred from chromosomal maps.

Table 2.2 Examples of genetic maps published for various crop plants/species indicating the marker and population types (and map length in cM where given).

Crop/ plant	Marker	Population	Map	Reference
Brassica rapa	RFLP	F_2	1876	Chyi *et al.*, 1992
Arabidopsis thaliana	RAPD	RI	630	Reiter *et al.*, 1992
Conifers	RAPD	MG*	8873.8	Tulsieram *et al.*, 1992
Phaseolus vulgaris L	RFLP	F_2	827	Nodari *et al.*, 1993
Vicia faba	Isozyme, RFLP,	F_2	350	Torres *et al.*, 1993
Oryza sativa	RFLP, RAPD,	F_2	1575	Kurata *et al.*, 1994
Pinus spp.	RAPD	F_1	1380	Kubisiak *et al.*, 1995
Theobroma cacao L.	Isozymes,	F_2	759	Lanaud *et al.*, 1995
Pinus pinaster	RAPD	MG*	1380	Plomion *et al.*, 1995
Beta vulgaris	RAPD,	RI	738	Uphoff and Wricke, 1995
Brassica napus	Isozyme,	DH	1765	Foisset *et al.*, 1996
Pacific Yew (*Taxus brevifolia*)	RAPD	MG*	305.8	Göcmen *et al.*, 1996
Mimulus	RAPD	BC_1	2474±	Lin and Ritland, 1996a
Zea mays	Microsatellites	R_I	33bins**	Senior *et al.*, 1996
Carica papaya L.	RAPD	F_2	999.3	Sondur *et al.*, 1996
Brassica rapa	RFLP, RAPD	F_2	519	Tanhuanpää *et al.*, 1996
Eucalyptus	RAPD	F_1	1415	Verhaegen and Plomion, 1996
Glycine max	RFLP, RAPD	RI	3441	Keim *et al.*, 1997

* denotes haploid megagametophytes, ** denotes bins, about 20cM each (Coe *et al.*, 1995)

2.2.2.2 Choice of population

The selection of an appropriate mapping population is critical to any mapping exercise. The choice of parents for constructing mapping populations in most crop species is often based on maximising the number of polymorphisms (Paterson and Wing, 1993). Many genetic linkage maps are based on segregating backcross or F_2 populations, in which the gametes have undergone only a single cycle of recombination, and are rarely recombinant between two loci (Paterson and Wing, 1993).

This means that, assuming the species is diploid, the number of segregating alleles is restricted to two, and the linkage phase can be determined *a priori*. Most of the primary genetic maps of crop species (O'brien, 1990, Table 2.2) are based on segregation of randomly selected molecular markers of single populations. The choice of wide crosses for mapping populations is necessary, since at the initiation of genetic mapping such strong linkage disequilibrium is necessary to facilitate the establishment of linkage groups between widely dispersed loci (Paterson and Wing, 1993).

2.2.2.3 Choice of a marker system

The type of a marker system to be utilised for genetic mapping is also crucial, and linked to the type of population. For co-dominant marker systems like RFLPs, a completely classified F_2 population usually suffices (Tanksley *et al.*, 1988a). This is because these populations provide maximum genetic information (Mather, 1938, Allard, 1956). On the other hand dominant markers such as RAPDs, only provide maximum information in populations where most or all loci are homozygous since they are not able to differentiate homozygosity from heterozygosity. Such populations are recombinant inbred lines (RI) or doubled haploids.

The marker system used is also crucial for efficient genome coverage, since different marker systems may preferentially amplify different regions of the genome. For example, the telomeric

regions with an abundance of highly repetitive sequences (McEarchen and Blackburn, 1994) may not be accessible to RFLP analysis (Foolad *et al.*, 1993).

The effect of marker type on genome coverage, and hence the length of the genetic map obtained, is also a function of the number of markers and individuals used. For example, the length of the sugar beet map described by Uphoff and Wricke (1995) is estimated at 738 cM with 94 markers, while Pillen *et al.* (1993) estimated the size of the map at 1057cM with 177 marker loci.

2.2.3 Local mapping

As genetic maps of plant crops become more saturated with molecular markers the need to focus on specific regions, controlling important agronomic traits, has become more apparent. However, the paradox is that the choice of parents for mapping populations in most crop species has been based on maximising the number of polymorphisms that could be generated genome wide, rather than on the polymorphisms between parents for genetic loci controlling traits of agronomic importance (Paterson and Wing, 1993).

One of the earliest examples of the efficacy of using DNA pools, as a method for rapid location of markers linked to specific genetic regions, was demonstrated by Arnheim *et al.* (1985) when they examined RFLP polymorphisms within *HLA* class II loci associated with susceptibility to insulin-dependent diabetes mellitus in humans. Since then, several important strategies have been developed to identify DNA markers linked to specific genetic regions.

Some of the earlier DNA pooling strategies in plant crops were based on the comparison of near isogenic lines (NILs) differing for a specific phenotype in order to target markers to a specific

genetic region (Young *et al.*, 1988, Paran *et al.*, 1991). NILs are time consuming to make in species for which controlled backcrosses are possible, and may be impossible in slowly maturing species (Kesseli *et al.*, 1993).

In species where NILs are readily available as part of the breeding strategy or simple to construct, such as some cereals like wheat, oat (*Avena sativa* L.) and rice, pooling strategies using these populations have found widespread use. Some examples of the use of NILs for successful identification of markers linked to traits such as disease resistance in cereals are the rice bacterial blight disease resistance locus *Xa21* (Ronald *et al.*, 1992), powdery mildew resistance gene *Mla* in barley (Schüller *et al.*, 1992), the oat stem rust resistance gene *Pg3* (Penner *et al.*, 1993b), leaf rust resistance genes, *Lr9* and *Lr24* in wheat (Schachermayr *et al.*, 1994, Schachermayr *et al.*, 1995 and Dedryver *et al.*, 1996) and the wheat *Bt-10* common bunt resistance gene (Demeke *et al.*, 1996).

Translocation lines have also been successfully used in wheat to identify and map a RAPD marker linked to the resistance gene *Pm21* for powdery mildew (Qi *et al.*, 1996). This gene was introduced from *Haynaldia villosa*, a wild relative of wheat. Doubled haploid populations have also been widely utilised in conjunction with pooling strategies to identify markers linked to loci controlling important agronomic traits in crop plants. Examples include markers for *H1* gene conferring resistance to the potato cyst nematode *Globodera rostochiensis* in potato (*Solanum tuberosum*) (Pineda *et al.*, 1993) and common root rot and spot blotch resistance in barley (Kutcher *et al.*, 1996).

Michelmore *et al.* (1991) described a method, BSA, for rapidly isolating molecular markers linked to traits of interest in plants as long as there was a population segregating for the trait of interest.

BSA has emerged as an efficient way to map target regions instead of randomly mapping many markers. It has also allowed target mapping in species where NILs are difficult or impossible to create since F_2 progeny can be used to construct bulks. In fact, any type of population can be used for BSA, as long as it is segregating for the trait of interest and the phenotype or genotype can be accurately assayed. The added advantage of BSA is that a genetic map is not a pre-requisite for tagging the genetic locus.

Backcross lines in conjunction with their recurrent parents have also been used with BSA to locate markers linked to traits like root-knot nematode (*Meloidogyne arenaria*) resistance in peanut (*Arachis hypogaea* L.) (Burow *et al.*, 1996). The utility of BSA has also been demonstrated by its use in conjunction with comparative mapping to identify molecular markers linked to the gene for extreme resistance to potato virus Y (Hämäläinen *et al.*, 1997). This gene (Ry_{adg}) was tagged with an RFLP marker TG508, derived from a subset of 64 tomato and RFLP markers.

Perhaps one of the most impressive demonstrations of the utility of bulk segregant analysis has been in identifying molecular markers linked to traits in forest trees, which are classed among the most genetically complex plant species because of their large DNA content (Sobral and Honeycutt, 1994) and allogamy (Plomion and Durel, 1996). Until fairly recently there was no genetic framework or suitable populations for constructing genetic maps in this species (Sobral and Honeycutt, 1994). However, with the advent of BSA several putative markers linked to important genes have already been identified, viz. the genes for: sex determination in *Pistacia vera* (Hormaza *et al.*, 1994), resistance to white pine blister rust (Devey *et al.*, 1995), leaf colour in Copper beech (*Fagus sylvatica* L. var. atropunicea) (Heinze and Geburek, 1995), and pendula phenotype in Norway spruce (*Picea abies* (L.) Karst. f. *pendula*) (Lehner *et al.*, 1995).

Bulk segregant analysis to detect markers linked to quantitatively inherited traits is dependent upon a number of factors, the most important of which is determination of the number of loci controlling the trait *a priori*, or the heritability estimate as an indication of the presence of QTLs with large effects. Since classical methods of identifying quantitative traits have not been able to estimate the number of genes involved in the expression of most of these traits, application of BSA to identify markers linked to QTLs has been somehow hampered. However, despite this limitation, success has been reported with the identification of RAPD markers linked to genes controlling linolenic acid and erucic acid in rapeseed (Jourdren *et al.*, 1996a, Jourdren *et al.*, 1996b). Jourdren *et al.* (1996a) isolated six RAPD markers, in two linkage groups, linked to genes implicated in the control of the level of linolenic acid in rapeseed. The two QTLs ascribed 24% and 30.7% linolenic acid respectively. The success of this particular experiment was due to the fact that low expression of this trait is caused by two independent mutations. These studies demonstrate the potential of the BSA technique to successfully identify genetic markers linked to QTLs, provided the locus ascribes a major portion of the trait phenotype.

2.2.4 Comparative mapping

Comparative mapping is the study of similarities and differences in gene order along the chromosomes of taxa that cannot be hybridised (Paterson, 1997). This kind of mapping study is based on analysis of orthologous genetic loci in different plant taxa. This method was pioneered in studies where tomato RFLPs were used to map the tomato, chilli pepper (*Capsicum annuum* L.), and potato genomes (Bonierbale *et al.*, 1988, Tanksley *et al.*, 1988b, Gebhardt *et al.*, 1991)

One of the benefits of comparative mapping is that it offers a common genetic framework for interpreting genetic information such as evolutionary divergence. The early studies by Bonierbale *et al.* (1988), Tanksley *et al.* (1988b) and Gebhardt *et al.* (1991) revealed that more distantly

related species accrue a greater number of re-arrangements between their genomes than do closely related species. They were able to determine the location of three paracentric inversions differentiating potato from tomato. Furthermore, resources used in one species can be used in another, for example, cross utilisation of probes among different species in the same taxonomic family. This enables an increase in the number of genetic markers and map density for many genera simultaneously, thus providing tools, among others, for the detailed studies of transmission genetics and the analysis of quantitative trait loci.

One of the most studied families using comparative mapping is the *Brassicaceae* (Slocum, 1989, McGrath and Quiros, 1991, Kowalski *et al.*, 1994). The most extant forms of *Brassica* are highly duplicated (even diploid forms) which makes it difficult to study their genome organisation (Slocum, 1989, McGrath *et al.* 1993). Comparative mapping of this family has been facilitated by using the model crop, *Arabidopsis thaliana* (n = 5), which is a close relative to plants within the genus Brassica (Kowalski *et al.*, 1994).

In the comparative analysis of genetic maps of different *Brassica* species a recurring picture of conservation of gene order punctuated by occasional chromosome re-arrangements is evident (Paterson, 1997). This relationship is further supported by the extensive conservation of coding sequences between *Brassica* and *A. thaliana*. Kowalski *et al.* (1994) found that although extensive chromosomal rearrangements have occurred since the divergence of *B.oleracea* and *A. thaliana*, islands of conserved organisation are still discernible. They found a total of eleven regions of conserved organisation detected from 3.7 to 49.6 cM spanning 158.2 cM (24.6%) of the *A. thaliana* genome, and 245 cM (29.9%) of the *B. oleracea* genome. At least 17 translocations and 9 inversions distinguish the two genomes.

While certain large regions of *Brassica* chromosomes revealed common gene order in different species, both inversions and translocations were found to differentiate among the diploid *Brassica* genomes (Slocum, 1989, Slocum, *et al.*, 1990, McGrath and Quiros, 1991). Tetraploid genomes show considerable correspondence with the diploid genomes although there are also significant differences (Paterson, 1997). Cultivated *Brassica* strains appear to have the same gene order as wild strains of the same species, but it is clear that chromosomal rearrangements subsequent to the origin of polyploid *Brassica* have differentiated the tetraploid subgenomes from their corresponding diploid genomes (Paterson, 1997). This rapid evolution of *Brassica* chromosomes is in stark contrast to that of other species (Kowalski *et al.*, 1994).

The family *Poaceae* includes most of the world's major food and feed crops such as rice, maize, wheat, sorghum, sugarcane, barley, oat, rye, millet, and others. Whitkus *et al.* (1992) investigated comparative mapping in sorghum and maize to find out if: 1) the maize and sorghum genomes maintained collinearity with respect to RFLP and isozyme loci and 2) the two genomes shared duplicated loci and the same constellation of duplicated linkage groups, or if the independent loss of duplicated sequences (random diploidisation) eliminated previously shared duplicated genomic regions. Maize low copy number nuclear DNA probes and isozyme loci in an F_2 population derived from a cross of sorghum were used (Whitkus *et al.* 1992). Results showed that many linkage groups are conserved between sorghum and maize and that the amount of recombination is roughly equivalent in the two species (Whitkus *et al.* 1992). The pattern of conserved linkages between maize and sorghum is such that most sorghum linkage groups are composed of loci that map to two maize linkage chromosomes (Whitkus *et al.* 1992).

The average size of the conserved linkage groups is 52.2cM in sorghum versus 53.9cM in maize with nine inversions of locus order within shared linkage groups (Whitkus *et al.*, 1992). Thirty

eight percent of loci were duplicated in the sorghum genome versus 72% in maize, while 31% showed duplication in both sorghum and maize genomes. A total of 40% showed duplication in maize but not in sorghum with 6.7% duplicated in sorghum but not in maize and 21% specific for a single locus in both species (Whitkus *et al.*, 1992). This data suggests that the primary processes involved in the divergence of maize and sorghum were duplications.

In the family *Poaceae* the utility of comparative mapping has been demonstrated by several sugarcane mapping efforts that have already benefited from using RFLP probes originally developed for mapping other members of this family such as maize, sorghum, oat, etc (Da Silva *et al.*, 1993, D'Hont *et al.*, 1994, Grivet *et al.* 1996). An RFLP linkage map of a *S. spontaneum* anther-derived, haploid population SES208 was constructed using probes from heterologous DNA libraries of oat, barley, rice and maize (Da Silva *et al.*, 1993). This was in addition to probes used from sugarcane DNA libraries. Eight of the 15 maize probes utilised had also previously been mapped in sorghum. Three rice cDNA probes, previously mapped on maize were also mapped in sugarcane. Oat and barley cDNA probes produced poor signals and constituted the smallest proportion of probes hybridising to multiple fragments, which is in line with the evolutionary relationship of these species with sugarcane (Hamby and Zimmer, 1988).

Previous studies by Hulbert *et al.* (1990) similarly showed that a high proportion of maize genomic probes, as well as cloned DNA fragments from characterised maize genes, hybridised to DNA of a sugarcane interspecific hybrid, as well as to that of sorghum. More recent application of the comparative mapping principle in sugarcane has seen Grivet *et al.* (1996) utilise probes from maize genomic libraries (BNL and UMC), oat cDNA library, and two cloned genes from wheat rDNA

and maize (*Adh1* gene) to successfully map a population of a selfed progeny of the commercial sugarcane variety R570.

Finally, in addition to providing tools for analysis of evolutionary genetics, transmission genetics and map saturation in the form of markers, comparative mapping also generates information about location of genetic loci controlling important traits mapped in one species, to be translated and used in another related species. This is made possible by the large amount of conservation at the overall gene order among distantly related plants such as rice, maize and *Triticeae* (Paterson, 1997).

In addition, studies on tomato QTLs have revealed that few genes may be responsible for the bulk of phenotypic variation, this observation forms one of the pretexts for using comparative mapping to dissect quantitative trait loci (Paterson *et al.*, 1988). The potential for utilising comparative mapping, *inter alia*, to locate quantitative trait loci in crop species is discussed in the next section.

2.2.5 Dissection of Quantitative Trait Loci (QTLs)

The identification of molecular markers linked to traits of agronomic importance is crucial for plant breeders because selection for such traits can then be performed at early stages of development (Winter and Kahl, 1995). This would have the effect of increasing the number of progeny to be assessed, since plants which do not fulfil the selection criteria can be discarded in the early stages instead of carrying these wrong genotypes until the late stages of selection (Allen, 1994).

The possibility of locating genes affecting quantitative trait loci was first demonstrated when Sax (1923) associated seed coat characters with seed size in bean. However, the dearth of genetic markers, due to reliance on morphological and isozyme markers, had considerably slowed progress.

This meant that characterisation and genetic analysis of quantitative traits for breeding purposes had largely been based on biometrical approaches (Mather and Jinks, 1974, Falconer and Mackay, 1997). The limitations of these were that they dealt mainly with collective characterisation of multiple factors affecting a quantitative trait by partitioning the overall phenotypic variance into its genetic and environmental components approaches (Mather and Jinks, 1974, Falconer and Mackay, 1997). Consequently, they were not able to characterise or manipulate specific loci.

The advent of molecular markers, coupled with the development of genetic maps, has meant that specific quantitative loci can now be characterised and manipulated. The mapping of QTLs can be broadly divided into three steps viz. 1) detection and location of markers linked to the quantitative trait, 2) estimation of the effect of the quantitative trait locus, and 3) characterisation of the number of loci controlling the trait. Most of the recent developments in QTL mapping have incorporated the first two steps, while the development of the third awaits comprehensive map saturation and statistical tools.

2.2.5.1 Detection and location of QTLs

The ability to detect QTLs with molecular markers is dependent upon a number of factors such as the magnitude of the effect of the QTL on the trait, the type and size of the experimental population, and the recombination frequency between the marker and trait locus. Several algorithms or statistical models have been developed for QTL mapping in a wide range of pedigrees and experimental designs, including F_2, backcross, recombinant inbred and other designs (Soller *et al.*, 1976, Lander and Botstein, 1989, Knapp, 1991, Haley and Knott, 1992, Jansen, 1992, Martinez and Curnow, 1992, Moreno-Gonzalez, 1992, Jansen, 1993, Zeng, 1994, Charcosset and

Gallais, 1996, Doerge and Rebai, 1996, Fisch *et al.*, 1996, Jansen, 1996, Lin and Ritland, 1996b, Muranty, 1996, Plomion and Durel, 1996, Satagopan *et al.*, 1996).

The basic premise of QTL mapping rests on the idea that genetic markers that tend to be transmitted together with specific values of the trait are likely to be close to the gene affecting the trait (Soller *et al.*, 1976, Lander and Botstein, 1989). Markers themselves cannot be considered as QTLs. In essence, the classical linkage analysis involves marker variants and genetic loci controlling a phenotype. This means that associations between marker variants and trait values are sought using various statistical methods, which take into account the various parameters that affect this association, such as recombination between marker locus and trait locus, gene/locus dosage, linkage phase and gene action (Soller *et al.*, 1976, Lander and Botstein, 1989).

2.2.5.2 Statistical parameters in QTL detection and location

Statistical methods used to detect QTLs generally suppose that QTL act on the phenotypic trait mean (Muranty, 1996). In the traditional approach (Soller *et al.*, 1976) marker genotypes are considered as factors in a t-test or a one way ANOVA, and the test for significance of the factor is the test for the presence of a QTL in the vicinity of the marker locus. In other words the search for QTLs is done using one marker at a time.

One of the disadvantages of this method is that the phenotypic effects of QTLs are systematically underestimated, and the genetic locations of the QTLs are not well resolved because the distant QTLs of large effect cannot be distinguished from proximal QTLs of small effects (Lander and Botstein, 1989). In addition, the number of progeny required for QTL detection is generally high (Soller and Genizi, 1978).

Another disadvantage is that the chance of detecting individuals with the marker, but without the genetic locus, or false positives (type 1 error) with single-marker methods is significantly high (Lander and Botstein, 1989). This is because for a given threshold value, e.g. α=0.05, the assumption does not take into consideration the presence of other markers, such that the chosen alpha value is only significant at that locus but genome wide the chances of false positives is higher (Lander and Botstein, 1989).

The failure of the single-marker analyses to accurately locate QTLs, and the problem associated with estimation of the QTL effect, can be circumvented by interval mapping i.e. use of two flanking markers (Lander and Botstein, 1989). However, despite these limitations, this method may be the only applicable one in situations where genetic maps have very few markers, and in cases where local mapping is done using only phenotypic data as in BSA.

The availability of a large number of markers, and the saturation of genetic maps to varying degrees have provided opportunities to develop statistical methods that are more efficient in the detection and location of QTLs. More effective methods of QTL detection utilise flanking markers to detect and locate QTLs (Lander and Botstein, 1989, Knapp, 1991, Haley and Knott, 1992, Martinez and Curnow, 1992, Jansen, 1993, Jansen and Stam, 1994, Zeng, 1994). By definition, these methods presuppose a marker map with varying degrees of density. They either employ a maximum likelihood (Lander and Botstein, 1989, Zeng, 1994), or the regression approach (Knapp, 1991, Haley and Knott, 1992, Martinez and Curnow, 1992, Jansen, 1993, Jansen and Stam, 1994, Zeng, 1994). One advantage of the regression approaches is the computing speed (Xu, 1995). There is little difference between the maximum likelihood method and regression analyses, with regard to their efficiency (Martinez and Curnow, 1992, Haley and Knott, 1992).

The success of these methodologies for QTL resolution is demonstrated by the increasing number of reports detailing associations between markers and traits of agronomic importance such as yield and quality traits (Table 2.3).

Table 2.3 Examples of QTL mapping for phenotypic traits in crop plants

Crop plant	**PT**	**Trait**	**PVE (No. of QTLs)**	**Reference**
Loblolly Pine (*Pinus taeda* L.)	F_3	Wood specific gravity	23% (5)	Groover *et al.*, 1994
Maize (*Zea mays*)	F_1	Grain yield	35% (4)	Ajmone-Marsan *et al.*, 1995
	RIL	Anthocyanins	18.5% (8)	Sourdille *et al.*, 1996
Soybean (*Glycine max*)	F_2	Maturity	57.4% (2)	Lee *et al.*, 1996
		Plant height	67.7% (1)	
		Lodging	56.4%	
Barley (*Hordeum vulgare*)	DH	Callus growth	47.8% (2)	Mano *et al.*, 1996
		Shoot regeneration	49.8% (4)	
Rice (*Oryza sativa* L.)	F_2	Seedling vigour		
Wheat (*Triticum aestivum* L.)	RIL	Tan spot resistance	49% (4)	Faris *et al.*, 1997
Oat (*Avena sativa* L.)	RIL	Grain yield	39% (3)	Siripoonwiwat *et al.*, 1997
		Test weight	26% (4)	
		Groat percentage	17% (2)	
		Days to heading	15% (3)	
		Plant height	34% (5)	

PT = type of population, PVE = total phenotypic variation ascribed by markers,

The basic methodology for detection of markers linked to QTLs is based on two experimental designs where two inbred lines are chosen as parents. Usually the two parents differ substantially

for the quantitative trait value, presumably because they carry different alleles at the trait locus. Segregating progeny are scored both for the trait, and for a number of genetic markers. The two experimental designs involve the scoring of segregating progeny in a B_1 backcross (F_1 x Parent) or an F_2 intercross (F_1 x F_1). Figure 2.1 is an illustration of this design.

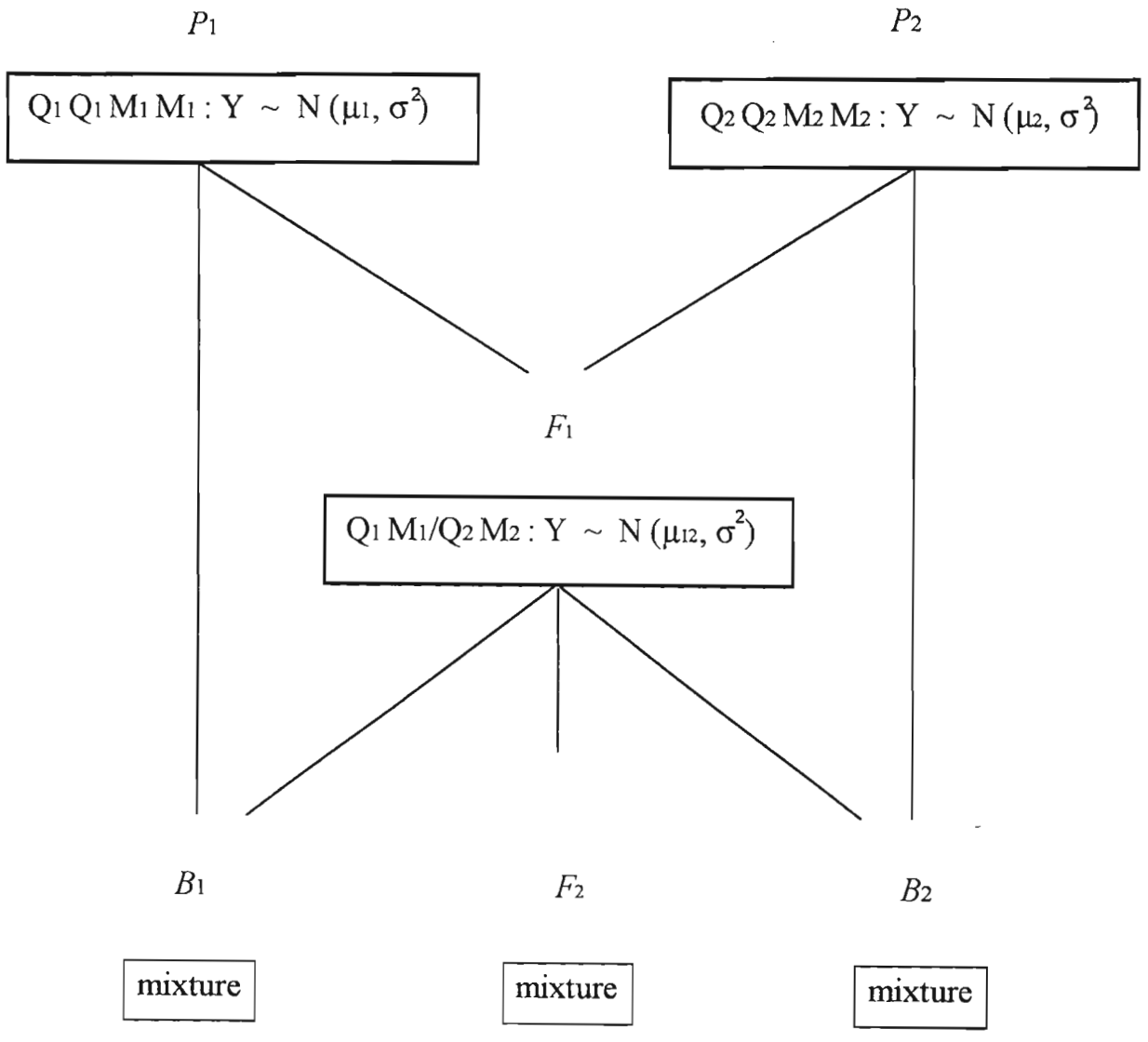


Figure 2.1 Backcross and F_2 designs for QTL mapping. P_1 and P_2 are the parents, M_1 and M_2 are the marker classes and B_1 and B_2 refer to backcrosses (Doerge *et al.*, 1997).

In the single marker at a time approach, mentioned above, QTL detection is carried out through the comparison of marker means using the t-test or a single factor ANOVA to declare linkage of

the marker to a genetic locus controlling the trait (Soller *et al.*, 1977). The difference between the marker means is taken as an estimation of the phenotypic effect of substituting the allele from parent 1 (P_1), with that of parent 2 (P_2) (Soller *et al.*, 1977). Under a completely additive model, the trait means of the F_1 progeny is the average of the two parental means. Since all three groups, P_1, P_2 and F_1 progeny are genetically uniform, they are assigned the same trait variance, σ^2 (Doerge *et al.*, 1997). In the backcross and F_2 generations the individuals have mixtures of trait and marker genotypes depending on the recombination fraction between the two loci (Doerge *et al.*, 1997).

Interval mapping as described by Lander and Botstein (1989), makes use of the maximum likelihood method to estimate the phenotypic effect and LOD score for a putative QTL at any given genetic locus. Trait loci are postulated to occur at a series of positions within a set of adjacent marker intervals. The LOD score is an indication of the probability of the distribution of the phenotypic data being due to the observed genotypes. Declaration of a significant QTL is made when the LOD score exceeds a predetermined threshold, T. This is equivalent to regressing the observed trait data on the number of trait alleles (Doerge *et al.*, 1997).

The basic methodology has been refined to cater for different marker densities in genetic maps (Feingold *et al.*, 1993, Dupuis, 1994, Rebaï *et al.*, 1994). Threshold values for declaring a significant QTL are dependent on the size of the genome and on the density of the markers genotyped (Doerge and Rebaï, 1996). Deviations from normality of the trait distribution and sample size also affect the LOD score (Doerge and Rebaï, 1996). Simulation results indicate that the Rebaï *et al.* (1994) method is the most appropriate for intermediate density maps (markers every 10 cM or more), while the Lander and Botstein (1989), Feingold *et al.* (1993), and Dupuis (1994) methods are the most appropriate for high density maps i.e. one marker every 10 cM or less

(Doerge and Rebaï, 1996). These thresholds provide stringent values that ensure that the type 1 error rate is less than the chosen significance level (Doerge and Rebaï, 1996).

Different experimental models affect the power of detection in QTL mapping. Rebaï *et al.* (1994) computed the relationship between phenotypic variation associated with a marker locus and the power of tests for various experimental situations, namely the type of population and the number of individuals used in the experiment. It was found that for a given value of r^2 and a given number of individuals, F_2 populations are generally inferior to recombinant lines because of their degrees of freedom, two as opposed to one for recombinant inbred populations (Rebaï *et al.*, 1994, Rebaï and Goffinet, 1993). This is because F_2s have three marker classes, whereas recombinant inbreds have two, thereby increasing the number of parameters.

Precision of QTL estimation is also a function of the number individuals in the mapping experiment (Soller and Genizi, 1978, Lander and Botstein, 1989, Darvasi *et al.*, 1993). The larger the number of individuals included in the experiment, the more precise the estimation of the QTL effect, as well as chances of detecting QTLs with small to medium effects. This is especially true since most marker loci do not lie at the actual QTL. Selective genotyping has been suggested as a method to increase the precision of QTL estimation, while simultaneously decreasing the number of individuals genotyped (Lander and Botstein, 1989, Darvasi and Soller, 1992), the rationale being that individuals at the phenotypic extremes provide the most linkage information.

However, selective genotyping can bias the estimates of recombination frequency between markers and linked QTLs, upwardly in repulsion phase or downward in coupling phase (Lin and Ritland, 1996b). The bias is a function of the proportion of individuals selected (25% of each extreme

according to Lander and Botstein, 1989), the magnitude of the QTL effects, distance between the QTLs, and the dominance effects (Lin and Ritland, 1996b).

Another bias associated with interval mapping is that of 'ghosting', where a QTL can be incorrectly detected in an adjacent marker interval, in addition to the correct interval (Paterson *et al.*, 1991, Doerge *et al.*, 1997). One way to avoid this bias is to use the multiple QTL method (Jansen, 1993, Jansen and Stam, 1994, Zeng, 1994), where QTLs are simultaneously evaluated throughout the genome (Jansen, 1996). An added advantage of this method is that it takes into consideration the possible effects of other loci on the trait of interest, which reduces the unexplained variance and increases the power of QTL mapping. Furthermore, the simultaneous use of all marker data compensates for missing marker information by utilizing information from neighbouring markers (Jansen, 1996).

2.2.5.3 Influence of environment on QTL detection

Paterson *et al.* (1991) reported sensitivity of some QTLs to different environments. They detected a total of 29 QTLs across three environments, 14% of which were common in all three environments, 34% were detected in two environments, and the majority (52%) were specific to a single environment. Since then several reports of QTLs specific to particular environments have been published (Bubeck *et al.*, 1993, Tinker *et al.*, 1996, Brummer *et al.*, 1997, Siripoonwiwat *et al.*, 1997, Mather *et al.*, 1997). However, other studies have revealed a higher proportion of QTLs that are consistent across environments for yield traits (Stuber *et al.*, 1992, Schön *et al.*, 1994, Ajmone-Marsan *et al.*, 1995, Veldboom and Lee, 1996).

2.2.5.4 Non-genetic map based methods of QTL resolution

In addition to using genetic maps specifically constructed for a particular population, species wide QTL location has been enhanced by the advent of comparative mapping (Lin *et al.*, 1995, Liu *et al.*, 1996, Paterson *et al.*, 1995a,b). This has enabled information of loci controlling important traits mapped in one species to be translated in another related species.

To reiterate, studies carried within grasses have revealed a large amount of conservation at the overall gene order level among distantly related species such as rice, maize and *Triticeae*. A close correspondence among QTLs affecting complex traits such as seed size has been shown for sorghum, sugarcane, maize, wheat, barley, and rice (Lin *et al.*, 1995, Paterson *et al.*, 1995a)

Paterson *et al.* (1995a) reported QTLs controlling three traits namely seed mass (size), seed dispersal (shattering) and flowering in sorghum, maize and rice. It was found that three QTLs that affect seed mass correspond closely in these three crops. Seven QTLs accounting for 52% phenotypic variance explained (PVE) in sorghum, correspond to five of the eight QTLs that account for 78% PVE in rice (Paterson *et al.*, 1995a). Four of the sorghum QTLs correspond to four of the eight QTLs that account for 69% of PVE in maize and five maize QTLs correspond to rice QTLs (Paterson *et al.*, 1995a). The probability that seed mass in sorghum, maize, and rice would correspond so frequently by chance was conservatively estimated at 0.1-0.8% (Paterson *et al.*, 1995a). QTLs for seed dispersal mapped to a single locus (100%) in sorghum, three loci (24%) in rice, and ten loci (60%) in maize (Paterson *et al.*, 1995a). The discrete sorghum locus corresponds to a rice QTL on chromosome 9, and to maize QTLs duplicated on chromosomes 1 and 5. Rice QTLs on chromosomes 2 and 3 correspond to maize QTLs on chromosomes 1 and 4 (Paterson *et al.*, 1995a).

Comparative mapping analysis has also been used to investigate correspondence among QTLs affecting height and/or flowering across five races of sorghum, an interspecific sorghum F_2 population, and 32 previously published sorghum, maize, rice, wheat, and barley populations (Lin *et al.*, 1995). One chromosomal region, putatively the *Dw2* gene, explained 54.8% of height variation in the interspecific F_2 and 85.7% of flowering time variation (putatively *Ma1*). The corresponding region of maize chromosome 10 accounts for up to 26% of PVE in the flowering of a temperate x tropical cross (Paterson *et al.*, 1995a). Of particular interest to sugarcane is that short flowering is closely associated with the DNA probe pSB188 which lies near *Mal*. In barley and wheat, the corresponding regions in the short arm of group 2 homeologues harbour photoperiodic mutants

Another method that has great potential in terms of speeding up the resolution of QTLs uses the large amount of information available from metabolic studies. Byrne *et al.* (1996) have used the wealth of information regarding the flavonoid pathway to dissect the QTL for corn earworm resistance in maize. The major objective of their study was to identify the genetic factors controlling the concentration of maysin, a C-glycosyl flavone produced in maize silks that is toxic to the corn earworm (Byrne *et al.*, 1996). A population developed from a cross of a high maysin inbred line (GT114) by one producing only trace amounts of maysin (GT119) was used. Maysin concentrations and RFLP genotypes at flavonoid pathway loci were determined using 285 F_2 plants. A major additive effect in the p1 region of chromosome 1 was detected, accounting for 58% of the phenotypic variance explained. The p1 locus is a Myb-transcription activator of the portion of the flavonoid pathway leading to maysin (Byrne *et al.*, 1996).

Recently, several studies have reported successes with the identification of molecular markers linked to QTLs using germplasm pools (Virk *et al.*, 1996, Beer *et al.*, 1997). The potential

problems associated with this approach is that the different recombination distances between trait and marker loci, in different accessions, may confound location and effects of QTLs. The implications of this are that estimation of the QTL effect is almost always biased. However, as a method for rapidly identifying linkages between markers and loci, and useful DNA markers that can be cross utilised, it provides a good starting point for QTL mapping (Beer *et al.*, 1997).

2.2.6 Marker Assisted Selection (MAS)

2.2.6.1 Prospects for MAS

Lee (1995) has summarised some of the conditions under which the efficiency of MAS can be enhanced: (1) trait(s) under selection has low heritability, (2) tight linkage between QTL and markers (<5cM) with additional efficiency realised when coupling linkages predominate (Gimelfarb and Lande, 1994) and (3) in earlier generations of selection prior to fixation of alleles at or near marker loci and recombinational erosion of marker - QTL associations.

MAS is only expected to accrue greater genetic gains at early stages of selection as at later generations MAS approximates, or is even less, than conventional breeding (Edwards and Page, 1994). It may be an important tool for introgressions (exotic x elite) and germplasm conversion programs due to coupling phase linkages and very close linkage between markers and QTLs (Edwards, 1992). MAS has been demonstrated to be an effective means of transferring genes (regions) for hybrid grain yield in elite maize inbreds (Stuber, 1994). Empirical studies were conducted on two maize inbreds, B73 and Mo17, which were backcrossed for three generations using MAS (Stuber, 1994). Chromosomal regions affecting grain yield had been identified earlier with RFLPs through QTL mapping (Stuber and Sisco, 1991). Each backcross-derived B73 and Mo17 was top-crossed to the normal B73 and Mo17 for comparative purposes. It was found that

51 of the 141 enhanced B73 conversions and 51 of the 114 converted Mo17's exceeded the grain yield of the normal hybrid by at least one standard deviation. Only 10 and 15, respectively, of the converted lines had lower yields. This study and others on processing traits (Edwards and Johnson, 1994) have demonstrated the potential for MAS to manipulate complex traits.

The greatest utility of molecular markers is likely to be in introgression of disease resistance genes and this is also the area that molecular markers are likely to find the most immediate application. This is because most of the successes with tagging of molecular markers closely linked to genes of economic importance has been in this area particularly with DNA pooling and BSA investigations (Wang and Paterson, 1994). One reason for these successes is the fact that disease resistance is usually a simply inherited or oligogenic trait, hence the ease with which linked markers have been found (Kesseli *et al.*, 1993). Furthermore, in most cases resistance in commercial cultivars is mostly introgressed from wild species which makes the tagging of the genomic regions flanking the resistance genes more amenable (Michelmore *et al.*, 1992).

Molecular markers tagging disease resistance genes, used in conjunction with highly saturated genetic maps will help with the identification of genes that are unlinked and therefore can be combined easily in breeding lines (Maisonneuve *et al.*, 1994). Furthermore, molecular markers will eliminate the need to test plants for resistance to pathogen isolates that are not locally available as this may have the effect of introducing the pathogen to a new area (Melchinger, 1990).

The potential for molecular markers to tag individual resistance genes that are located in clusters has already been demonstrated (Michelmore *et al.*, 1991, Kesseli *et al.*, 1993, Maisonneuve *et al.*, 1994). The immediate benefit of this is that they can be used to screen recombinants that carry

most or all of the resistance genes since in introgression programmes these linkages may be broken during backcrossing, resulting in their loss. This is especially crucial in cases where the action of some of the genes is not apparent because epistatic genes mask their effect.

Map position indicates which resistance genes are unlinked and therefore can be combined easily (Michelmore *et al.*, 1992). It also indicates which genes might be lost when introgressing another when the resistance genes are linked. Molecular markers can be used to identify recombinants with new combinations of resistance genes linked in coupling which can then be taken as single Mendelian blocks (Kesseli *et al.*, 1993). Additionally, markers can be used to monitor the introduction of multiple resistance genes, especially in cases where it is difficult to use a foreign isolate of the pathogen in field testing (Melchinger, 1990), or to test for multiple resistance simultaneously (Young and Kelly, 1996).

2.2.6.2 Technical feasibility of MAS

Since MAS is expected to be more effective in the early stages of selection, its application as a practical plant breeding tool is dependent on the automation of most of the methods hitherto used for genetic analyses. This is primarily due to the fact that large numbers of progeny are involved in the early selection stages of most breeding programmes. For example, in the SASEX sugarcane breeding programme, 180 000 individuals are evaluated in the first selection stage (Blose, 1992). Since PCR is expected to be the method of choice in MAS, and its automation is at an advanced stage, the bottlenecks in the analytical process are the DNA extraction methods, and the resolution of PCR products.

Traditional DNA extraction and purification methods are very time consuming and expensive,

which mitigates against their application as practical tools for MAS. Recently, Clancy *et al.* (1996) reported the use of alkali-treated barley leaf tissues as templates for direct PCR. Previously direct PCR on plant tissues had been reported (Berthomieu and Meyer, 1991, Langridge *et al.*, 1991, Klimyuk *et al.*, 1993, Wang *et al.*, 1993). These methods are likely to find increasing usage in MAS application.

The recent advances in the design of PCR thermocyclers, has integrated laser scanning tools that monitor the progress of PCR throughput (Livak *et al.*, 1995). These kinetic thermocyclers, which require the use of fluorescent dye-labelled primers are expected to dispense with the need for time consuming electrophoretic systems for PCR product resolution, and will provide efficient screening tools in MAS.

2.3 GENETIC ANALYSIS IN SUGARCANE

2.3.1 Origins, taxonomic classification and transmission genetics of the major *Saccharum* species

Sugarcane is a large grass of the family Gramineae belonging to the Andropogoneae tribe (Jeswiet, 1925). The Andropogoneae include some of the most important food crops such as cereals (sorghum, maize), fodder (*Iseilema*, *Themeda*, etc), sugar (sorghum, *Saccharum*, maize), essential oils (*Cymbogon*, etc) (Simmonds, 1959, Bor, 1960). In addition, some members of the tribe Andropogoneae (*Imperata*, *Miscanthus*, etc) are used for archery, thatching, and housebuilding (Simmonds, 1959, Bor, 1960).

Cultivated sugarcane belongs to the genus *Saccharum*, thought to have arisen initially as a result of natural hybridisation between the genera *Saccharum*, *Erianthus* (section=Ripidium),

Sclerostachya, and *Narenga* (Mukherjee, 1957). These genera constitute a closely related interbreeding group referred to as the *Saccharum* complex (Mukherjee, 1957). Sugarcane and members of the *Saccharum* complex are thought to have arisen as a result of polyploidization and hybridisation events (Price, 1963). Chromosome numbers in these species ranges from 2n = 20 in some *Erianthus* (Michx. section Ripidium Henrard) to 2n = 200 in some hybrid *Saccharum* clones (Daniels and Roach, 1987).

Members of the genus *Erianthus* have chromosome numbers in the range of 2n=20, 30, 40, and 60 (Daniels and Roach, 1987). The genus *Sclerostachya* is characterised by chromosome numbers which are mainly 2n=30 (Nair and Ratnambal, 1967), although some 2n=34 cytotypes have been found (Sreenivasan and Sreenivasan, 1984). *Sclerostachya* is generally in the middle range of the *Saccharum* complex, and is readily distinguished morphologically from *Saccharum* by the absence of root eyes and poorly developed buds (Daniels and Roach, 1987). The genus *Narenga* is closely related to *Sclerostachya* (Daniels and Roach, 1987).

The most important genus in the *Saccharum* complex as far as sugarcane is concerned is *Saccharum*, which comprises six species (Brandes, 1958). The cultivated species are *S. officinarum* L. characterised by thick stalks, *S. barberi* Jeswiet (the sugarcanes of India) and *S. sinense* Roxb. (the sugarcanes of China) (Daniels and Roach, 1987). The latter two species are readily distinguished from *S. officinarum* by their floral characteristics, thin to medium stalks, low fibre, low to moderate sucrose concentrations and their tolerance to stress (Daniels and Roach, 1987). The most widely cultivated and important in terms of genetic contribution to modern cultivated sugarcanes, are the *S. officinarum*s, and a subgroup of these was originally termed the nobles for their high sucrose content (Mukherjee, 1954).

The wild species of the genus are the *S. robustum* Brandes and Jeswiet ex Grassl found primarily in Indonesia and New Guinea (Grassl, 1946). *S. robustum* has been suggested as the wild progenitor of *S. officinarum* (Daniels, 1973). This species is characterised by two forms with chromosome numbers 2n=60 and 2n=80 (Price, 1965, Daniels and Roach, 1987). The second one is *S. edule* whose clones are characterised by their edible inflorescence and widely used as a vegetable in New Guinea and Melanesia (Massal and Barrau, 1956). *S. edule* has been suggested as a mutant form of *S. robustum* because of their close morphological likeness (Lennox,1939 and Brandes *et al.*,1939). *S. edule* is a polyploid series of 2n=60, 70, and 80 with some aneuploid forms (Roach, 1972).

The third, and the most important, in terms of genetic contribution to the modern sugarcane germplasm is *S. spontaneum* L. This wild cane is the most widely distributed species of the genus and its geographical distribution stretches from New Guinea to the Mediterranean and Africa (Panje and Babu, 1960). The wide geographical distribution of *S. spontaneum* attests to its high adaptability (Stevenson, 1965). Its chromosome number ranges from 2n=48-128 (Panje and Babu, 1960).

2.3.2 Application of molecular markers in the study of sugarcane genetics

In addition to the fact that many of today's commercial varieties are a product of a few crosses between *S. officinarum* and *S. spontaneum*, the practice of clonal propagation for sugarcane using stem cuttings has been the only form of cultivation for millenia. Consequently, sugarcane as a crop has a narrow genetic base in terms of potentially useful new alleles from its wild relatives (Berding and Roach, 1987). This is despite the fact that it can be argued that high ploidiness of sugarcane bequests it with substantial variation in terms of combinations of alleles available for each cross (Skinner *et al.*, 1987).

2.3.2.1 Taxonomy, phylogenetics and genetic diversity of sugarcane

Assessment of genetic relationships between members of the *Saccharum* complex and modern hybrids is important in the collection and maintenance of sugarcane breeding germplasm. The estimation of genetic similarities of the varieties in the germplasm is essential in the crossing section of the breeding programme.

Before the advent of DNA markers, flavonoid chemotaxonomic markers as a form of differentiating between the *Saccharum* species were widely utilised (Daniels and Roach, 1987, section 2.1.4). A leaf flavonoid marker, F13, specific to *Erianthus* section Ripidium has been detected in some *S. spontaneum* clones indicating the possible introgression of *Erianthus* in the genome of *S. spontaneum* (Daniels *et al.*, 1980). Isozymes have also been used to study the genetic diversity of sugarcane and assess the evolutionary relationships between different *Saccharum* species (Glaszmann *et al.*, 1989, Eksomtramage *et al.*, 1992). Electrophoretic variation of nine enzymes, in conjunction with multivariate analyses in a sample of 39 wild and noble sugarcane clones, was used to isolate an *Erianthus* clone because of its unique pattern for most enzymes (Glaszmann *et al.*, 1989). In addition, *S. spontaneum* was separated from the *S. robustum* and *S. officinarum* clones on the basis of their isozyme banding patterns (Glaszmann *et al.*, 1989). The latter two were not differentiated from one another, supporting the contention that *S. robustum* is the progenitor of *S. officinarum* (Daniels, 1973).

Eksomtramage *et al.* (1992) also used isozyme markers to assess genetic variability in sugarcane. In a sample of *S. officinarum*, *S. barberi*, and *S. sinense* clones and some commercial varieties, they identified six varietal groups. Their results indicated that the *S. officinarum* clones displayed

a restricted diversity, and the variation in the commercial varieties sampled was mainly attributed to the *S. spontaneum* component.

Some of the limitations of protein based markers have already been discussed (section 2.1.4). In sugarcane these shortcomings are amplified by the high ploidy. The presence of many alleles, co-existing in the same plants at different dosages, may result in a high number of bands migrating at very close distances, thus confounding analyses (Glaszmann *et al.*, 1989). Furthermore, since the detection of some of the isozymes is dependent on developmental stage of the plants, which may in turn be environmentally modulated, it means plants may have to be cultivated locally for efficient analysis (Eksomtramage *et al.*, 1992).

DNA based markers provide a better alternative to circumvent these limitations and several studies have already demonstrated their utility to assess genetic diversity in sugarcane (D'Hont *et al.*, 1993, Al-Janabi *et al.*, 1994a, Harvey *et al.*, 1994, Lu *et al.* 1994a, Lu *et al.*, 1994b, Sobral *et al.*, 1994, Harvey and Botha, 1996, Besse *et al.*, 1997). The results of Lu *et al.* (1994a) using maize RLFP probes allowed the separation of the three basic *Saccharum* species implicated in the evolution of modern sugarcane hybrids, *S. spontaneum*, *S. robustum* and *S. officinarum*. The placing of the secondary species *S. barberi* and *S. sinense* between *S. officinarum* and *S. spontaneum*, supported the hypothesis that *S. barberi* and *S. sinense* are derivatives of natural hybridisation between *S. officinarum* and *S. spontaneum* (Brandes, 1958). The genotypes of *S. spontaneum* were also distinguished on the basis of their geographical origin, i.e. the Southeast Asian clones were separated from those of Indian origin (Lu *et al.*, 1994a). *Erianthus* and *Miscanthus* clones were confirmed as representatives of distinct genera. Genetic variability was found to be highest in *S. spontaneum*, followed by *S. robustum*, and lowest in *S. officinarum* which

is consistent with the results of isozyme analyses (Glaszmann *et al.*, 1989, Eksomtramage *et al.*, 1992).

RFLP analysis, using maize DNA probes, has also been used to estimate average genetic similarity of 40 cultivated sugarcane varieties (Lu *et al.*, 1994b). The average genetic similarity was found to be 0.61, with cultivated clones closely related to *S. officinarum* clones (Lu *et al.*, 1994b). The results further confirmed the observation that the principal component of varietal diversity is due to the *S. spontaneum* contribution.

PCR-based DNA markers have also been used to estimate the genetic diversity in the *Saccharum* varieties (Harvey *et al.*, 1994, Harvey and Botha, 1996). In a sample of 20 commercial varieties, including *S. spontaneum*, *S. officinarum* and early interspecific hybrids, RAPD, microsatellite, and telomere marker data was used to show that there has been an erosion of DNA diversity (84%) from the early interspecific crosses to the modern varieties (Harvey and Botha, 1996).

The variation in cytoplasmic DNA has been used in several studies to assess genetic variation and evolutionary relationships in the *Saccharum* complex (D'Hont *et al.*, 1993, Al-Janabi *et al.*, 1994a, Sobral *et al.*, 1994). The results of these studies indicate a lack of variation in the chloroplast genome, while differences were detected in mitochondrial DNA. Mitochondrial DNA is known to evolve faster than chloroplast DNA. D'Hont *et al.* (1993) showed that *S. spontaneum* could easily be differentiated from *S. robustum* and *S. officinarum*. A large diversity was detected in the *S. spontaneum* clones, which is in line with the results of nuclear DNA analyses, however, there was no clear relationship between the diversity and the geographical origin of the clones (D'Hont *et al.*, 1993). The *S. officinarum* and *S. robustum* clones generally displayed the same patterns, which confirms the closeness of the two species (Daniels, 1973).

The results of Sobral *et al.* (1994), using 12 chloroplast DNA probes from rice with 15 restriction enzymes on 32 genotypes representing eight genera and 19 species from the Andropogoneae, showed that the genera *Narenga*, *Sclerostachya* and *Saccharum* form a closely related monophyletic group. This is a confirmation of their grouping in the *Saccharum* complex. In contrast, another presumed member of the *Saccharum* complex, *Erianthus*, was found to have significantly different chloroplast genomes. The results showed that the *Saccharum* species have the same chloroplast restriction sites. The only variation found within the *Saccharum* genus was in *S. spontaneum*, which is concordant with the results of D'Hont *et al.* (1993). Phylogenetic analyses using chloroplast and mitochondrial DNA sequences, confirmed the limited variability in the maternal genomes in the *Saccharum* species (Al-Janabi *et al.*, 1994a). The lack of variability in chloroplast DNA in cultivated sugarcane is indicative of a worldwide cytoplasmic monoculture (Al-Janabi *et al.*, 1994a).

The separate grouping of *Erianthus* is in agreement with the results of Besse *et al.* (1997) who found that *Erianthus* species are not closer to either *S. officinarum* or *S. spontaneum*. The two genera i.e. *Erianthus* and *Saccharum* had more than 40% of their RFLP bands in common, and an intergeneric distance of 0.748 was found. These results show that the involvement of *Erianthus* in the evolution of *S. officinarum* is more complex than originally hypothesised (Besse *et al.*, 1997).

2.3.2.2 Genetic mapping in sugarcane

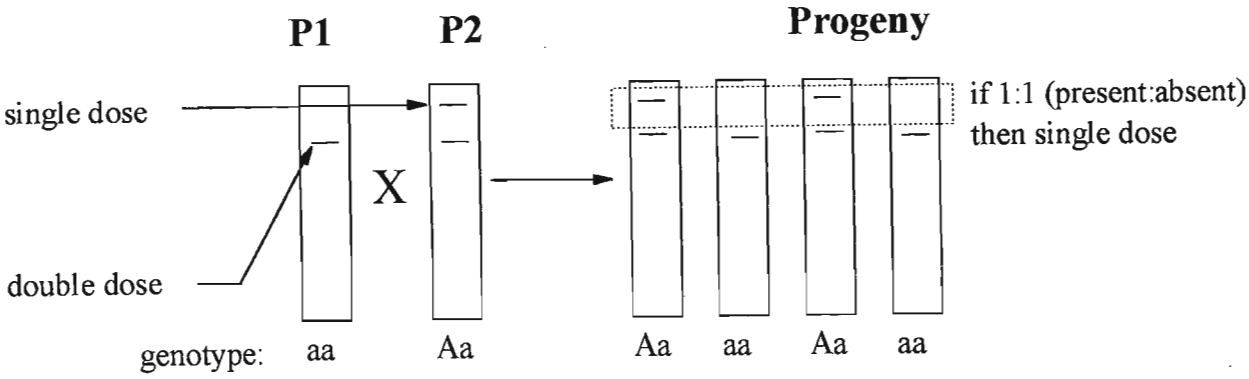
Sugarcane has lagged behind most major crop species in terms of the development of genetic maps because of its high ploidy and complex genetics. Some of these complexities, as mentioned above, are the aneuploid nature of sugarcane and the possible multiple origin of its genome. Construction of genetic maps on polyploids is difficult because of the high number of possible marker classes,

co-migration of fragments during agarose gel electrophoresis, and the unknown ploidy, making it difficult to determine patterns of inheritance. Furthermore, there are no suitable diploid relatives that could be used to develop maps that could be extrapolated for use in sugarcane genetics. These factors meant that the usual linkage theory of Mendelian inheritance had to be modified to cater for the peculiar genetics of sugarcane.

In 1992 Wu *et al.* described a genetic framework that could be used to map sugarcane. Single dose restriction fragment analysis has enabled the mapping of such markers in polyploids. Basically, in a polyploid where the genotypes are unknown, the segregation of a polymorphic fragment in a 1:1 ratio (absence: presence) in a population is inferred *a posteriori* as representing a single dose allele. If the fragment segregates in a 3:1 ratio it is inferred as representing a double dose allele in one of the parents (Figure 2.2). The population size required to detect single dose fragments with 98% level of confidence for four ploidy levels (2n=4x, 6x, 8x, and 10x) was calculated to be 75 individuals (Wu *et al.*, 1992). This size is adequate in detecting and estimating linkages in the coupling phase for both autopolyploids and allopolyploids, but linkages in repulsion can only be estimated in allopolyploids. The size of the progeny for estimating linkages in repulsion in autopolyploids was estimated at 750.

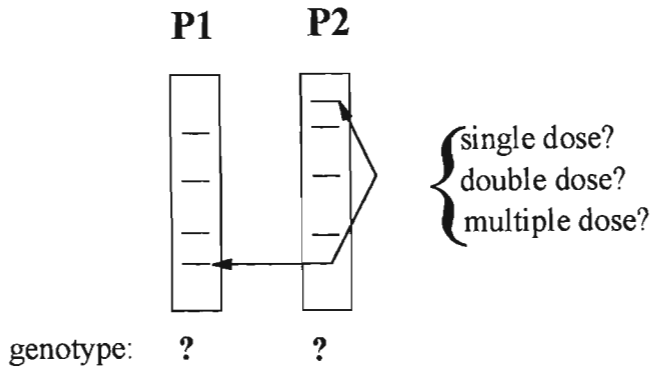
The linkage analysis described by the theory of single dose restriction fragments is accurate within a recombination fraction of 25cM, at 95% confidence limit, for 75 individuals (Wu *et al.*, 1992). A consequence of single dose restriction fragment analysis is that the mapping information content of RFLP markers is equivalent to PCR-based markers, like RAPD markers (Da Silva and Sobral, 1996). This is due to the fact RFLP fragments revealed by single copy probes are analysed separately. To increase the information content, the mapping theory has been extended to include double dose, and other higher dosage fragments (Da Silva, 1993, Da Silva *et al.*, 1995).

A. Diploid Analysis

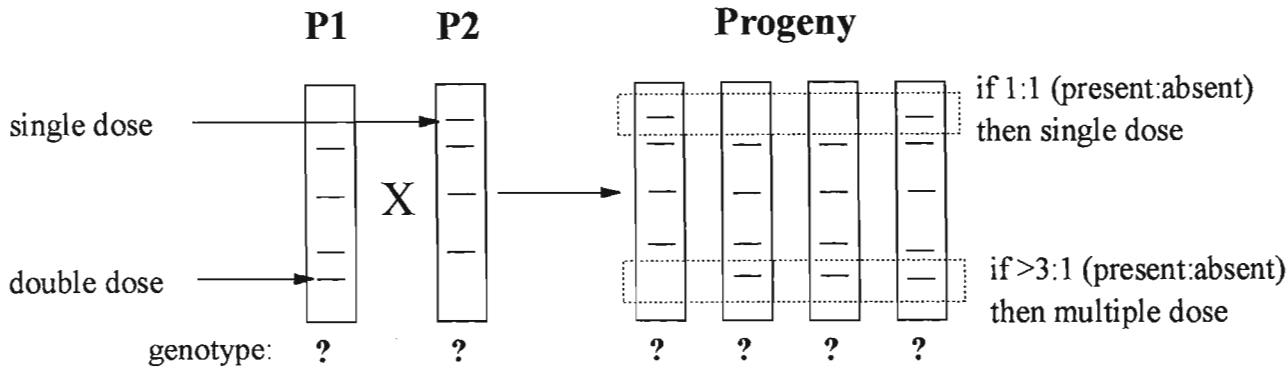


B. Disomic Polyploid Analysis

Parent Analysis



Progeny Test



'igure 2.2. Comparison of DNA polymorphisms between diploid and polyploid parents and among their progeny (Moore nd Wu, 1991).

. Diploid Analysis. Parents and progeny genotypes are recognised as either homozygous (aa or AA) or heterozygous (Aa) y the number of hybridising fragments. Linkage analysis is usually based on segregation of the genotypes, but it could also e on segregation of the single dose fragments.

. Disomic Polyploid Plant Analysis. The genotype of polyploids is not known. If a polymorphic band segregates as 1(present:absent) in either diploid or polyploid progeny populations, then that allele is single dose. However, if the olymorphic band segregates as >3:1 (present:absent), then it is due to multiple doses in one of the parents. The progeny test r distinguishing segregation ratios of 1:1 from >3:1 will have a 98% confidence level on a progeny population of 75 dividuals. This population size is adequate for distinguishing linkages with recombination fractions of 0.36 to 0.50 at the)% level of confidence.

In a preliminary demonstration of the efficacy of the theory of single dose fragments, Burnquist (1991) identified thirty-two single dose fragments in the cross SES208 (2n = 64) x ADP0068 (2n = 64), a doubled haploid derived from an anther culture of SES208 (Fitch and Moore, 1983). A total of 18 RFLP markers were placed on eight linkage groups (Burnquist, 1991). The first genetic maps of sugarcane were described by Al-Janabi *et al.* (1993) and Da Silva (1993) for a *S. spontaneum* clone, 'SES 208'. These maps represented the first time that genetic maps were directly constructed on a polyploid species with neither a classical linkage map nor a well-defined genetic system, and no diploid relatives. Both maps were based on the segregation of single dose markers in the progeny of a cross between a *S. spontaneum* clone and an anther culture derived doubled haploid of itself ('SES 208' x ADP85-0068) (Fitch and Moore, 1983).

Al-Janabi *et al.* (1993) used the arbitrarily primed polymerase chain reaction (AP-PCR) to screen 88 progeny of the cross for single dose fragments. Forty two linkage groups, with at least two markers each, spanning a genetic distance of 1500 cM were detected with two and three point linkage analyses (LOD scores of 9.0 and 6.0 respectively). The average distance between markers was estimated to be 30 cM, and 85% genome coverage was inferred based on an estimation of the total genome size of 2547 cM. None of the linkages were in repulsion phase, which is indicative of autopolyploidiness.

Da Silva *et al.* (1993) screened 91 progeny of the same cross with RFLP probes from the sugarcane genomic and cDNA libraries. Maize, rice, and oat probes were also used. The genetic map was constructed with 216 RFLP markers, distributed over 44 linkage groups at an average density of 25 cM per marker. The map covered approximately 86% of the genome, which was estimated to be 2107 cM, which is slightly smaller than Al-Janabi *et al.*'s estimation of 2547 cM.

No repulsion linkages were detected, indicating that the ploidy type of this *S. spontaneum* clone is autopolyploid. The marker segregation information of 71 common progeny for the two genetic maps was utilised to create an integrated map of this *S. spontaneum* clone (Da Silva *et al.*, 1995). This integrated map had over 500 markers at an average density of 5 cM per marker.

In a preliminary analysis, Andersen and Fairbanks (1994) scored a sample of 84 mapping progeny for the segregation of 108 single dose RAPD fragments in a cross of *S. officinarum* (La Purple) x *S. robustum* (Mol 5829). The preliminary map of *S. officinarum* contained 51 markers over 21 linkage groups in coupling. Repulsion phase linkages were also detected, which implies that *S. officinarum* is at least a segmental allopolyploid. Another implication is that the complex genome of *S. officinarum* may be in the process of diploidisation (Andersen and Fairbanks, 1994).

Most of the described mapping efforts have been based on the base members of the genus *Saccharum*. However, for genetic maps to have more impact in terms of identifying markers linked to important traits, elite crosses have to be mapped because they have higher linkage disequilibrium between markers and trait loci (Da Silva, 1993). Recently, Grivet *et al.* (1996) used the selfed progeny of a commercial cultivar, R570, to generate a 2008 cM genetic map, with 128 RFLP probes and one isozyme. The map had 480 markers, which were placed in 96 cosegregation groups, based on linkage in coupling only. These were then coalesced into 10 linkage groups on the basis of common probes, and the average distance between marker loci was 6.1 cM, though their distribution was irregular, with both sparsely and densely populated regions. Eighty markers were specific to *S. officinarum*, and 66 markers were *S. spontaneum* specific; and occasional recombination was detected between the two genomes. Over half of the bands detected were present in both *S. officinarum* and *S. spontaneum*.

3.2.3.3 Resolution of the ploidy type in Saccharum

The recent advances in the acquisition of molecular markers and construction of genetic maps in sugarcane, has enabled detailed study of sugarcane genetics with regard to ploidy type (auto or allo) and inheritance patterns (di- or polysomic).

Wu *et al.* (1992) described the theoretical framework under which disomic and polysomic inheritance could be inferred in sugarcane. Briefly it involves investigation of the marker linkage relationships to compare the expected and observed number of linkages in repulsion-phase. This is done by calculating the ratio of the probabilities for linkage in repulsion and coupling, and then applying it to the total number of two point linkages observed. Disomic inheritance can then be inferred with a Chi-squared test. In the polysomic inheritance situation, detection of repulsion: coupling linkages is dependent on the sample size (Wu *et al.*, 1992).

Segregation data can also be used to compare expected and observed proportions of single dose to multiple dose fragments (Da Silva *et al.*, 1995). The expected proportion of single dose to multiple dose fragments can be calculated from the expected segregation ratios of double, triple and quadruple dose fragments depending on the ploidy type. Under polysomic inheritance calculated values for single dose : multiple dose are 70:30, while in the case of disomic inheritance the values are 56:44 in a doubled haploid population. Da Silva *et al.* (1995) used both these analytical tools to conclude that *S. spontaneum* is an autopolyploid. No linkages in repulsion were detected with over 590 two point linkages that were examined in the SES208' integrated genetic map. Furthermore, the ratio of single dose to multiple dose markers was indicative of polysomic inheritance.

Linkages in repulsion were detected for both *S. officinarum* and *S. robustum* (the presumed progenitor of *S. officinarum*) genomes in a sample of 44 progeny of a cross between these two species (Al-Janabi *et al.*, 1994b, Sills *et al.*, 1995). In the *S. officinarum* genome, a map hypothesis gave 7 linkage groups with 17 linked and 33 unlinked markers. Four of the 13 pairwise linkages were in repulsion, and nine in coupling phase. In the *S. robustum* genome two out of nine pairwise linkages, in five linkage groups, were in repulsion (Al-Janabi *et al.*, 1994b). The detection of repulsion linkages, despite the small number of progeny and linkage groups indicates that these two species may be at least segmental allopolyploids (Al-Janabi *et al.*, 1994b, Sills *et al.*, 1995).

The data from the mapping of the commercial cultivar R570 (Grivet *et al.*, 1996), revealed repulsion phase linkages in 21 of the 96 co-segregation groups, indicating preferential pairing. The preferential pairing seemed to be stronger between chromosomes with a *S. spontaneum* origin (D'Hont *et al.*, 1994, Grivet *et al.*, 1996).

2.3.2.4 Transmission genetics in sugarcane

It has been known for some time that in interspecific crosses between the two species (*S. officinarum* x *S. spontaneum*), the female *S. officinarum* transmits its somatic number of chromosomes, while the male *S. spontaneum* only transmits its haploid number hence the 2n + n transmission (Bremer, 1923). Normal n + n chromosomal transmission does occur, but it is rare (Berding and Roach, 1987). However, D'Hont (1996) using the genomic *in situ* hybridisation (GISH) technique observed the n + n chromosomal transmission in an interspecific F_1 cross. Chromosome transmission between *S.officinarum* x *S.officinarum* is the normal n + n type (Berding and Roach, 1987).

Chromosome transmission in the crosses involved in the evolution of modern commercial hybrids follows the 2n + n pattern in the interspecific F1 hybrids of the *S. officinarum* x *S. spontaneum* crosses (Berding and Roach, 1987). The 2n + n transmission is maintained in the first backcross, after which, in the subsequent backcrosses, chromosome transmission reverts to the normal n + n type. This type of chromosomal transmission allows rapid recovery of *S. officinarum* agronomic traits such as high sucrose, and has been implicated in the added vigour of modern hybrids (Gill and Grassl, 1986). A consequence of this type of chromosome transmission is that only about 10% of the genome of modern sugarcane cultivars is derived from *S. spontaneum* (Simmonds, 1976, D'Hont *et al.*, 1996).

Recent work by Burner and Legendre (1993) has confirmed the 2n + n transmission in the *S. officinarum* x *S. spontaneum* crosses, and the endosperm balance number (EBN) concept (Johnston *et al.*, 1980) has been suggested as a way of explaining the possible mechanism for this type of transmission. Briefly, successful endosperm development in wide crosses depends on the endosperm receiving two EBNs from the female parent and one from the male parent (Johnston *et al.*, 1980). Two species with unlike EBNs can be crossed by doubling the chromosome number or use of the functional 2n megaspores of the female (Johnston *et al.*, 1980). The ploidy level of the elite sugarcane clones, unlike that of *S. officinarum*, is apparently at the EBN level of *S. spontaneum* which means that functional megaspores are unnecessary for zygote survival (Burner and Legendre, 1993).

2.3.2.5 Comparative mapping in sugarcane

In addition to the use of common probes from maize and other related species for the genome mapping effort in sugarcane (reviewed in section 2.2.4), a detailed study has recently been carried

out by Dufour *et al.* (1997) to ascertain the degree of synteny between sugarcane and sorghum genomes. This is of particular interest because of the evolutionary relationship between the two species. Whereas sugarcane is thought to have diverged from maize 60 million years ago, its divergence from sorghum is presumed to be closer to 5 million years which is a relatively short time on the evolutionary scale (Al-Janabi *et al.*, 1994a). Given the advanced coverage of the sorghum genome and QTL analyses (Lin *et al.*, 1995) it means that the search for important QTLs in sugarcane can be enhanced. It also means that the evolutionary mode of divergence between the two species can be deduced, which may provide some answers to the events leading up to the polyploidization of sugarcane.

The construction of the sorghum composite map, with two recombinant inbred line populations, was based on 188 probes previously mapped on maize and sugarcane (Dufour *et al.*, 1997). The probes revealed 199 loci on 13 linkage groups. The comparison between the sorghum composite map and a partially complete sugarcane map (10 linkage groups) was based on a set of 84 common probes. The results indicate that a straight collinearity exists between two pairs of linkage groups (linkage groups E and IV and linkage groups H and IX) (Dufour *et al.*, 1997). Large arrays of collinear probes of sugarcane were also observed along the other sorghum linkage groups (Dufour *et al.*, 1997)

2.3.2.6 Molecular markers linked to traits in sugarcane

Searching for markers linked to phenotypic traits in sugarcane is an almost intractable task because of the complexity of the genome and the high ploidy levels. The dearth of information on simply inherited traits (Hogarth, 1987), has also contributed to the slow progress in finding markers linked to phenotypic traits. Daugrois *et al.* (1996) reported the first monogenic inherited trait in

sugarcane that could be tagged with DNA markers. The segregation of rust resistance in the selfed progeny of the commercial cultivar, R570, displayed a clear cut 3:1 pattern indicative of a dominant resistant gene. An RFLP marker revealed by the probe, CDSR29, was estimated to be linked to the rust resistance gene at a distance of 10 cM. Other minor factors associated with resistance were also detected.

Sills *et al.* (1995) investigated associations of 84 single dose AP-PCR markers with recorded traits in 44 random clones using single degree of freedom analysis of variance and multiple regression. All markers, found to be linked to traits by ANOVA ($P<0.10$), were subjected to backwards elimination regression until only significant markers ($P<0.05$) remained in the model. These analyses were performed using the software package JMP (Stam, 1993) and Mapmaker/QTL (Lander *et al.*, 1987) to link QTL to previously mapped markers (Al-Janabi *et al.*, 1994). Significant associations between markers and the traits of stalk number, stalk diameter, plot weight, percent tasselled, %fibre cane, which ranged from 32% to 76% of the phenotypic variation explained, were detected (Sills *et al.*, 1995). These results represent some of the first attempts to detect associations of markers with phenotypic traits using molecular markers in sugarcane.

Chapter 3

Bulk Segregant Analysis

3.1 Introduction

Bulk segregant analysis (BSA) or pooling strategies have been used to identify markers linked to simply-inherited traits (Michelmore *et al.*, 1991) and in populations where NILs are available (Young *et al.*, 1988, Martin *et al.*, 1991). The majority of BSA or BSA- type studies have been on plants where genetic maps already exist, with pooling strategies based on known or characterised genomic intervals (Giovanonni *et al.*, 1991, Martin *et al.*, 1991, and Michelmore *et al.*, 1991). In sugarcane there are no NILs, and Kandasami *et al.* (1980) have shown that it is virtually impossible to create NILs in sugarcane due its high ploidiness and aneuploid nature.

Bulk segregant analysis based solely on phenotypic information has mostly been used to study simply-inherited traits (Wang and Paterson, 1994). In sugarcane there is a dearth of information on traits that are simply inherited. In fact, only a few attempts have been made to accurately characterise simply-inherited traits (Raghavan and Govindwasamy, 1956, Skinner, 1956, Batcha and Palanichamy, 1978, Wu *et al.*, 1983). One investigation was carried out by Skinner (1956) who studied the inheritance of hairiness, concentrating on Jeswiet hair groups 56, 57 and 60. However, doubts have been cast on the validity of these findings due to problems associated with the control of pollination contamination during the study, and no other trials have been carried out since (Stevenson, 1965).

Sobral and McClelland (1993) defined a trait most amenable to a bulk segregant analysis in sugarcane as one that would meet the following requirements. Firstly, it has to have high heritability; in this case heritability is taken to mean the reliability of the phenotype as a measure

of the genotype (Falconer and Mackay, 1997). Secondly, it has to have a large genetic spread, in other words there has to be a fairly large difference between the phenotypic extremes. Finally, it has to be an uni- or oligogenic trait i.e. controlled by one or a few major genes.

The population that can be used for a BSA needs to meet a few criteria, the most important of which is segregation for the trait of interest. Furthermore, such a population has to be the product of a good genetic contrast i.e. the parents of the cross have to differ sufficiently for the trait of interest. In addition, a population has to have at least 75 progeny so that the putative BSA markers can be tested for their segregation ratios and dosage (Wu *et al.*, 1992).

These considerations are very important, in view of the fact that in sugarcane genetic maps are still under construction, and the bulking would be based solely on the phenotype. However, it is important to note that some of these requirements could not be met in the present study primarily due to two limitations. Firstly, to restate, very limited work has been done to characterise the genetic control of sugarcane traits. Secondly, the populations that were available for the investigation had just been planted, and not sufficiently replicated. This meant that there was not enough information on the behaviour of traits over several seasons and environments, a very important consideration for determining heritability.

However, the possible distinct role played by the introgression of the *S. spontaneum* genome, coupled with the probability that modern sugarcane varieties could be hybrids of as many as four species (Price, 1963) may imply that some of the phenotypic traits in sugarcane are under the genetic control of loci introduced by single introgression events, and therefore readily resolvable. The example of the initial introgression of *S. spontaneum* genome as a means of introducing resistance to the Sereh disease in Java provides a case in point (Stevenson, 1965). It has been hypothesised that strong linkage disequilibrium may exist in the elite germplasm with regard to genetic loci controlling some of the important agronomic traits (Da Silva, 1993). If

this is the case, then it is possible that some of these genetic loci can be tagged, and BSA provided a potential method to rapidly investigate this possibility.

Finally, the extensive study of the heritability of traits in the South African breeding germplasm (Anon., 1988, Blose, 1992) provided an estimate of the behaviour of traits, especially under the uniform environmental conditions where the populations available for this study were planted. This means that the current phenotypic data could be used in conjunction with that of past studies to evaluate the most suitable trait for the BSA.

3.2 Materials and Methods

3.2.1 Population analysis

Six populations of 150 F_1 individuals each were planted in the field at Pongola, Mpumalanga. Planting was done using a randomised block design with lines of stools a metre apart. The parents of these crosses are part of the SASEX breeding collection (Appendix A). The fields at Pongola were irrigated, and the conditions kept as close to uniform as possible. This means that clonal variation due to the environment should be minimal, so that differences in phenotypic expression are largely genetically based.

3.2.1.1 Phenotypic traits

Six stalks from each clone were taken, ground and analysed with various millroom techniques (Anon., 1985) to yield phenotypic data for the traits of brix % cane, fibre % cane, pol % cane, estimated recoverable sucrose (ers) % cane and dry matter % cane (Appendix B). Phenotypic data was also available for purity, brix dry matter, and cane and sucrose grams per stalk. It was decided to analyse the four traits of brix %cane, fibre %cane, pol %cane and purity for their segregation in the populations. Brix % cane is defined as the mass of all soluble matter as a percentage of fresh cane, fibre %cane is defined as the mass of water-insoluble dry matter as

a percentage by mass of cane (Anon., 1985). Pol % cane is the apparent sucrose content of any substance expressed as a percentage by mass and purity is the percentage ratio of sucrose (or pol) to the total soluble solids (or brix) in a sugar product (Anon., 1985). The choice of these four traits was based on past studies of heritabilities and general behaviour of traits (Blose, 1992, K.Nuss, pers comm[1]) and the availability of phenotypic data.

Simple statistical analyses were performed on these populations to determine 1) the trait that was going to be investigated, and 2) the best population in terms of phenotypic differences between the extremes for that particular trait. The means, standard deviations, standard errors, and the coefficients of variation (cv %) were calculated for the phenotypic data using Lotus 1-2-3.

Using the phenotypic data the best normal distribution curves were drawn for each trait, the main aim being to find a trait that shows the best segregation pattern, and a desirable phenotypic spread. For each trait the data was divided into arbitrary classes and these were consistent only within traits so that the best distribution curves could be drawn.

Once the appropriate phenotypic trait had been chosen, and a suitable population selected, it was then replanted in three replicates in the same area. These were then assayed for the phenotypic trait at six and eight months and the individuals ranked according to their trait values so that those at the extremes could be identified for bulking.

3.2.2 Genomic DNA Isolation

Six to 10 grams of leaf roll tissue was used to isolate genomic DNA according to a modified method of Honeycutt *et al.* (1992). The leaf roll tissue was homogenised in 40ml of 50mM Tris-HCl buffer (pH 8.0) containing 5mM ethylenediaminetetraacetic acid (EDTA), 0.5mM spermidine, 1% (m/v) polyethylene glycol (PEG)-8000, 0.1% (v/v) 2-mercaptoethanol, and

1 K. Nuss, Head: Plant Breeding Department, SASEX, Mount Edgecombe, 4300

0.35M sucrose for 2 minutes, and filtered through two layers of mutton cloth. The filtrate was centrifuged at 5000x g for 20 minutes in a rotor pre-cooled to 4°C. The pellet was resuspended in 10ml wash buffer (same as homogenisation buffer except for 25mM EDTA and no PEG) and placed on ice. To this solution the following were added sequentially: 5M NaCl to a final concentration of 0.7M, 10% (m/v) sodium dodecyl sulphate (SDS), final concentration 0.7% (m/v) and 10% (m/v) cetyltrimethylammonium bromide (CTAB), final concentration (0.9% (m/v)). The resulting mixture was incubated at 60°C for 30 minutes and cooled at room temperature for 15 minutes. The proteins were extracted by adding an equal volume of 24:1 chloroform: isoamyl alcohol and centrifuged at 3500 x g for 10 minutes at 4°C. The aqueous phase was collected and the protein extraction repeated. The aqueous phase was collected and the DNA precipitated by addition of an equal volume of isopropanol. Precipitated DNA was then spooled out, dried and resuspended in 1.0ml of TE buffer (10mM Tris-HCl (pH 8.0), 1.0mM EDTA).

3.2.3 Quantification of the DNA

The DNA was quantified with a Beckman DU 7500 spectrophotometer at 260nm, and the relative purity assessed by the 260/280nm ratio. The DNA was then analysed on a 1% (m/v) agarose gel to confirm its concentration and quality. DNA samples were stored at -20°C until use.

3.2.4 Preparation of template DNA and primer solutions for the PCR

Stock genomic DNA solutions of the individuals and the two parents were diluted in TE buffer, to a final concentration of 3ng/µl. Bulks of five and 10 individuals from the extremes of the population were made to a final DNA concentration of 1ng/µl or 3ng/µl. A total of 60 random decamer primers from Operon Technologies, 20 each of the A, B, and C series, were diluted in sterile distilled water to a final concentration of 6µM and stored at -20°C.

3.2.5 Random amplification of polymorphic DNAs (RAPDs)

PCR was done as described by Al-Janabi *et al.* (1993). The final reaction volume of 21µl contained: 2 units of AmpliTaq (Stoffel fragment), 0.22µM primer, 0.1mM of each deoxynucleotide triphosphate (dNTP), 4.0mM $MgCl_2$, 10mM KCl, 10mM Tris-HCl (pH 8.3) 2µg acetylated bovine serum albumin and 7 or 21ng of template DNA. The reactions were carried out in a 96-well microtiter plate. Each reaction mixture was overlaid with 50µl mineral oil. Thermal cycling was done in a Hybaid OmniGene thermal cycler, using the following thermal profile 94°C/3minutes, 35°C/1 minute and 72°C/2 minutes with a ramp of $2.4°Cs^{-1}$; then 40 cycles of 94°C/1 minute, 35°C/1 minute and 72°C/2 minutes with a $2,4°Cs^{-1}$ ramp. The final elongation step was at 72°C for 7 minutes. All the reactions were performed, at least, twice.

3.2.6 Gel electrophoresis.

The PCR products were analysed on a 2% (m/v) agarose gel which was developed at 4 V/cm for 2-3 hours in 0.5 x TBE buffer (45mM Tris-HCl, 44mM boric acid and 1 mM EDTA, pH8). The gel was stained using ethidium bromide at a concentration of 1µg/ml for 30 minutes with constant shaking, and destained for 40 minutes in fresh TBE buffer. The gel was visualised under ultra violet (UV) light and photographed with a COFU camera.

3.2.7 Scoring of the RAPD fragments

The polymorphisms between the bulks were visually scored, first under UV light and then in the gel photographs. Two approaches were employed in the scoring of polymorphic fragments between bulks. Firstly, fragments were scored as polymorphic on the basis of their presence or absence in the bulks. Secondly, they were scored solely on the basis of their presence in the individuals comprising the bulk. The criterion was that a fragment should be present in four individuals in the bulk of five on one extreme of the trait, while one aberrant individual could be tolerated in the opposite bulk. The molecular weights of the fragments were calculated from

a calibration curve constructed from the relative mobility of molecular weight markers (2000bp – 150bp).

3.2.8 Marker segregation analysis

The bulk specific polymorphisms were then analysed for their segregation in 80 randomly selected individuals of the same cross. In addition, segregation of the other polymorphisms between the parents was investigated.

Chi-squared tests were carried out on the data for the segregation of RAPD markers to test the goodness of fit for the 1:1 and 3:1 segregation ratios, according to Suzuki (1989). Pairwise and multipoint linkage analyses of the RAPD markers, for construction of genetic linkage groups, were investigated using Mapmaker 2.0, an interactive software programme (Lander *et al*, 1987).

3.2.9 RAPD marker - trait linkage analysis

An analysis of variance (ANOVA), using one marker at a time, to determine whether any of the putative markers were linked to the fibre trait was carried out. This was done using data from the sample of 80 individuals of the population AA157. The marker data (presence or absence) was fitted as the independent variable, and the phenotypic trait value as the dependent variable. This work was performed using the GRNORMAL procedure of Genstat (1993).

Markers found to be significantly linked to the fibre trait were then subjected to a regression analysis, to determine the percentage variation (as a measure of R^2) each of them contributed to the fibre trait in this population. The regression analysis was of the form:

$$y = b_0 + b_1x + b_2x + \ldots + b_nx_n$$

where y is the dependent phenotypic value, and x is the independent marker variable and b_0 is the constant. This was carried out using the GRNORMAL procedure of Genstat (1993).

Analysis of variance and regression analyses were also carried out using 22 random RAPD markers.

3.3 Results

3.3.1 Population and phenotypic trait analysis

The frequency distributions of the four traits (purity, pol %cane, brix %cane, and fibre %cane) were similar between the six populations. The distribution of purity is consistently left-skewed (Figure 3.1). This was true for all six populations analysed. There are very few individuals on the lower extreme of the purity distribution.

The frequency distribution curves of pol %cane were normally distributed with a bias towards high levels. Most of the curves are narrow i.e. the extremes are close to the mean (Figure 3.2). In some cases the phenotypic spread is distorted by the presence of individuals that can be considered as outliers and therefore confound the analysis of the 'true' trait distribution. This was observed in the crosses AA1292 and Z927.

The results of frequency distributions for both purity and pol % cane are probably due to the intense selection for high yield of cane and sugar content in the plant breeding programme. The parents of these crosses are part of the elite germplam, therefore they are highly selected cultivars specifically chosen for high sugar content. This means that at the genetic level there is little chance of detecting loci that are associated with low purity and pol % cane in the progeny, because of skewed gene frequencies in favour of high yield as observed in the frequency distribution curves. Years of selecting for high purity and pol % cane may have eroded the genetic factors associated with low extremes of these traits.

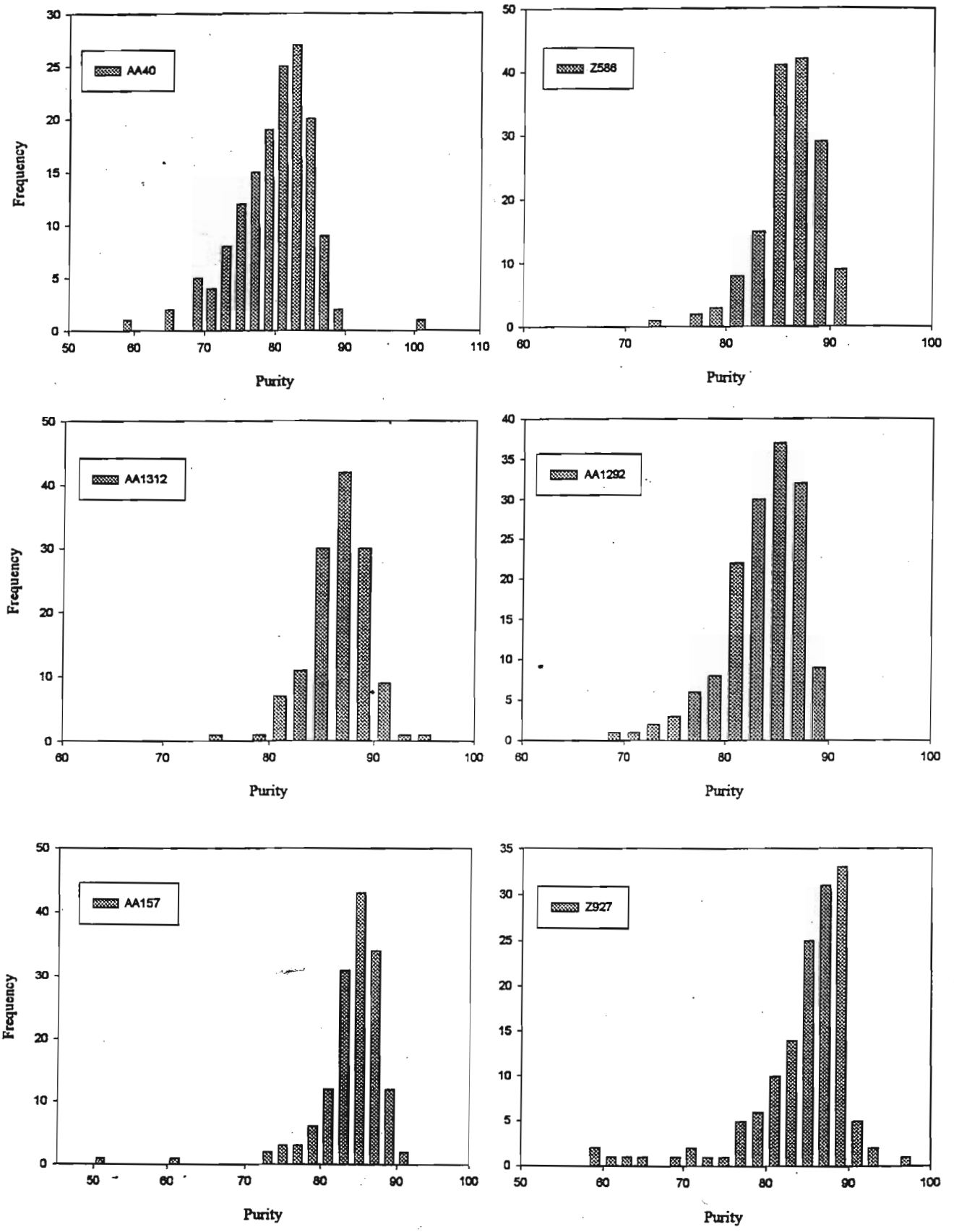


Figure 3.1 The frequency distributions of the six crosses for purity.

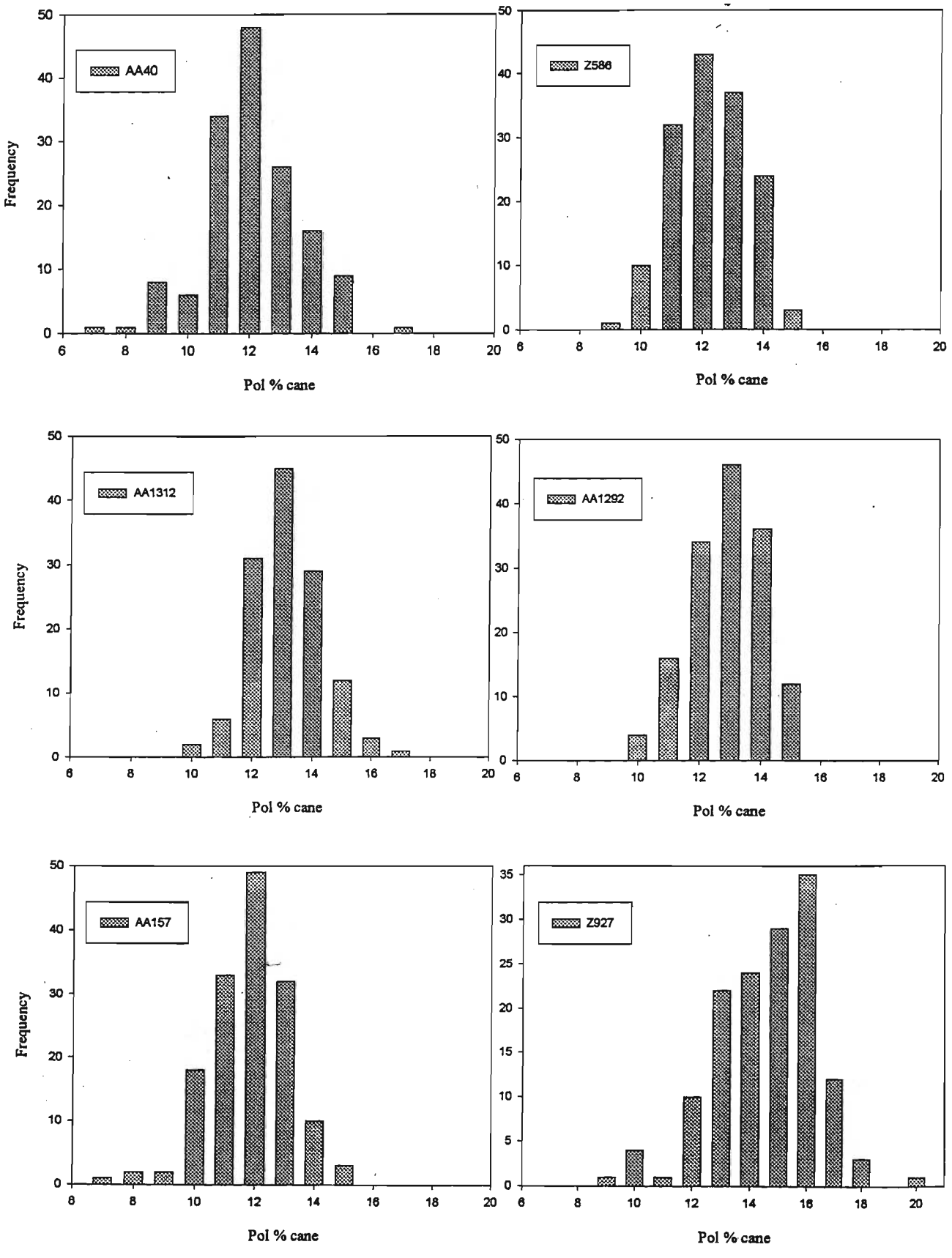


Figure 3.2 The frequency distributions of the six crosses for pol %cane.

Brix %cane showed good normal distributions, except for cross Z927, which was left-skewed (Figure 3.3). The range in values for brix %cane was narrow, implying that there are few genetic factors controlling the extreme phenotype. Even though brix %cane is, by definition, a component of sucrose yield, it is not directly selected for in the breeding programme. Brix %cane has been shown to be positively correlated to pol %cane (Blose, 1992). This means that at the genetic level it may be controlled by the same loci, or tightly linked to the ones controlling pol %cane and purity, hence the skewness and/or limited gene frequencies.

Fibre %cane had the best normal distribution curves of the four traits investigated (Figure 3.4). It is the least selected of the studied traits in the breeding population and the least correlated to any of the other traits. This implies that the gene frequencies for its expression are the least skewed. The well spread out frequency distributions of the six populations with clearly definable extremes confirm this assertion. Thus a relatively large number of the opposite genetic factors for the extreme phenotype are still present. Since high fibre is derived from *S. spontaneum* (Stevenson, 1965), genetic loci controlling its expression may still be present in its remaining genetic complement.

Furthermore, among the four traits it is the least affected by environmental variation (Blose, 1992). The implication of this observation is that genetic loci (QTLs) controlling fibre may be environmentally stable. Environmentally stable QTLs usually ascribe a large percentage of phenotypic variation , which is translatable to high heritability. Based on the results of analysis with frequency distribution curves, past studies on general behaviour of traits in the elite germplasm (Anon., 1988, Blose, 1992, K. Nuss, pers comm[2]) and the above considerations, fibre %cane was chosen for further analysis to determine the most suitable population.

2 K. Nuss, Head: Plant Breeding, SASEX, Mount Edgecombe, 4300

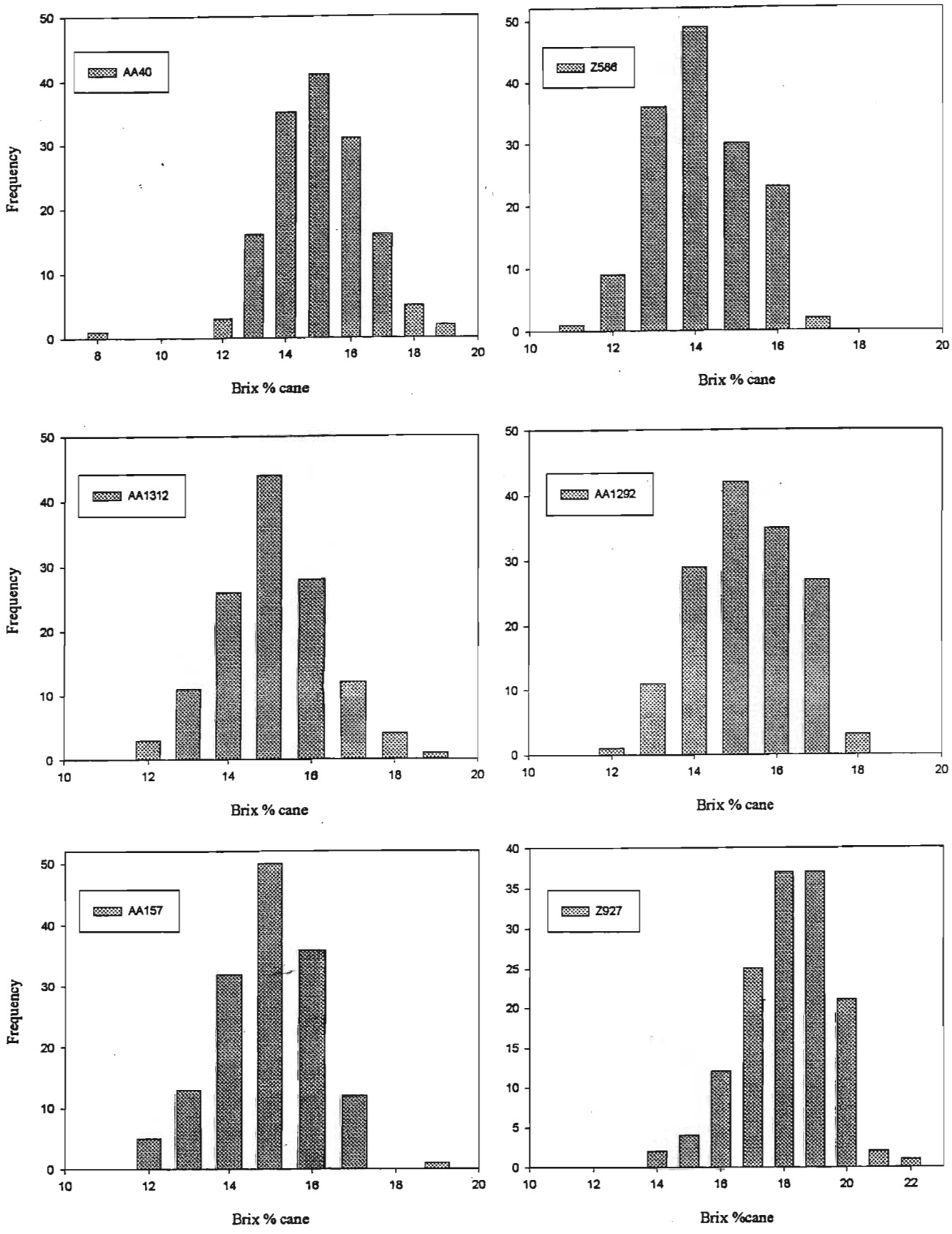


Figure 3.3 The frequency distributions of the six crosses for brix % cane.

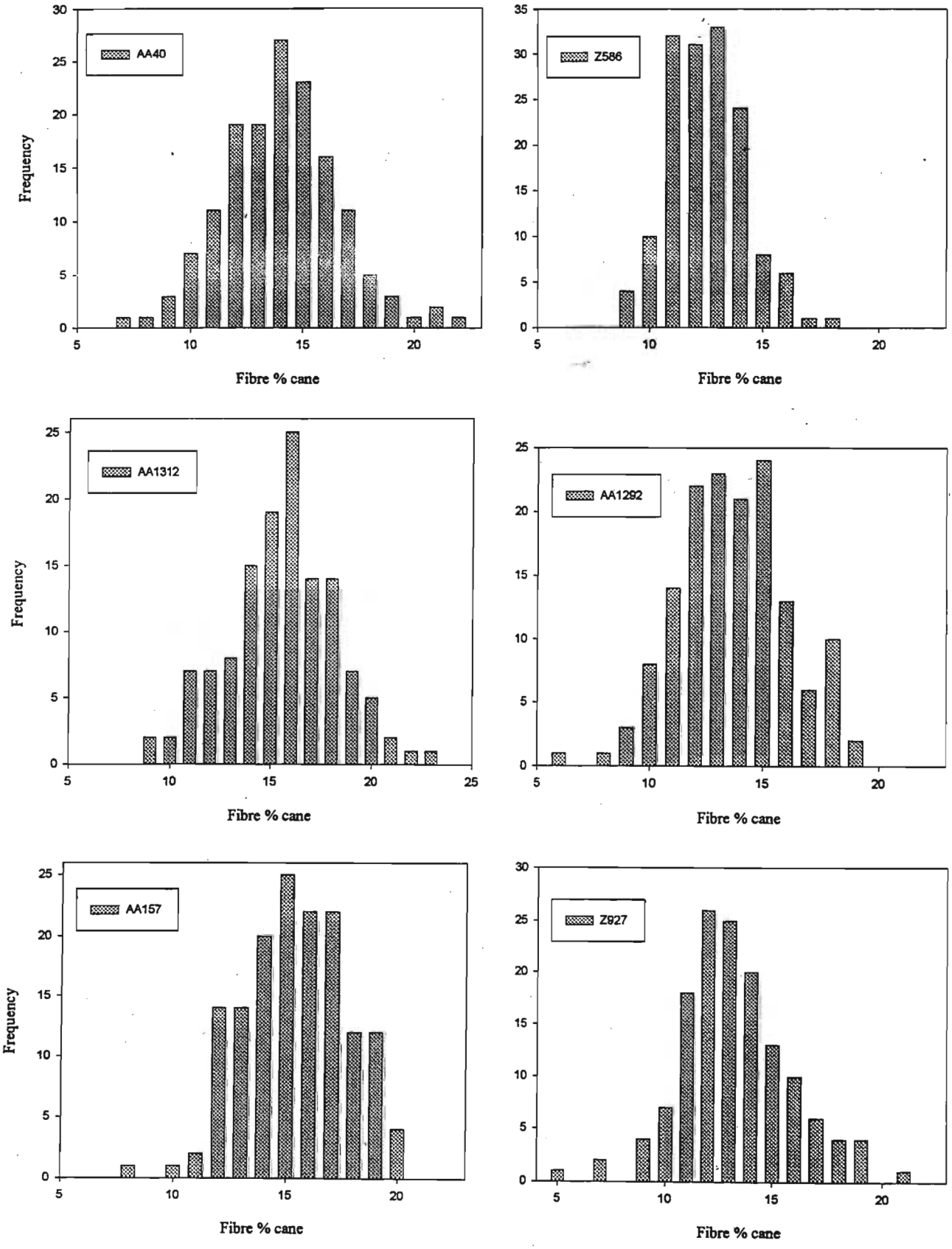


Figure 3.4 The frequency distributions of the six crosses for fibre % cane.

The phenotypic range of the six populations was from 8.8% for Z586 to 15.5% for AA40 (Table 3.1). The standard deviations ranged from 1.7 for Z586 to 2.6 for AA40 and Z927.

Table 3.1 A summary of the primary statistics for fibre of the six populations studied

	Populations					
	AA 40	Z 586	AA 1312	AA 1292	AA 157	Z 927
G Mean	14.5	13.0	16.0	14.1	15.8	13.7
MaxVal	22.5	18.0	23.9	19.4	20.9	21.3
MinVal	7.0	9.2	9.6	6.4	8.9	5.9
Range	15.5	8.8	14.3	13.0	12.0	14.4
SD	2.6	1.7	2.7	2.4	2.3	2.6
CV%	17.6	13.1	17.0	17.1	14.6	18.9

G Mean refers to the grand mean, MaxVal = maximum phenotypic value, MinVal = minimum phenotypic value, SD = standard deviation and CV% is the co-efficient of variation for the measurements.

Population AA40 had the best range indicating a large phenotypic spread, however, it had a large coefficient of variation (17.6). Population AA157 was chosen for the BSA on the basis of its comparatively low cv % value (14.6). Since the coefficient of variation standardises the deviation from the mean it was taken as an indication of the reliability of phenotypic measurement.

Population AA157 was replanted in three replicates in the same area and its phenotype assayed, using the same methods as above, when they were six months and eight months old. The means of the three replicates for both sets of data were calculated, and then used to rank the clones according to their phenotypic content (Tables 3.2 and 3.3).

Table 3.2 Analysis of fibre%cane in the lower fibre group of the bottom 20 individuals over three replicates. The analysis was done two months apart and the combined means of replicates used to rank individuals according to their fibre content.

Plot No.	Ranking (6 months)	Rep.1	Rep.2	Rep.3	Mean	Ranking (8 months)	Rep.1	Rep.2	Rep.3	Mean	Combined Mean	Final Ranking
117	1	7.1	8.7	7.3	7.7	3	7.5	10.1	8.3	8.6	8.2	2
106	2	8.0	7.2	8.5	7.9	1	7.5	7.8	8.5	7.9	7.9	1
21	3	8.5	7.8	7.8	8.1	12	11.2	9.9	9.6	10.2	9.1	10
79	4	8.5	9.3	6.4	8.1	9	10.8	9.6	9.2	9.9	9.0	9
56	5	9.1	8.0	7.2	8.1	2	8.7	9.3	7.5	8.5	8.3	3
51	6	8.9	7.7	9.1	8.6	4	9.1	8.6	8.2	8.6	8.6	4
34	7	8.1	8.9	8.7	8.6	14	10.2	11.8	9.0	10.3	9.5	12
41	8	8.3	8.9	8.6	8.6	6	8.4	9.5	8.5	8.8	8.7	5
108	9	8.6	8.1	9.4	8.7	8	8.3	9.3	9.7	9.1	8.9	7
93	10	8.3	9.3	8.7	8.7	7	9.1	8.7	8.8	8.9	8.8	6
120	11	8.6	11.4	6.3	8.8	17	10.1	11.8	11.7	11.2	10.0	15
15	12	7.7	9.6	9.2	8.8	13	10.4	9.6	10.7	10.2	9.5	11
65	13	10.6	10.1	6.6	9.1	18	11.4	11.8	12.5	11.9	10.5	17
122	14	8.6	10.9	7.9	9.1	5	7.5	10.5	8.3	8.8	9.0	8
147	15	10.8	9.0	7.5	9.1	19	10.0	15.3	11.8	12.4	10.7	18
75	16	9.0	8.2	10.4	9.2	16	10.3	11.9	10.7	11.0	10.1	16
135	17	9.8	9.6	8.4	9.3	15	9.6	11.6	10.1	10.4	9.9	14
96	18	10.1	8.6	9.4	9.4	10	8.8	11.7	9.3	9.9	9.7	19
63	19	10.1	8.7	9.4	9.4	20	18.1	11.0	10.5	13.2	11.3	20
130	20	9.4	10.3	8.7	9.5	11	9.9	10.4	9.5	9.9	9.7	13

Rep = replicate

Table 3.3 Analysis of fibre%cane in the high fibre group of the top 20 individuals over three replicates. The analysis was done two months apart and the combined means of replicates used to rank individuals according to their fibre content.

Plot No.	Ranking (6 months)	Rep.1	Rep.2	Rep.3	Mean	Ranking (8 months)	Rep.1	Rep.2	Rep.3	Mean	Combined Mean	Final Ranking
119	1	14.3	16.1	15.5	15.3	2	14.6	16.9	15.3	15.6	15.5	2
140	2	17.2	13.8	14.5	15.2	1	16.8	14.7	16.8	16.1	15.6	1
94	3	15.2	14.9	14.2	14.7	3	14.6	14.4	15.4	14.8	14.8	3
57	4	15.2	13.3	14.8	14.4	5	14.4	14.2	14.1	14.2	14.3	4
64	5	14.8	13.3	15.0	14.4	9	13.6	15.1	12.8	13.8	14.1	5
111	6	13.0	16.4	12.0	13.8	12	14.1	12.5	14.2	13.6	13.7	8
10	7	14.2	14.2	12.7	13.7	18	11.6	12.9	12.6	12.4	13.0	15
6	8	13.0	13.6	14.2	13.6	8	11.4	15.0	15.7	14.0	13.8	7
126	9	13.3	13.4	13.8	13.5	14	11.8	13.8	13.8	13.1	13.3	13
129	10	14.9	11.7	13.8	13.5	4	14.6	14.5	14.2	14.4	14.0	6
133	11	14.4	12.8	12.4	13.2	16	11.9	13.6	12.3	12.6	12.9	16
50	12	12.8	14.3	12.2	13.1	6	13.4	14.8	14.3	14.2	13.6	9
61	13	12.8	13.5	12.8	13.0	10	12.4	14.8	14.3	13.8	13.4	11
62	14	12.0	14.7	12.3	13.0	19	12.6	11.8	12.4	12.3	12.6	18
70	15	13.1	11.8	13.9	12.9	13	13.9	13.2	13.3	13.5	13.2	14
84	16	13.5	11.0	14.1	12.9	17	11.3	14.3	11.8	12.5	12.7	19
74	17	13.7	12.7	12.3	12.9	7	12.6	15.9	13.9	14.1	13.5	10
107	18	13.9	10.3	14.2	12.8	11	13.3	13.2	15.0	13.8	13.3	12
2	19	12.3	14.1	11.8	12.7	20	10.7	12.9	11.9	11.8	12.3	20
76	20	14.0	13.0	11.1	12.7	15	14.0	12.3	12.2	12.8	12.8	17

Rep = replicate

The final rankings were inferred by calculating the combined means of the two sets of data for each clone. In general, the rankings of the top and bottom 20 individuals were stable, and did not move out of this group during the two separate phenotypic assays. In the low fibre group the order of ranking within the lowest 10 individuals was stable between the two phenotypic measurements.

Only clone 122 originally ranked 14 at six months, showed a significant shift in its ranking to number five at eight months. This was as a result of the reduction in its measured fibre content in replicates one and two, which was also observed in replicates of other clones e.g. replicates one and three of clone 51. However, most clones showed an increase in fibre content from six months to eight months e.g. clones 117 and 106. This may be due to either sampling error or development/physiological processes under genetic control, which reduce fibre content.

The rankings in the high fibre extreme were also stable in terms of movement in or out of the group except for two clones, numbers 10 and 126, which dropped out of the group of 10. The trends with regard to the means after two months were such that some of the clones had a decrease in fibre content. These were clones 10, 57, 64, and 111. However, these reductions in the fibre content were marginal.

On a group basis the top and bottom five individuals were more stable than the corresponding groups of 10 individuals. Sampling of the individuals for the construction of DNA bulks was therefore based on both the extreme five and 10 individuals. The rationale for the two sets of bulks was that the appropriate genetic window could not be readily ascertained *a priori*, since bulking was based solely on the phenotype.

3.3.2 DNA concentration and quality

The spectrophotometric measurements of the DNA concentration were generally accurate and the 260/280 ratios ranged between 1.6 and 1.82 indicating that there was little contamination in the samples.

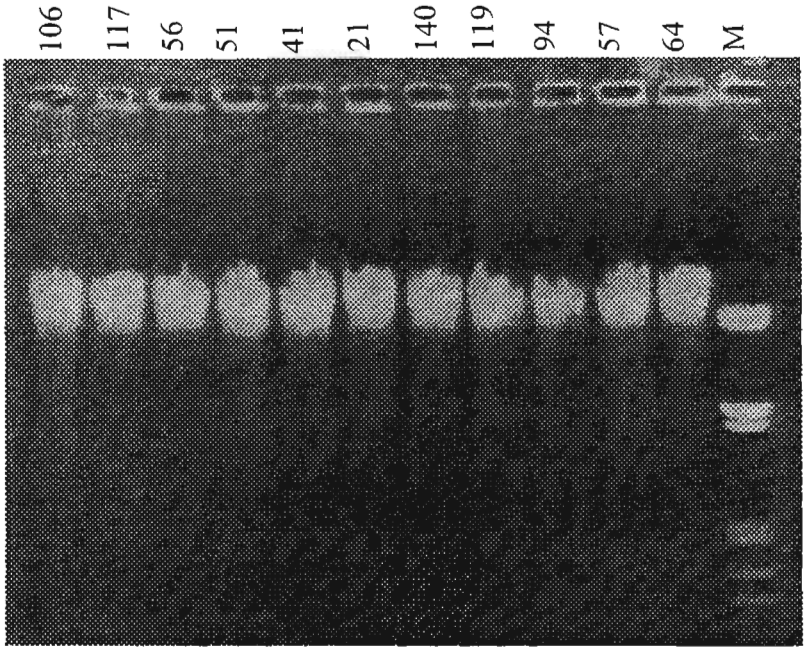


Figure 3.5 A 1% agarose gel for confirmation of DNA concentrations. A total of 500ng was loaded in each lane, according to the spectrophotometric analysis. Lanes 1to 11contain the DNA of different progeny of cross AA157, designated as clone number, and lane 12 is the EcoRI/HindIII Lambda DNA marker.

In addition, the results of the 1 % agarose gel electrophoretic analysis confirmed that the DNA was of high quality (not degraded), and that there was no RNA contamination (Figure 3.5). The results also indicate that the concentration of the DNA samples as determined by spectrophotometric readings were fairly accurate.

3.3.3 RAPD analysis

3.3.3.1 Scoring of the RAPD fragments

RAPD fragments were visually scored on the agarose gels under UV light, and on photographs. There were no discrepancies either in the number of bands or bulk specific polymorphisms between the two methods. Figure 3.6 is an example of RAPD amplification for two primers, OPB1 and OPB16, which yielded bulk specific polymorphisms.

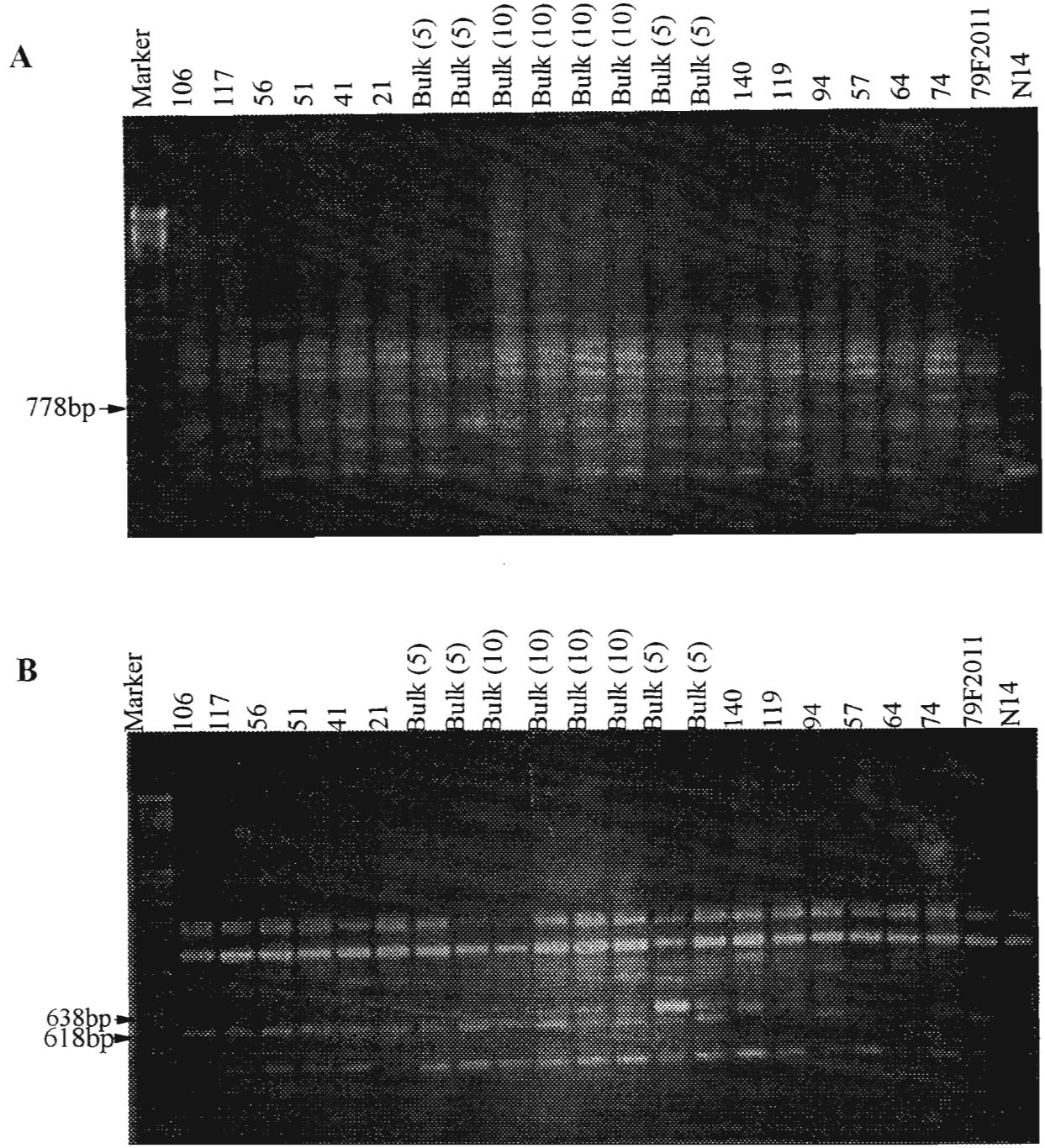


Figure 3.6 The RAPD profiles of the progenies of cross AA157and their bulks, amplified with two primers, OPB1(A) and OPB16 (B). Lane 1 is the Lambda DNA-EcoRI/HindIII marker, lanes 2-6 are five progenies of cross AA157, with the lowest fibre content, lane 7 is clone 21 ranked 10th lowest in fibre content, and lanes 8-11 are the bulks of low fibre (number of individuals is indicated in brackets). Lanes 12-15 are the bulks of high fibre individuals, lanes 16-20 are five progenies of cross AA157 with the highest fibre content, lane 21 is clone 74 ranked 10th highest in fibre content and lanes 22 and 23 are the parents of cross AA157, 79F2011and N14. **A** represents a bulk specific polymorphism ($OPB1_{778}$) which is absent in all low fibre bulks, despite a contaminating individual (clone 21) in the bulk of ten. **B** represents a bulk specific polymorphism ($OPB16_{618}$), which is scored on the basis of the segregation of progeny, since it is present in all bulks because of a contaminating individual (clone 94) in both high fibre bulks.

3.3.3.2 RAPD marker output

A total of 749 loci were amplified with the 60 primers in the bulks. This indicates that on average the primers amplified 12 loci (Table 3.4). There were 79 polymorphisms between the parents, N14 and 79F2011, which represents about 10.5% genetic variation of the genome sampled by the random primers.

Table 3.4 Summary of the PCR-RAPD amplification output with random decamer primers

Primers used	60
Loci amplified	749
Average loci per primer	12
Parental polymorphisms	79
Detected genetic variability between parents	10.5%

3.3.4 Detection of bulk specific polymorphisms

Eight polymorphic fragments were detected between the four sets of bulks. Four of each were specific to low and high fibre. Six of these fragments originated from the female parent 79F2011, while the other two came from the male parent, N14.

Two criteria were used to declare polymorphisms as bulk specific. Firstly, fragments were scored solely on the basis of presence or absence in the bulks, for example $OPA17_{438}$ present in both low fibre bulks and absent in the two high fibre bulks. Six of the fragments were scored in this manner (Table 3.5). The second criterion was that fragments were scored on the basis of their presence or absence in the individuals used to construct the bulks, the criterion being that one aberrant individual could be tolerated as in the case of $OPB16_{618}$.

Table 3.5 The distribution of bulk specific polymorphisms in the bulks, five individuals from which the bulks were constructed and the parents of the cross AA157.

Clone	OPA17 438bp	OPB1 778bp	OPB9 1832bp	OPB11 464bp	OPB16 618bp	OPC4 321bp	OPC11 352bp	OPC16 889bp
				Low Fibre				
106	+	-	+	+	+	-	+	-
117	+	-	+	+	+	-	+	-
56	+	-	+	+	+	-	+	-
51	+	-	+	-	+	-	+	-
41	+	+	+	+	+	-	+	+
Bulk5	+	-	+	+	+	-	+	-
Bulk10	+	-	+	+	+	+	+	+
				High Fibre				
140	-	+	-	-	-	+	-	-
119	-	-	-	+	-	+	-	+
94	-	+	-	-	+	-	+	+
57	-	+	-	-	-	+	-	+
64	-	-	+	-	-	+	+	+
Bulk 5	-	+	-	+	+	+	-	+
Bulk10	-	+	+	-	+	+	-	+
N14	-	+	-	+	-	-	-	-
79F2011	+	-	+	-	+	+	+	+

The second criterion for declaring bulk specific polymorphism was based on the observation that in some cases the presence of one aberrant individual resulted in the loss of polymorphism in the bulks. This approach has the effect of reducing the bulk, and may therefore increase the likelihood of detecting false positives. However, since the putative markers are always validated for their segregation in the progeny of the population for linkage to the trait, false positives can be identified.

The amplification of the rare allele in the bulk is either a function of the dilution factor or loci specific since it is not possible to predict whether the contaminating loci will/will not amplify in the bulk (Table 3.5). The results of the amplification with primers OPB1, OPB9 and OPC16 in the null bulks of five indicate that the amplification of the rare allele is dependent on competition with other unrelated loci.

In the case of primer OPB1, there are nine other loci competing with the bulk specific locus. This means that the locus constitutes 10% of loci, per individual, available for primer binding. The contamination levels of the locus in a bulk of five, therefore amounts to 2.2% of the bulk. Contamination levels can be illustrated by a simple equation:

$$\text{Dilution factor} = \frac{A}{N\,(T-A) + A} \times 100\% \qquad \text{(Appendix C)}$$

where A is the number of contaminating bands, N is the total number of individuals in the bulks, and T is the total number of loci amplified. The equation is based on the assumption that the amplified loci are in equal dosage per individual. and therefore represents a maximum for possible contamination levels.

For both primers OPB9 and OPC16, each amplified a total of 13 loci, indicating that aberrant loci constitute 7.7% of loci available for primer binding. As a function of their contribution in the bulk of five individuals, this translates to approximately 1.63%. Non-amplification with primer OPC11 in the high fibre bulk of five is a further confirmation of this observation, where two individuals were aberrant and a total of 18 bands amplified. In this case contamination levels were calculated to be 2.2% which is similar to that of the other primers.

The RAPD amplification of the rare allele in the null bulk of five with primers OPB11 and OPB16 (contamination levels ~ 1.4% and 1.8% respectively) indicates that the amplification of the rare allele is also locus specific. In both cases the dilution factor did not prevent non-amplification in the bulks.

3.3.5 RAPD marker segregation and linkage analyses

The segregation of the eight bulk specific polymorphisms was analysed in 80 randomly selected progeny of the population AA157. Seven of these reproducibly segregated as single dose fragments (1:1 ratio) at 95% probability level (Table 3.6).

Table 3.6 The segregation of the seven bulk specific RAPD fragments in 80 progeny of the cross AA157.

RAPD	Present	Absent	χ^2 (1:1 ratio)*	Parental origin
OPA 17_{438}	39	41	0.05	79F2011
OPB 1_{778}	34	44	1.28	N14
OPB 11_{464}	45	32	2.19	N14
OPB 16_{618}	33	36	0.13	79F2011
OPC 4_{321}	31	41	1.39	79F2011
OPC 11_{352}	43	30	2.31	79F2011
OPC 16_{889}	37	41	0.205	79F2011

The χ^2 values are given at 95% probability level and 1 degree of freedom (3.86 is the threshold value)

In addition, 15 other parental polymorphisms segregated in a single dose fashion. This indicates that at least 27.8% of the parental polymorphisms in this cross are single dose, or 2.93% of the total loci amplified. The other polymorphic fragments either segregated with a 3:1 or 4:0 ratio, indicating multiple dosage. According to Wu *et al.* (1992), the fragments that exhibit a 3:1 segregation ratio are most probably double dose fragments, and depending on the ploidy type of the species (allo or auto), the ones that segregate 4:0 can either be double or multiple dose.

Segregation ratios for three fragments, $OPB13_{500}$, $OPC7_{940}$, and $OPC17_{1822}$ could not be classified under any of the three ratios of 1:1, 3:1 or 4:0. In fact, they exhibited a ratio that is close to 1 (presence): 2 (absence). Their chi-square values were 0.155, 0,156, and 0.306 respectively which are significant for the 1:2 segregation ratio at 95% probability level. This may be due to segregation distortion as sugarcane is an aneuploid. Segregation distortion as a result of the RAPD technique has also been reported (Reiter *et al.*, 1992, Echt *et al.*, 1992).

Each parent contributed 50% of the total number of single dose fragments. Linkage analyses were done with Mapmaker 2.0 (Lander *et al.*, 1987) using a minimum LOD score of 3.0 to detect linkage between the single dose fragments of either parent. Only two RAPD fragments were linked. These were $OPA15_{667}$ and $OPB16_{618}$ with a genetic distance of 9.1 Haldane centiMorgans, and a LOD score of 25.7. These two fragments were from the female parent 79F2011. No linkages could be detected with the RAPD fragments from the male parent N14. Repulsion phase linkages were not detected in RAPD markers derived from any of the two parents.

3.3.6 RAPD marker - trait linkage analysis

Using a sample of 80 individuals from the population AA157, the segregation of the eight putative markers was investigated, and this information together with the phenotypic data was used to establish their linkage to the fibre trait. This was done by a single factor analysis of variance using Genstat (1993). It was found that three of the putative markers were significantly linked to the fibre trait, $OPA17_{438}$ and $OPC16_{889}$ at 99% significance level and $OPB11_{464}$ at 95% significance level (Table 3.7). The three linked markers were then subjected to regression analyses to establish the percentage fibre variation they ascribed in this population.

Table 3.7 A summary of the results for the analysis of variance (the F probabilities) and the regression analysis.

RAPD Marker	Linkage	F Prob	%Fibre variation ascribed
OPA 17_{438}	YES	0.001**	14.6
OPB 1_{778}	NO	0.281	0.3
OPB 11_{464}	YES	0.046*	4.8
OPB 16_{618}	NO	0.639	-
OPC 4_{321}	NO	0.080	3.4
OPC 11_{352}	NO	0.93	-
OPC 16_{889}	YES	0.009**	9.2

** indicates significant at 99% probability level, * indicates significant at 95% probability level

Although markers OPB1_{778} and OPC4_{321} seem to ascribe some fibre variation, their linkage could not be ascertained by the ANOVA. The results also indicate that more than 50% of bulk specific polymorphisms were false positives, as they were not significantly linked to the trait.

To ascertain whether the bulk segregant analysis was more efficient in finding molecular markers than the conventional method where random markers are subjected to multiple regression analyses to find marker-trait associations, all 22 single dose markers were subjected to statistical analyses to determine their linkage to fibre. Only the three previously detected RAPD markers were significantly linked. This implies that BSA was more sensitive and could be utilised for finding markers linked to traits provided they are highly heritable and there is accurate phenotypic data.

3.3.7 Validation of putative RAPD markers linked to fibre

The availability of phenotypic data for fibre from the previous season (1992) and the following season (1994) presented an opportunity to carry out validation of the linkage of these markers on a year to year basis. Furthermore, a *post hoc* analysis on the composition of bulks in any one

season could be carried out. In sugarcane breeding, both the year and time of harvest are important components of environmental variation (Milligan *et al.*, 1990, Blose, 1992).

Both time of harvest and different seasons were variables in the data for the present study. The time of harvest in 1992 was 12 months, eight months in 1993 and nine months in 1994. Furthermore, the 1992 trial was not replicated, while the 1993 and 1994 trial had three replicates each. For this reason the 1992 data was treated with caution as non-replication may increase sampling variation.

Briefly, validation involved two approaches. The first approach involved simulation of bulking using clonal rankings based on trait values, comparing the bottom and top 10 and 20 individuals between the three data sets, using the 1993 data set as the reference point. In addition, the RAPD genotypes of the simulated bulks were compared on a year to year basis. The RAPD genotypes were based on the three putative RAPD markers, $OPA17_{438}$, $OPB11_{464}$, and $OPC16_{889}$. The second approach involved the analysis of variance and multiple regression on the phenotypic data of 1992 and 1994 and comparing it to the 1993 data, the year in which the phenotypic data for the study was based.

3.3.7.1 Investigations of seasonal variation of fibre as a function of clonal composition in the bulks.

The analysis to determine the reliability of bulking according to the rankings based on phenotypic data indicates that there was less correlation in the rankings for the low fibre group compared to the high fibre group. In fact, the low fibre group, for all three sets of data showed less than 50% correlation in the rankings of the first 10 individuals (Table 3.8).

Table 3.8 Comparison of clonal rankings of the individuals in the low and high fibre groups on a year to year basis, using the 1993 phenotypic data as the reference point.

Number of clones	Low Fibre		High Fibre	
	1992	1994	1992	1994
10 clones	30%	40%	60%	50%
20 clones	45%	50%	55%	55%

When the analysis was extended to include 20 individuals, which is 25% of the sample of 80 individuals, the results indicated a slight increase in correlation, for the 1992/1993 and 1994/1993 sets of data in the low fibre group. The correlation value for 1994 phenotypic data is higher than the 1994 season for the low fibre group. This may be due to the replication of the trials. Alternatively, it could be due to the cane in 1993 and 1994 being approximately the same age (eight and nine months respectively).

The high fibre group was generally better than the low fibre one with correlation values as high as 60% for the first 10 individuals in 1992. Translated in terms of bulk composition it means that 60% of individuals that were included in the BSA bulk of 10 would have been the same in 1992. An increase in correlation was observed from the group of 10 to 20 individuals in 1994 (from 50% to 55%). The correlation between phenotypic data sets for 1993 and 1992 decreased from 60% in the 10 individuals to 55% in the 20 individuals.

Clones 24 and 27, ranked in the low fibre group of 20 in 1994 were found in the high fibre group of 20 in the 1992 data set. However, these individuals fall outside the group of 10 that was used for bulking. The data for the middle 50% of all the samples indicate that there is no reliable ranking in this group, individuals change their ranking at random. The shows that since the phenotypic extremes differed from year to year, individuals chosen for bulking in any one year would have been different.

The results indicate that bulk composition for low fibre in any of these seasons would have been different, and this would have the effect of reducing the chances of identifying putative markers. The results for high fibre are better than those of low fibre, which may be an indication of the stability of genetic factors involved in the expression of high fibre.

This raises the question of the reliability of phenotypic assaying, or more importantly, the choice of trait that was used for bulking. Environmental modulation and growth stage factors also need to be considered. The differences in correlation values in the groups are an indication that low fibre is more susceptible to environmental variation.

3.3.7.2 Investigation of simulated bulk genotypes

Simulated bulks were constructed according to the rankings from year to year. The results for low fibre groups (Table 3.9) indicate the genotypic composition of bulks of 10 individuals would have been highly accurate for putative marker $OPC16_{889}$ for all three seasons.

Table 3.9 The % correct RAPD genotypes of the simulated bulks for 10 low and high fibre individuals for the three seasons.

	Low Fibre			High Fibre		
	OPA 17	OPB11	OPC 16	OPA 17	OPB 11	OPC 16
1992	50%	30%	70%	90%	70%	90%
1993	70%	70%	90%	90%	70%	90%
1994	50%	40%	70%	60%	60%	70%

As expected the bulks from 1993 are the most accurate because it is the year in which the phenotypic data for bulk construction was assayed. Compared to the clonal rankings correlations, RAPD markers $OPA17_{438}$ and $OPB11_{464}$ are slightly better. The implication of these results is that putative marker $OPC16_{889}$ may be the most stable in terms of ascertaining fibre variation across seasons.

The result of the simulated high fibre bulks confirm the stability of the putative marker $OPC16_{889}$ across all three seasons. It also shows that putative marker $OPA17_{438}$ is also stable in the three seasons. The general trend is that the stability of the genotypes of the high fibre group is higher than those of low fibre. These findings corroborate the clonal ranking results in that the high fibre group is more stable on a year to year basis than the low fibre group. Furthermore, the results indicate that putative marker $OPC16_{889}$ would have been detected as a bulk specific polymorphism in all three seasons. To investigate this point further an analysis of RAPD marker–fibre linkage was carried out for all three seasons using the same statistical methods as before.

3.3.7.3 Validation of putative RAPD marker–fibre linkages on a year to year basis

The results of the validation of the association of putative markers with the fibre trait (Table 3.10) indicate that putative marker $OPA17_{438}$ was the most stable across the three seasons. The percentage fibre ascribed by this marker in 1993 and 1994 is stable, and higher than that of $OPC16_{889}$ which was not significantly linked to the fibre trait in the 1994 season data.

Table 3.10 A summary of the fibre % variation ascribed by the bulk specific RAPD markers in the three seasons as a measure of R^2 in the multiple regression analysis.

Putative marker	1992[a]	1993	1994
$OPA17_{438}$	0.9	14.6	12.2
$OPB1_{778}$	-	-	17.2
$OPB11_{464}$	-	4.8	-
$OPB16_{618}$	-	-	-
$OPC4_{321}$	1.8	-	2.0
$OPC11_{352}$	-	-	-
$OPC16_{889}$	2.8	9.2	-

a indicates that the data was not replicated.

The percentage fibre ascribed by putative marker $OPB11_{464}$ was higher than that of any other marker ($OPA17_{438}$ and $OPC4_{321}$) in the 1992 season, however, the total phenotypic variation ascribed by the markers was generally low in this season.

3.3.7.4 Cross validation of the RAPD markers by sample partitioning

Obviously, the most efficient way of cross validating marker-trait linkages would be test the set of markers detected in one population in another population (Bridges and Sobral, 1996). If that option is not feasible due to practical limitations, as in this case, the best alternative is to use the method of 'jacknifing' where the results of the regression on a sample of (n –1) is used to predict the trait value of the individual left out.

In this study cross validation was carried out by a modified form of jacknifing. The sample size was reduced by 10 individuals taken at random using the random number generator prompt of Quattro Pro. The (N – 10) sample thus produced was subjected to the same regression analyses as the N sample from which it was derived. This enabled the comparison of 1) R^2 values and 2) to ascertain which markers were retained.

Normality tests were carried out to test if the data was not biased by sub-sampling and the results showed no significant deviation from the N sample (data not shown). Chi-squared tests were also performed to test if the segregation ratios of the putative markers still fitted the 1:1 ratio. All seven markers segregated as before. Finally, t-tests to compare if the means of the (N – 10) sample were significantly different from that of the N sample were performed. None of the means were significantly different. These results indicate that the basic statistics of the samples had not changed. Once these tests had been performed, the regression analyses were performed on the (N – 10) sample as before.

The summary of the results (Table 3.11) indicate that the RAPD markers $OPA17_{438}$ and $OPC16_{889}$ are still putatively linked to the fibre trait. With respect to marker $OPA17_{438}$ the putative linkage to fibre in 1992 has been lost. It is, however, worth pointing out that the putative linkages in 1992 were not strong (0.9%). The general trend is that the putative linkages between the markers and traits are maintained on the seasonal basis. Furthermore, the preliminary predictions that were apparent with the simulated bulks were confirmed in that the most stable markers, at least in this population, were $OPA17_{438}$ and $OPC16_{889}$.

Table 3.11 The calculated R^2 values for different seasons using only a subset of the progeny (53 individuals) with multiple regression analysis.

RAPD marker	1992	1993	1994
$OPC16_{889}$	7.6	6.0	NS
$OPB16_{618}$	NS	6.6	2.4
$OPA17_{438}$	NS	17.6	3.1
$OPB1_{778}$	NS	NS	3.8

NS indicates not significant

The previous linkages of marker $OPC4_{321}$ to the fibre trait in 1992 and 1994 could not be confirmed. Again, the amount of fibre variation ascribed by this putative marker had been low (1.8 and 2.0% respectively).

The new association of the putative marker $OPB16_{618}$ with the fibre trait in 1992 and 1993 can probably be explained as an anomaly resulting from the reduction in sample size. The putative marker $OPB1_{778}$ is still linked to the fibre trait, albeit greatly reduced. The other anomalous situation is the increase in % fibre variation ascribed by marker $OPC16_{889}$ in 1992.

3.4 Discussion

One of the main objectives of the present study was to find out whether a phenotype influenced by multiple genetic loci, together with the environment, could serve as a selectable marker for the construction of DNA pools which would permit the detection of DNA markers near genes influencing the target phenotype. This problem is compounded by the fact that for a trait controlled by multiple loci, extreme phenotypes are due to different sets of QTLs (sub-populations) plus the environment (Mendell *et al.*, 1991, Wang and Paterson, 1994).

The extreme phenotypes are not necessarily made up of the same allelic sub-populations. This means that for a QTL which has a large number of genes/alleles, it would be difficult to detect markers linked to genetic loci with significant phenotypic effect unless there has been an introgression of a QTL with a large effect in that crop. This is evident in the present study where all bulks, except for the bulk of five for putative marker $OPA17_{438}$, were contaminated. However, the genotyping of individuals comprising the bulks enabled the identification of putative markers. In general, the putative markers that were identified ascribed a small percentage of fibre variation (Table 3.7), which is contrary to the prediction of Wang and Paterson (1994). In their study they concluded that only QTLs with large effects can be detected with pooling strategies such as the present one. This may be an indication of the nature of the trait that was chosen for investigation.

The choice of the fibre phenotype in sugarcane was the most appropriate one. Firstly, in a previous study, Blose (1992) postulated that non-additive genetic effects are not important in determining fibre content. In trials from 1979, 1983, 1985 and 1987 the heritabilities (h^2_a) were all large ranging from 0.48 to 0.81 and their standard errors were low (less than 25%). The

important factor is that population AA157 is part of the same breeding programme that was investigated in the previous study.

Secondly, modern commercial varieties are hybrids of *S. officinarum* and *S. spontaneum*, and presumably high fibre is from *S. spontaneum* (Stevenson, 1965). This may explain why in the investigations of phenotypic stability from season to season as a measure of clonal ranking the high fibre group was significantly more stable than the low fibre group. It may be that some of the QTLs controlling high fibre have relatively large effects (possibly from *S. spontaneum*) and are largely independent of environmental factors. In maize some QTLs affecting yield and related traits have been shown to be largely independent of environmental variation (Stuber *et al.*, 1992, Ajmone-Marsan, 1995).

If high fibre in hybrid sugarcane varieties was contributed by the *S. spontaneum* progenitors, and taking into consideration that they now comprise 5-10% of the genome of modern varieties (D'Hont *et al.*, 1994) one would expect to be able to detect markers that are linked to fibre relatively easier than the other agronomic traits that were derived from the *S. officinarum.*

The detection of a marker linked to a specific locus in a large quantitative trait is also a function of the permutations of loci controlling it. Most of the individuals will carry different permutations of the alleles. The success of identifying a marker linked to a genetic locus is therefore dependent on the size of the effect of that locus, and the degree of linkage. A large effect genetic locus is more likely to be present or absent in most of the individuals on the extremes of the phenotype, provided dosage does not play a major role in the expression of the phenotype.

Evidence of gene silencing and subsequent diploidisation of some of the higher polyploids has been documented in several tribes, such as *Vicieae*, *Trifolieae*, and *Cicereae* (Weeden *et al.*,

1989). More importantly, for sugarcane, it has been shown in the closely related *Triticeae* (Hart, 1983). Gene silencing and diploidisation of polyploids has important implications in the search for molecular markers linked to traits in sugarcane, because it may mean that a bulk specific marker is linked to a genetic locus that is likely to be directly responsible for the expression of the trait. The results of the *post hoc* analysis with the genotypes of simulated bulks of high fibre indicate that RAPD markers, $OPA17_{438}$ and $OPC16_{889}$, are stable, and would have been detected with the 1992 phenotypic data.

With regard to allelic contamination there are two possible explanations. Certainly in NILs and mutant stocks in diploid species allelic contamination is mainly due to recombination between marker locus and the QTL or gene of interest (Williams *et al.*, 1993). The results of the amplification of the rare alleles in the bulks of sugarcane indicate that both the dilution factor and type of genetic locus are important determinants. With regard to the dilution factor, the calculated contamination levels in this study represent a maximum i.e. the highest possible % contamination levels.

Hallden *et al.* (1996) using 1:1 DNA mixtures of homozygotes in *Brassica napus*, to create artificial heterozygotes found that an expected band failed to amplify in 14% of all cases investigated. In sugarcane it is postulated that there are between eight and twelve homologues for each chromosome (Sreenivasan *et al.*, 1987, D'Hont *et al.*, 1995, D'Hont *et al.*, 1996), so that the competing alleles are potentially the number of homologues multiplied by the total number of RAPD bands. Since all the bulk specific polymorphisms were found to segregate in a single dose fashion, this means that the rare allele was competing with a large number of higher dosage loci for the primer.

Non-amplification of the rare allele in the bulks with primers OPB9, OPB11, and OPC16 are in agreement with previous results (Michelmore *et al.*, 1991, Wang and Paterson, 1994). According to Michelmore *et al.* (1991) the minimum dilution that can be detected of a rare allele is 4% in the bulk. Wang and Paterson (1994), in their *post hoc* analysis using data from previous mapping experiments, hypothesised that contaminating alleles contributing 5% or more of the bulk could be detected.

The level of false positives detected in each one of the three seasons was 50%, which is higher that the 30% detected in peach (Warburton *et al.*, 1996). The higher level of false positives in the study may be due to one of several reasons. The relaxed stringency applied in the declaration of bulk specific polymorphisms, may be one of the factors. However, since only two putative markers ($OPB16_{618}$ and $OPC11_{889}$) were not linked in any of the seasons, the more plausible explanations may be that some of the markers are linked to QTLs that are specific to environments (e.g. age of harvest, year).

Even though the alpha level was set at 0.01 (99% probability level), the use of the single-marker method for declaring associations of QTLs with putative markers may have increased the probability of the type 1 error i.e. declaration of significant associations between marker and trait loci when they are not linked (Lander and Botstein, 1989). However, since cross validation was performed with markers tested in different environments i.e. year and age of harvest, the associations of markers $OPA17_{438}$ and $OPC16_{889}$ are more likely to be real. Another consequence of using the single-marker method is that the effect of the trait loci may have been underestimated if the markers are distant from the genetic loci (Lander and Botstein, 1989).

The validation methods were able to independently show that the two putative markers $OPA17_{438}$ and $OPC16_{889}$ were probably linked to the genetic loci controlling the fibre trait, and

are not spurious linkages. Their retention across different seasons attests to their validity. The stability of putative marker $OPC16_{889}$ was further amplified by the fact that when tested against the mean of the combined data across all three seasons it showed significant linkage to the fibre trait with an R^2 of 0.0846 translating to 8.46% of fibre variation ascribed.

Chapter 4

The behaviour of RAPD based markers and their conversion to sequence characterised amplified regions (SCARs) in sugarcane.

4.1 Introduction

RAPD markers have several limitations, which may prevent their utilisation in genetic analyses, and marker assisted selection (MAS). Despite its ease of use the RAPD technique has been associated with difficulties in reproducibility (Penner, 1993a, Jones *et al.*, 1997). Different DNA extraction methods may result in different RAPD profiles (Talbert *et al.*, 1996). Using DNA preparation methods of Dellaporta *et al.* (1983) and Lassner *et al.* (1989), different amplification profiles were observed with the J15 RAPD primer (Talbert *et al.*, 1996).

Non-reproducibility of RAPD marker profiles among different laboratories has also been reported by Penner (1993a). Among the PCR variables implicated in this variability was the type of thermocycler used for RAPD analysis (Hallden *et al.*, 1996). The possible solutions suggested for this problem is that replicate runs should be done and discarding of all non-reproducible bands (Hu and Quiros, 1991, Reiter *et al.*, 1992). Alternatively all bands can be used and a certain level of error accepted when interpreting the results (Stiles *et al.*, 1993). The use of replicates is especially crucial in distinguishing PCR artefacts (Lamboy, 1994a,b). Another type of PCR artefact that has been observed in RAPDs is the formation of

heteroduplexes which are defined as "hybrid" double-stranded DNA that are formed following the PCR amplification of two DNA segments that have a high degree of homology to each other (Novy and Vorsa, 1996). These loci are detected in the segregating progeny but not in the parents (Ayliffe *et al.*, 1994, Novy *et al.*, 1994, Davis *et al.*, 1995).

The short RAPD primer is also susceptible to point mutations and mismatches, which may affect the reproducibility of products across species and varieties. Miklas *et al.* (1993) reported linkage of a RAPD marker to the Up_2 gene that could not be amplified across crosses of common bean originating from the Andean pool. The value of the marker was restricted to the Middle American gene pool.

RAPD markers are dominant by nature (Tingey and del Tufo, 1993) i.e. they detect only presence or absence of loci. This means that they are unable to discriminate between length variants of the same locus, and therefore dosage and copy number. In some crop plants it has been shown that the majority of RAPD amplification products are highly repetitive e.g. wheat (Devos and Gale, 1992).

Paran and Michelmore (1993) have shown that converting the RAPD markers to sequence characterised amplified regions (SCARs) may circumvent some of these problems in diploid lettuce. By sequencing the RAPD markers and designing longer specific primers they were able to distinguish length polymophisms and resolve dosage of the marker loci.

In sugarcane the problem of dosage resolution is compounded by the high ploidiness, with the basic chromosome number of modern varieties estimated to be between 8x and 12x

(Sreenivasan *et al.*, 1987, D'Hont *et al.*, 1994, D'Hont *et al.*, 1996). Furthermore modern varieties are interspecific hybrids with the genetic components of the original progenitors, mainly *S. officinarum* and *S. spontaneum*, behaving as separate entities, with limited recombination (~10%) between the chromosomes of the base species (D'Hont *et al.*, 1996, Grivet *et al.*, 1996). Since SCARs are more specific because their primers are longer there is a possibility that they may be cross utilised among species and varieties, and they may also be able to distinguish length variants even in a complex polyploid like sugarcane.

The present study was therefore an analysis of the behaviour of SCARs in a complex polyploid. Two RAPD markers, $OPA17_{438}$ and $OPB11_{464}$, which were known to segregate as single dose fragments were used for the analysis, and for comparison a SCAR marker, $SA11_{641}$ (S. Groenewald, pers comm)[1], derived from sugarcane cDNA as a stem preferential expressed sequence tag, was also included. A stably inherited RAPD marker, $OPC16_{889}$, linked to fibre %cane, and also segregating in a single dose fashion in this cross (Chapter 3) was also included in the analysis for comparative purposes.

4.2 Materials and Methods

4.2.1 Cloning of polymorphic fragments

RAPD PCR was carried out as previously described (section 3.2.5). The amplification products were analysed on 2% (m/v) agarose gels, stained with 1ug/ml ethidium bromide and visualised under UV light. RAPD polymorphic fragments, $OPA17_{438}$, $OPB11_{464}$, and $OPC16_{889}$ (see Chapter 3 for details) were excised from the agarose gel and re-amplified with the same RAPD primer. The PCR products were analysed on 2% (m/v) agarose gels, stained,

[1] S. Groenewald, Biotechnology Department, SASEX, Mount Edgecombe,4300

excised, and purified using a Qiagen gel extraction column (Anon., 1995a). The RAPD fragments were then cloned into the blunt ended *Srf I* restriction site of the pCR Script phagemid (Stratagene) according to the manufacturer's instructions (Anon., 1993). Fragment insertion was confirmed with PCR, directly on the white colonies, using the T7 and M13 reverse vector primers.

The PCR conditions were as follows: one unit of Taq polymerase, 0.22μM of each primer (T7 and M13 reverse), 0.1mM of each dNTP (dATP, dCTP, dGTP, and dTTP), 1.5mM $MgCl_2$, 50mM KCl, 10mM Tris-HCl (pH 8.3), 2.1μg gelatine, and 21ng of template DNA. The volume for each reaction was 21μl. The cycling parameters were as follows: 94°C for 2 minutes; then 94°C/1 minute, 50°C/1 minute and 72°C/1minute and 30 seconds, for 10 cycles; another 30 cycles of 94°C/30 seconds, 45°C/30 seconds, and 72°C/1minute. Finally an extension cycle of two minutes at 72°C was performed. The results were analysed on 2% (m/v) agarose gels.

4.2.2 Isolation of plasmid DNA

The positive clones were inoculated into 250ml LB-broth [1.0% (m/v) bacto-tryptone, 0.5% (m/v) yeast extract, 0.5% (m/v) NaCl, and 0.1% (m/v) glucose], and incubated overnight at 37°C with shaking at 225 revolutions per minute. The isolation of plasmid was carried out using the Nucleobond plasmid purification kit according to the manufacturer's instructions (Anon., 1995b).

4.2.3 Sequencing of polymorphic fragments

The plasmids with the polymorphic RAPD fragments, $OPA17_{438}$ and $OPB11_{464}$ were

sequenced. Dye terminator cycle sequencing of the clones was carried out using primers T7 and KS for the 5' end, and T3 and M13 reverse primer for the 3' end. The PCR conditions were as follows: one unit of Taq polymerase, 0.8μM primer, FS ready reaction mixture (Perkin Elmer), and 250ng of template DNA in a total volume of 15μl. The reactions were carried out in PCR tubes. Each reaction mixture was overlaid with 50μl mineral oil. Thermal cycling was done in a Hybaid OmniGene thermal cycler, using the following thermal profile 95°C/1minute; followed by 30 cycles of 95°C/30 seconds, 40°C/30 seconds and 60°C/3 minutes with a $2°Cs^{-1}$ ramp.

The extension products (15μl) were transferred from the PCR tubes to 1.5ml microcentrifuge tubes. Purification of the extension products was done by adding 55.5μl of 70% ethanol with 0.5mM $MgCl_2$, and vortexed briefly. The samples were left at room temperature for 15 minutes for precipitation, and then centrifuged in a benchtop microfuge at room temperature for 15 minutes. The supernatant was aspirated with a micropipette and discarded. The pellet was dried in a vacuum centrifuge for five minutes, followed by resuspension in 18μl template suppression buffer (Perkin Elmer). The extension products were denatured for 2 minutes in a boiling water bath, and immediately quenched on ice for one minute. The extension products were spun briefly in a benchtop microfuge, and then analysed in an automated sequencer (Perkin Elmer ABI 310 Genetic Analyser).

4.2.4 Design of specific primers

A pair of 21-mer oligonucleotide primers was designed according to the consensus sequences on the 5' and 3' ends of each cloned RAPD marker fragments.

4.2.5 Analysis of SCAR markers

Specific PCR was carried out on the parents of the AA157 cross, selected progeny, randomly selected commercial varieties and the bulks of *Saccharum officinarum* and *Saccharum spontaneum*. The bulk of *S. officinarum* comprised Black Cheribon, Black Tanna, and Badilla clones. The bulk of *S. spontaneum* was made up of Pasorean, Nigeria, Coimbatore, Kloet, Tab, and Mauritius clones.

The PCR conditions were as described previously (section 4.2.1). The cycling parameters were: 95°C/30 seconds, 45°C/30 seconds and 72°C/45 seconds for 10 cycles; another 30 cycles with the annealing temperature changed from 45°C to 40°C; and finally a single extension cycle of five minutes. RAPD PCR of the same individuals was carried out as previously (section 3.2.5).

4.2.6 Southern hybridisation

The RAPD and SCAR PCR results were analysed on 2% (m/V) agarose gel. The DNA was then transferred from the agarose gels to positively charged nylon membranes (Hybond N+, Amersham) using 0.4M NaOH, in an upward capillary manner, as described in Ausubel *et al.* (1995). The DNA was fixed to membranes by UV cross-linking at 2500kJ for two and a half minutes on both sides of the membrane.

4.2.6.1 Labelling of DNA probes

The incorporation of the DIG label was done using the PCR amplification of the whole plasmid, with the M13 reverse and T7 primers. The PCR conditions were as follows: one unit

TAQ polymerase, 1.75mM $MgCl_2$, 50mM Tris-HCl (pH 9.2), 16mM $(NH_4)_2 SO_4$, 0.22µM of each primer, 0.2mM of each dNTP (dATP, dCTP, dGTP, 0.13mM dTTP, 0.07mM DIG-11-dUTP, and 250ng of plasmid DNA. The total reaction volume was 15µl. The cycling parameters were: one denaturation cycle of 95°C/7 minutes; followed by 95°C/30 seconds, 45°C/30 seconds and 68°C/45 seconds for 30 cycles, and finally an extension cycle of 68°C for five minutes.

4.2.6.2 Hybridisation and detection

The membranes were pre-hybridised at 42°C for 4 hours using 10ml DIG Easy Hyb buffer (Boehringer Mannheim) per $100cm^2$ of membrane in Hybaid hybridisation bottles. Hybridisation was carried out, with 20ng of DIG-labelled DNA probe, at 42°C overnight.

The membranes were then washed in 2 x SSC (0.3M NaCl, 0.03M Na-citrate, pH 7.0), 0.1% SDS wash solution, twice for five minutes at room temperature. The membranes were then twice washed in 0.5 x wash solution (0.5 X SSC, 0.1% SDS) for 15 minutes each at 68°C.

Chemiluminescent detection with CDP-Star™ substrate was carried out at room temperature according to the manufacturer's instruction (Anon., 1996b). The membranes were wrapped in glad wrap, and exposed to X-ray film for five minutes. The X-ray film was developed for five minutes in the developer (Ilford Phenisol), one minute in 3% acetic acid, followed by 10 minutes in the fixative solution (Ilford Hypam), and washed under running tap water.

4.3 Results

4.3.1 Cloning of polymorphic fragments

The PCR results on the plasmids $NPA17_{438}$, $NPB11_{464}$, and $NPC16_{889}$ were positive i.e. they confirmed the correct fragment insertion. This observation was based on fact that all three fragments were approximately 200 base pairs longer than the original RAPD fragments, since the binding sites for T7 and M13 reverse primers are approximately 100 base pairs from the *SrfI* cloning site (Anon., 1993). On the basis of these results, the $OPA17_{438}$, and $OPB11_{464}$ RAPD fragments were then sequenced.

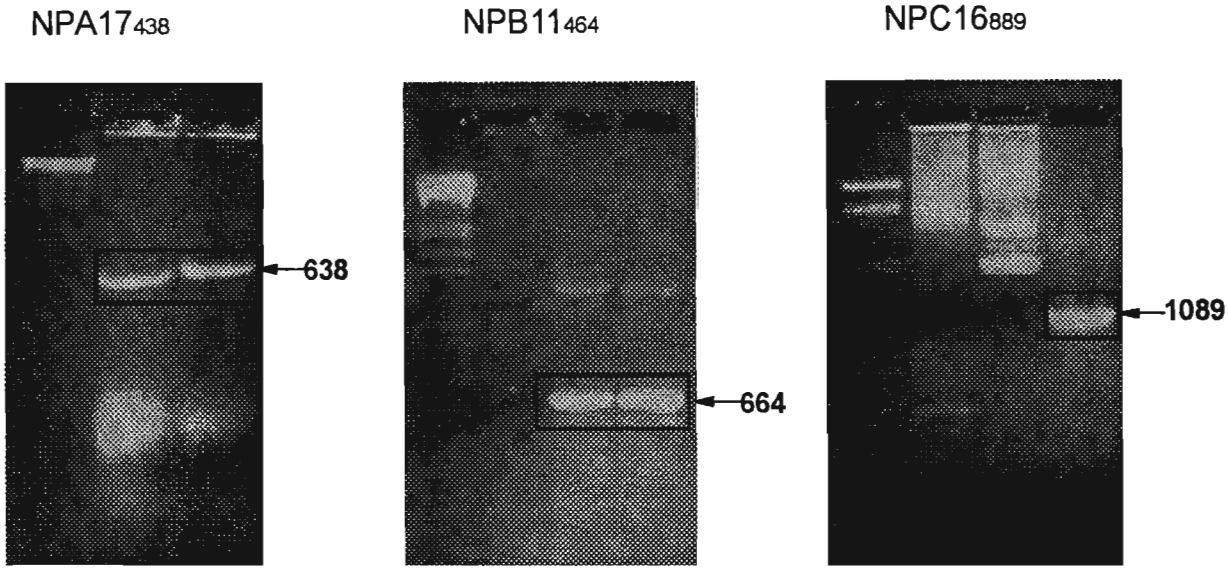


Figure 4.1 Electrophoretic separation on a 2% (m/v) agarose gel of the PCR amplification products from putative transformant clones with the T7 and M13 reverse primers. The boxed clones indicate the successful cloning of the RAPD fragment. Lane number 1, in all three results, represents lambda - DNA EcoRI/ HindIII marker.

4.3.2 Sequencing of the polymorphic RAPD fragments and the design of SCAR primers

The RAPD fragments were not fully sequenced, since only the 3' and 5' ends were needed for primer design. Since each fragment was sequenced with two different primers in the 3' and 5' end, the partial sequences were aligned to form a consensus sequence. For the $OPA17_{438}$ fragment, the alignments of the sequences were fully consensual around the RAPD primer

sequence (GACCGCTTGT) as the results of partial sequences (80 bases) for the forward T7 and KS primers indicate:

NPA17-T7 --5' GCAGCCCGGG GGCTCCGCCC **GACCGCTTGT GGAGGTAGTC**
NPA17-KS --5'GCAGCCCGGG GG*A*TCCGCCC **GACCGCTTGT GGAGGTAGTC**
NPA17-T7 AAGAGATGCT GGAGAAGGCG AAGGCGGCGG CACCATGGAT 3'--
NPA17-KS AAGAGATGCT GGAGAAGGCG AAGGCGGCGG CACCATGGAT 3'--

and the T3 and M13 reverse primers,

NPA17-T3 --5' TGCGGTGGCG GCCGCTCTAG CCC**GACCGCT TGTTATGTAA**
NPA17-M13 --5' TGCGGTGGCG GCCGCTCTAG CCC**GACCG CTTGTTATGTAA**
NPA17-T3 **GAAC**TACGTG A*C*GTAATAAC GATATGTATG TACATGGTGA 3'--
NPA-M13 **GAAC**TACGTG A*T*GTAATAAC GATATGTA TGTACATGGTGA 3'--

For the RAPD fragment $OPA17_{438}$ the SCAR primers for both the 3' and 5' ends were designed by adding 11 base pairs after the original RAPD primer sequence. The percentage GC content was 52% for the 5' end primer (designated as the forward primer) and 38% for the 3' end primer (designated as the reverse primer).

The alignment of the sequences for the $OPB11_{464}$ fragment did not contain the OPB11 sequence (GTAGACCCGT) at the 5' end as the results of partial sequences with T7 and KS primers (forward) indicate:

NPB11-T7 --5' CAGCCCG*GG* GNATCCGCCC GTAGNCCCG*T* **GCATACGAGT**
NPB11-KS --5' CAGCCCG*C*G GNATCCGCCC GTAGACCCG*A* **GCATACGAGT**
NPB11-T7 **TATAATAGCT C**GCGT*C*TTTA TCAATATGGA GTAGGAATC 3'--
NPB11-KS **TATAATAGCT C**GCGT*T*TTTA TCAATATGGA GTAGGAATC 3'--

At the 3' end of the OPB11$_{464}$ fragment, four nucleotides of the original primer sequence were not consensual as the results of partial sequences for T3 and M13 reverse primers indicate:

NP11-T3 --GGTGGCGGCC GCTCTAGCCC GTA**GACCCGT AATACTGTAT**
NPB11-M13 --GGTGGCGGCC GCTCTAGCCC GNN**GACCCGT AATACTGTAT**
NPB11-T3 **TTCC**TAAGCA TATTTCCCTC GTTAGAAGAG AAAAGGCAAAA--
NPB11-M13 **TTCC**TAAGCA TATTTCCCTC GTTAGAAGAG AAAAGGCNAAA--

Since there was no consensual sequence around the sequence of the OPB11$_{464}$ RAPD primer at the 5'end, the SCAR primer contains no sequence similarity to OPB11 primer. The 3'end SCAR primer was designed to start from the fourth nucleotide, at the start of the consensus sequence. Both primers had a %GC content of 38%. Table 4.1 is the summary of results of the design of SCAR primers, and the sequence of SA11$_{640}$ primers which were designed from a RAPD fragment amplified with the OPA11 primer (CAATCGCCGT). The reverse SA11$_{640}$ primer does not contain any sequence of the original RAPD primer.

Table 4.1 The summary of the SCAR primers design.

LOCUS	PRIMER	SEQUENCE[a]	% GC content
OPA17$_{438}$	SPA17-F	5' **GAC CCG TTG** TGG TGG TAG TCA	52
OPA17$_{438}$	SPA17-R	5' **GAC CCG TTG** TAA TAT AAG AAC	38
OPB11$_{464}$	SPB11-F	5' GCA TAC GAG TTA TAA TAG CTC	38
OPB11$_{464}$	SPB11-R	5' **GAC CCG** TAA TAC TGT ATT TCC	38
OPA11$_{640}$	SA11-F	5'**ATC GCC GTC** AAT GAG CAC TC	55
OPA11$_{640}$	SA11-R	5'AGG AGG TTG TAG TGA CAT CG	50

a - the highlighted sequences are the original RAPD decamer primer sequences.

4.3.3 Segregation analyses of SCAR markers

The segregation analysis of the SCARs was carried out in two parts 1) segregation in selected progeny and parents of cross AA157, and 2) selected varieties present in the genealogy of the cross and bulks of *S. spontaneum* and *S. officinarum*, the presumed progenitors of modern sugarcane varieties. In both cases, the SCAR marker was compared to its RAPD originator. The results are presented separately for each SCAR marker.

4.3.3.1 Segregation analyses of the RAPD $OPA17_{438}$ and SCAR $SPA17_{438}$ markers

The amplification profile of the SCAR marker when compared to its RAPD originator in selected commercial varieties reveals a reduction in the number of loci amplified (Figure 4.2 overleaf). The SCAR primer amplified the original polymorphic fragment but its segregation was monomorphic in the progeny of cross AA157, selected commercial varieties, bulks of *S. officinarum* and *S. spontaneum*. This indicates that the molecular basis of the original polymorphism was probably a point mutation in the priming site.

In addition to the non-segregating polymorphic fragment, the SCAR primer also amplified a second fragment (designated SPA_{430}) which was present and monomorphic in the RAPD profile. This may indicate that the original RAPD polymorphism was a result of a mutation in this fragment since its segregation was monomorphic in the RAPD profile. The results of the Southern blot analysis using the DIG labeled $NPA17_{438}$ plasmid confirmed the homology of the two bands. The results of Southern blot analysis also indicate that the RAPD primer amplified a number of partially homologous loci as evidenced by the weak hybridisation signals.

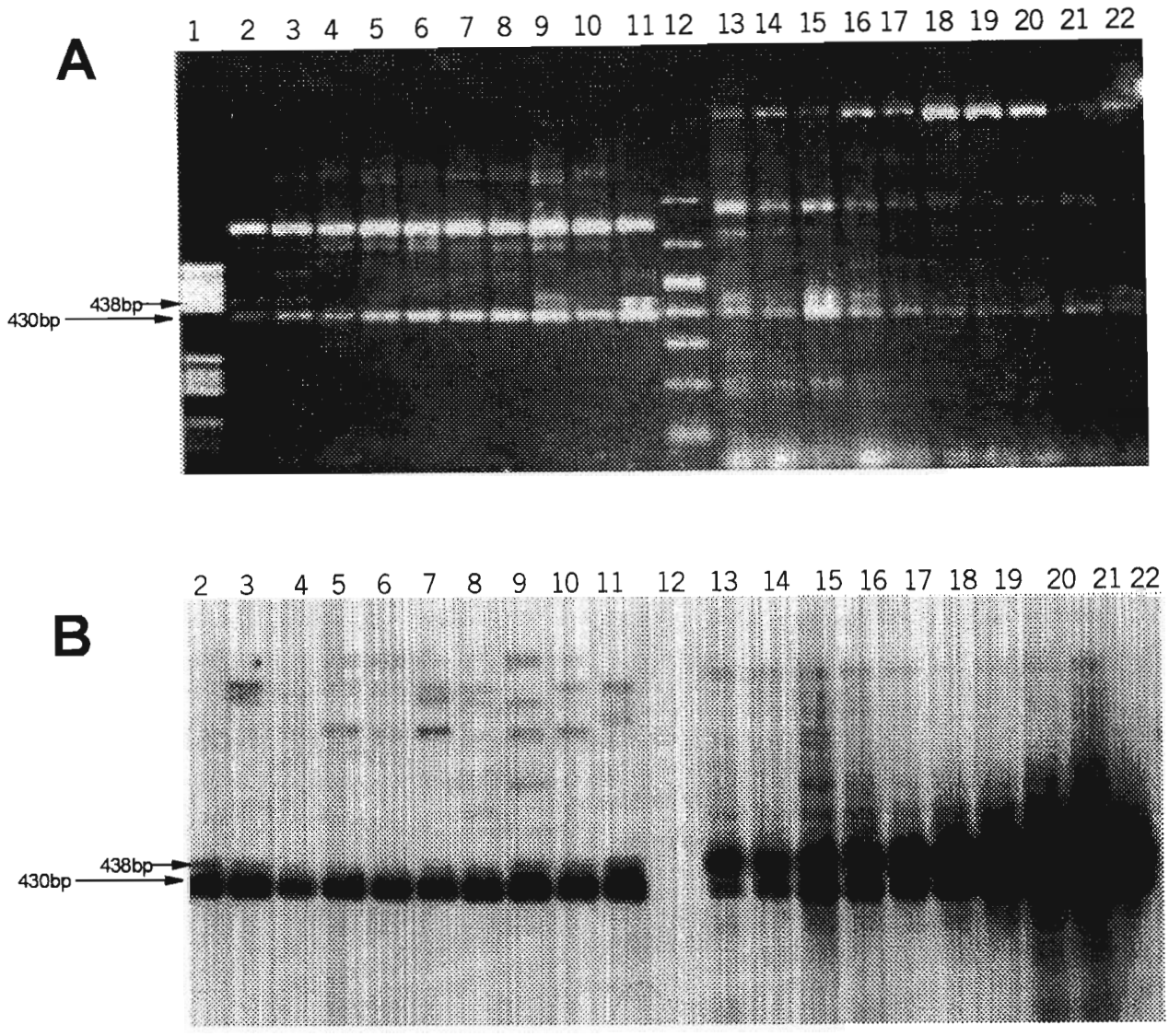


Figure 4.2 RAPD and SCAR amplification profiles for marker OPA17$_{438}$. A is the separation of RAPD and SCAR fragments and SCAR fragments on 2% agarose gel stained with ethidium bromide. B is the Southern blot analysis of the PCR results. Lane 1 is the pBR322 DNA-HaeIII marker, lane numbers 2-11 are the bulks of *S. spontaneum, S. officinarum,* and different commercial varieties amplified with the OPA17 RAPD primer and lane 12 is the 50-1000 base pairs ladder marker (Biomarker). Lane numbers 13-22 is the same array as lanes 2-11 amplified with the SPA17 SCAR primers. The array of the lanes is *S. spontaneum* bulk (2), *S. officinarum* bulk (3), NCo376 (4), NCo293 (5), J58/3 (6), N25 (7), 79H181 (8), N8 (9), POJ2878 (10) and Co285 (11). The bulks of *S. spontaneum* and *S. officinarum* , and Co285 appear to have the RAPD polymorphic fragment OPA17$_{438}$ however, the Southern blot results are negative.

The SCAR primer also amplified a number of bands, which did not correspond to any in the RAPD profile. These were revealed to be unrelated to the two loci. The analysis of the partial sequences of the OPA_{438} RAPD fragment for repetitive loci with the DNASYS software showed an abundance of these sequences. This raises the possibility that new priming sites may have been created as a result of the extension of the RAPD primer.

The fact that the co-migration fragment SPA_{430} is slightly smaller than OPA_{438} is an indication that the mutation differentiating the two loci is insertional in nature.

4.3.3.2 Segregation analyses of the RAPD $OPA11_{640}$ and SCAR $SA11_{640}$ markers

The RAPD profile of the selected progeny and parents of the cross AA157 did not amplify the expected size band using the original primer OPA11. When tested on the varieties only a weak amplification on variety J59/3 was detected, this was only evident in the Southern analysis. However, the SCAR primers were able to amplify a strong band in all the progeny tested and varieties (Figure 4.3).

Southern analysis of the segregation of this SCAR confirmed the PCR results. The results indicate that the RAPD priming site mutation that enabled the detection of the polymorphism in the first place was probably a point mutation or a mismatch, which was cross specific. It was not detected in the progeny of cross AA157, because neither of the parents had the RAPD priming site.

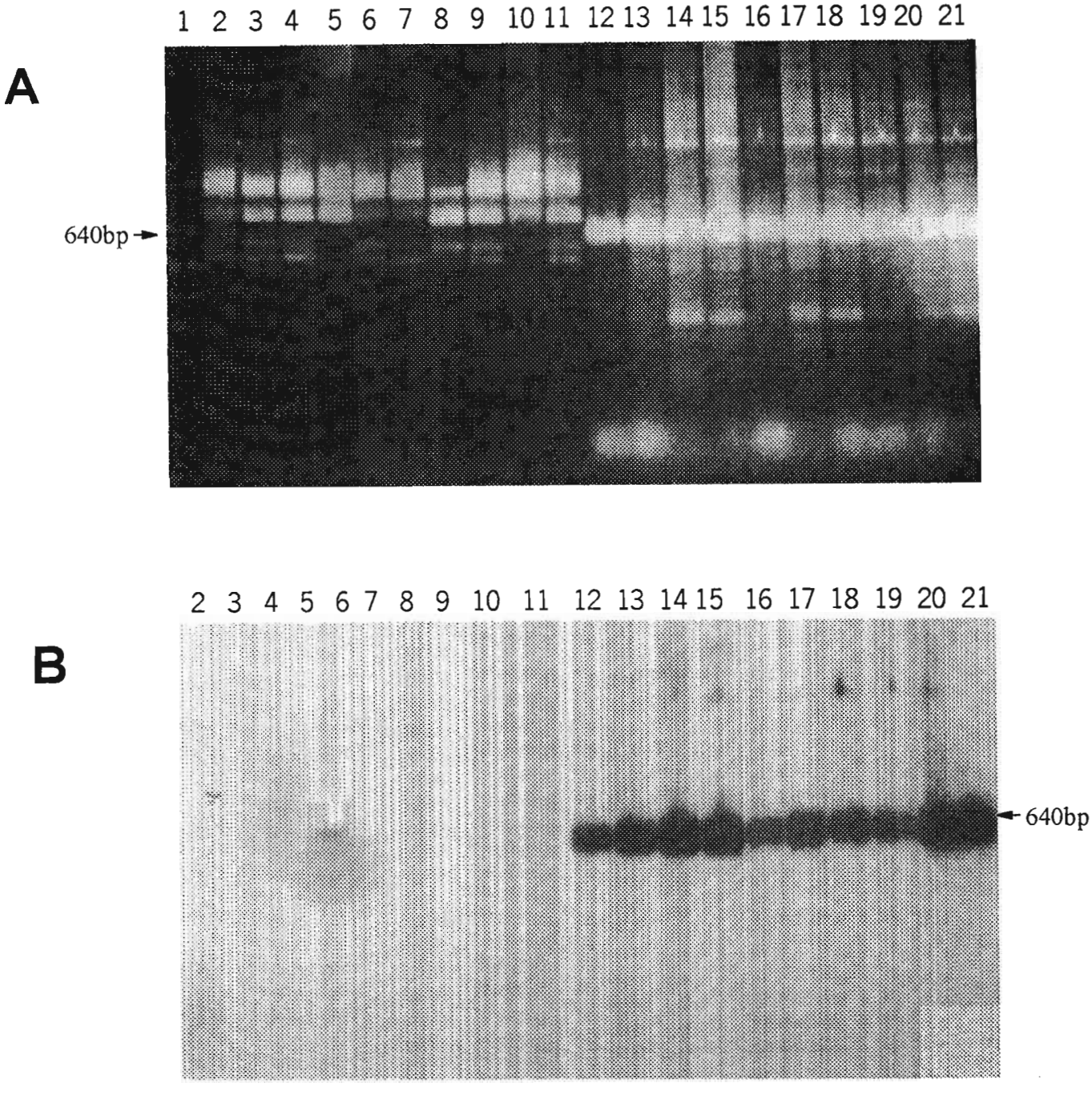


Figure 4.3 RAPD and SCAR amplification profiles of marker SA11$_{640}$. A is the separation of RAPD and SCAR fragments on 2% agarose gel stained with ethidium bromide. B is the Southern blot analysis of the PCR results with a DIG-labelled plasmid of the polymorphic RAPD fragment. Lane 1 is the 50-1000bp ladder marker (Biomarker), lane numbers 2-11 are the bulks of *S. spontaneum*, *S. officinarum* and different commercial varieties amplified with the OPA11 RAPD primer. Lane numbers 12-21 is the same array as lanes 2-11 amplified with the SA11 SCAR primers. The array of the lanes is *S. spontaneum* bulk (2), *S. officinarum* bulk (3), NCo376 (4), NCo293 (5), J58/3 (6), N25 (7), 79H181 (8), N8 (9), POJ2878 (10) and Co285 (11).

4.3.3.3 Segregation analyses of the RAPD OPB11 $_{464}$ and SCAR SPB11$_{464}$ markers

The results of the segregation analysis of OPB11$_{464}$ reveals two related amplified bands that are segregating in the progeny tested. The results in Figure 4.4 show a section of the individuals tested segregating for the polymorphism. Using Southern blot analysis the RAPD polymorphic fragment, OPB11$_{464}$ was found to have a high homology to the band below it, designated OPB11$_{440}$ indicating this may be its length variant.

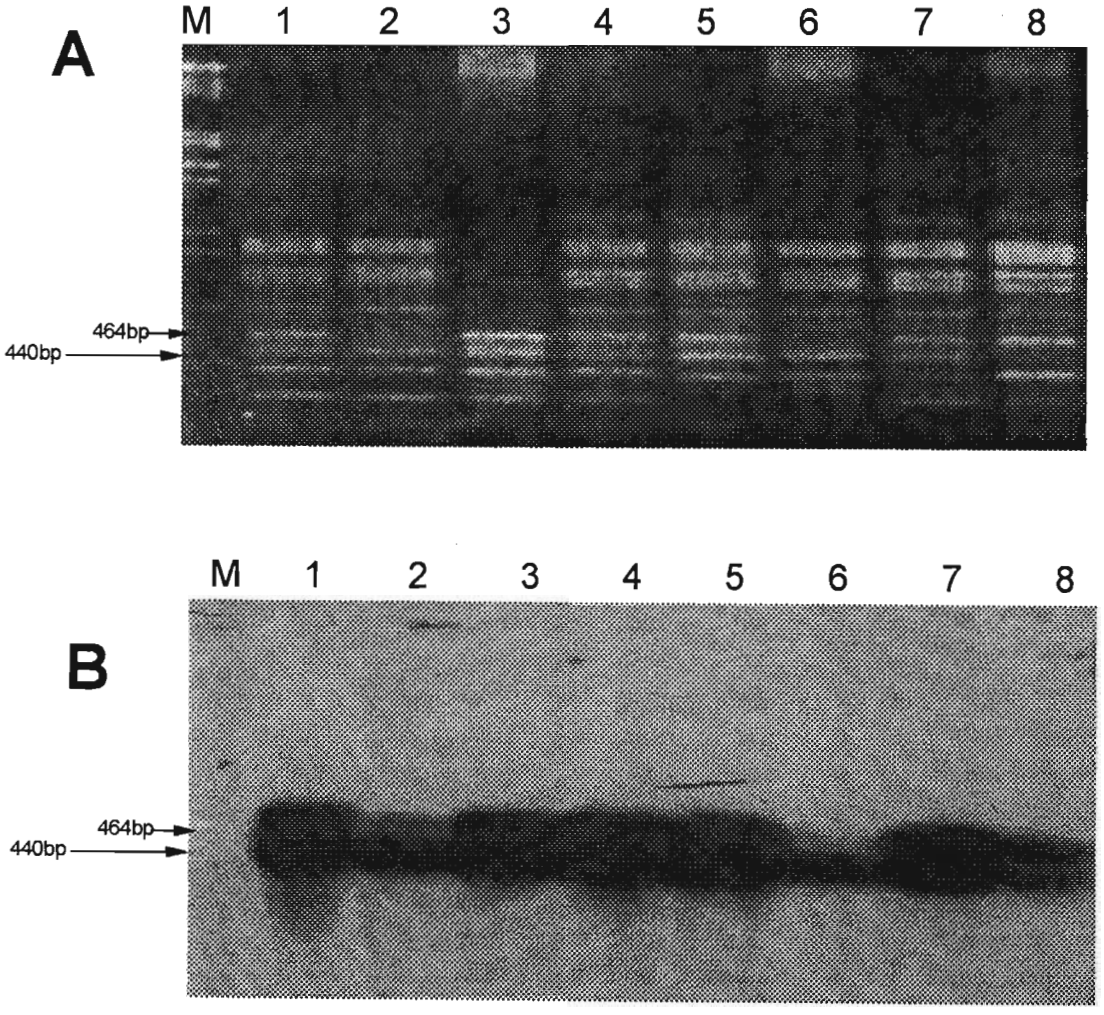


Figure 4.4 The RAPD profile of the polymorphic marker OPB11$_{464}$.. A. is the RAPD amplification profile of the segregating progeny of cross AA157, resolved on 2% agarose gel. B. is the Southern blot analysis of the RAPD results, prcbed with a DIG-labelled whole plasmid of the RAPD polymorphism. Lane 1 is lambda DNA -EcoR1/ Hind III marker. Lanes 2 - 9 are the selected progreny of cross AA157 (clones 1- 8). The arrowhead indicates the RAPD polymorphism and its length variant, OPB11$_{440}$.

The segregation analysis of the locus, $OPB11_{464}$, with the SCAR primers resulted in the loss of polymorphism. Furthermore, $OPB11_{440}$, the co-migrating band, was strongly amplified confirming that it is a length variant of the locus. The results in Figure 4.5 show the same individuals amplified with SCAR primers of this locus.

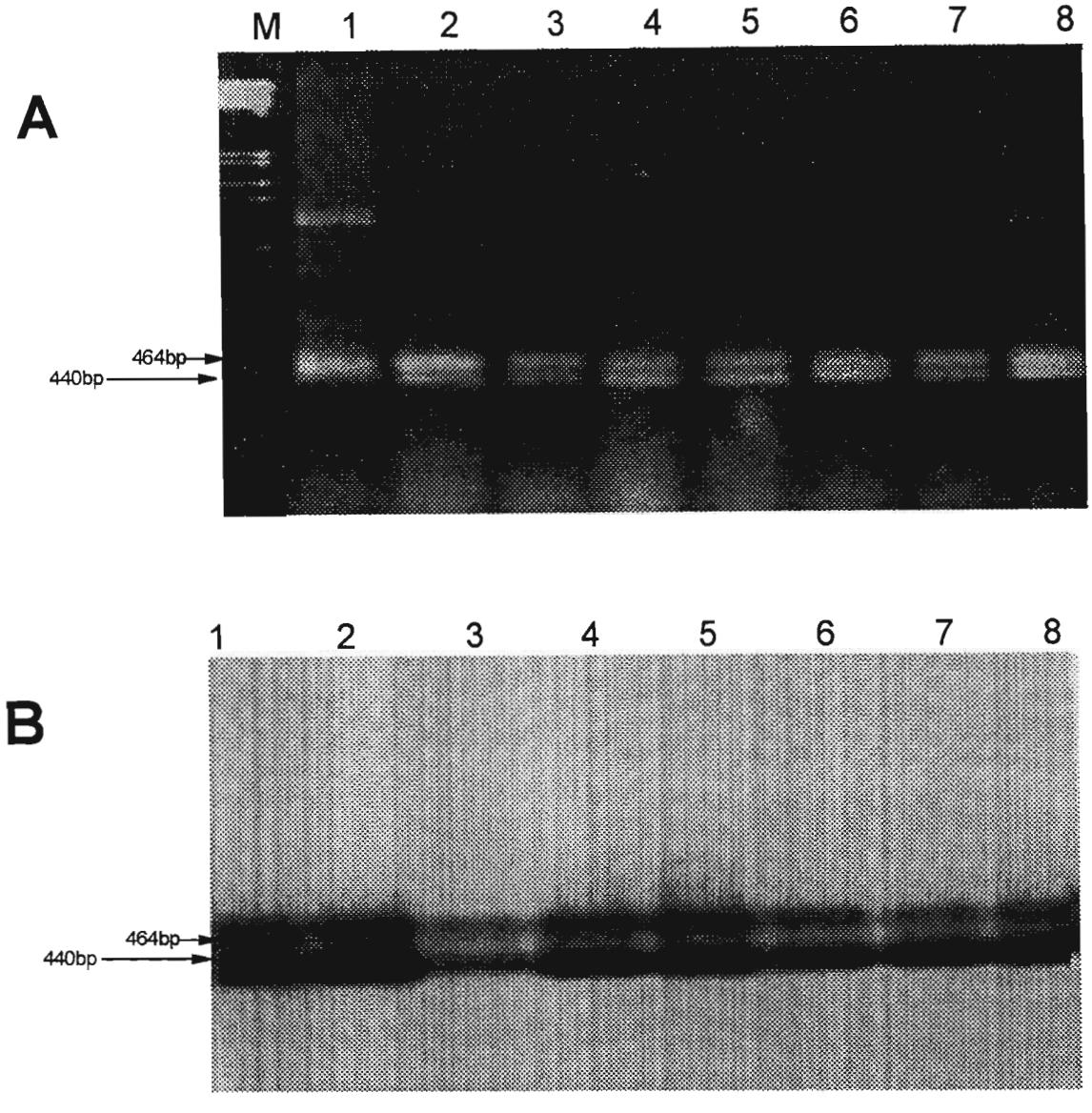


Figure 4.5 The SCAR amplification profile of the polymorphic marker $OPB11_{464}$.. A. is the SCAR amplification profile of the progeny of cross AA157, resolved on 2% agarose gel. B. is the Southern blot analysis of the SCAR results, probed with a DIG-labelled whole plasmid of the RAPD polymorphism. Lane 1 is lambda DNA -EcoR1/ Hind III marker. Lanes 2 - 9 are the selected progreny of cross AA157 (clones 1- 8). The arrowhead indicates the original RAPD polymorphism, now monomorphic and the co-migrating homologous locus.

4.3.3.4 Comparative analysis of the OPC16$_{889}$ RAPD marker

Figure 4.6 shows the results of the amplification of different varieties with the OPC16$_{889}$ RAPD primers. Southern blot analysis results indicate that this is a single copy locus.

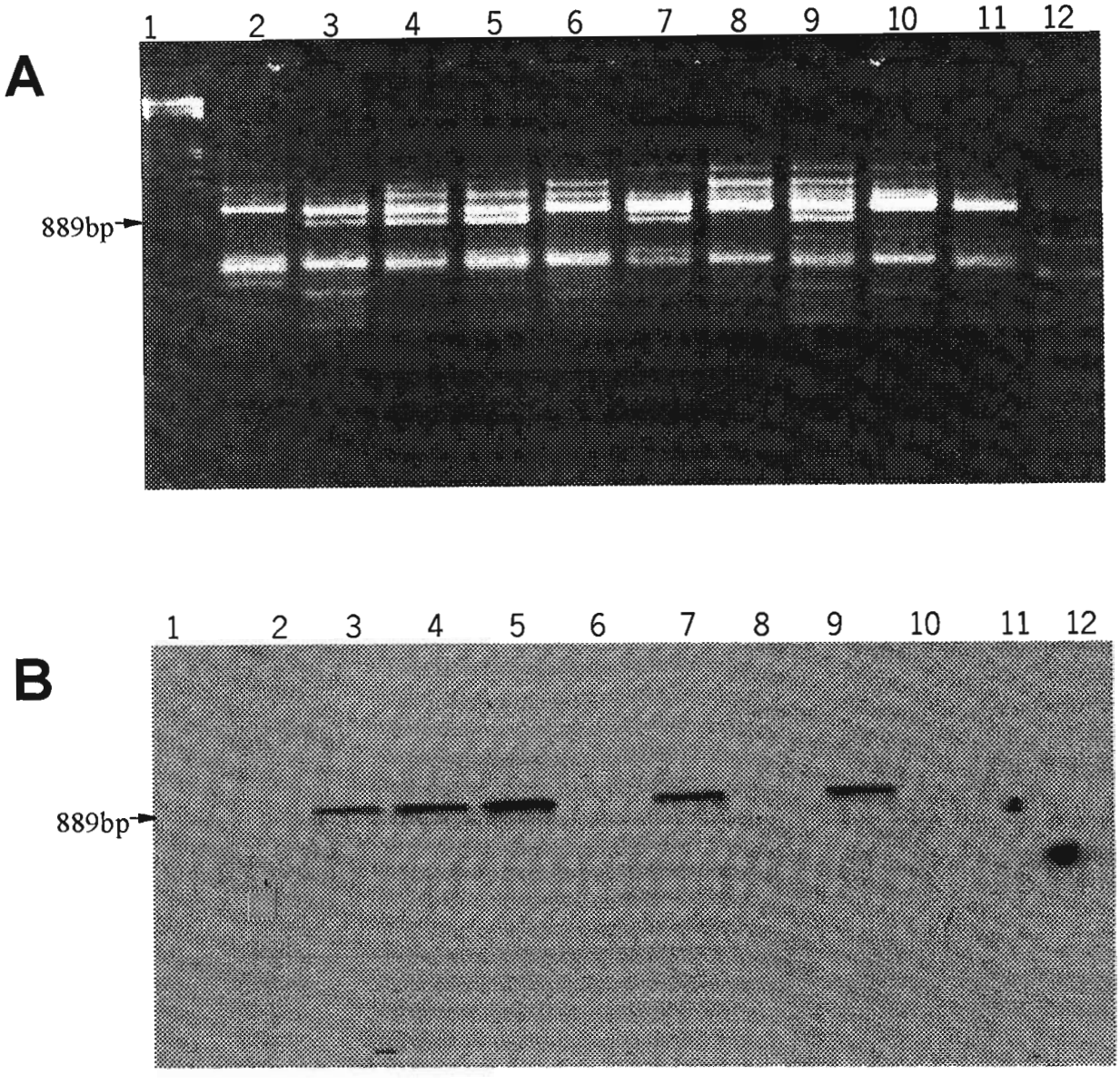


Figure 4.6 The RAPD profile of the stably inherited OPC16$_{889}$ marker of bulks of *S. spontaneum* and *S. officinarum*, and modern sugarcane varieties. A is the separation of RAPD fragments on 2% agarose gel stained with ethidium bromide. B is the Southern blot analysis of the PCR results with a DIG-labelled plasmid of the RAPD fragment. Lane 1 is the Lambda DNA EcoR1/Hind III marker, lane numbers 2-11 are the bulks of *S. spontaneum, S. officinarum* and different commercial varieties amplified with the OPC16 RAPD primer and lane 12 is the 50-1000bp ladder marker (Biomarker). The array of the lanes is *S. spontaneum* bulk (2), *S. officinarum* bulk (3), NCo376 (4), NCo293 (5), J58/3 (6), N25 (7), 79H181 (8), N8 (9), POJ2878 (10) and Co285 (11).

These results indicate that the RAPD marker $OPC16_{889}$ in not variety specific, and is stably inherited across the varieties tested. It segregated as a single dose fragment in the cross AA157 (section 3.3.5). Its presence in the *S. officinarum* bulk can be traced to its presence in two *S. officinarum* clones comprising the bulk, Black Tanna and Badilla.

Finally, the origin of the RAPD-based markers investigated in this study indicate that most are present in both *S. officinarum* and *S. spontaneum*, when assayed using the SCAR primers (Table 4.2).

Table 4.2 Summary of the origin of the polymorphic fragments with respect to the *S. officinarum* and *S. spontaneum* progenitors.

Marker	**Origin**	
	S. officinarum	*S. spontaneum*
$OPA17_{438}$	–	–
$SPA17_{438}$	+	+
$OPB11_{464}$	–	–
$SPB11_{464}$	+	+
$OPA11_{640}$	–	–
$SA11_{640}$	+	+
$OPC16_{889}$	+	–

The RAPD results with the three primers OPA11, OPA17, and OPB11, did not yield the expected band in both the *S. officinarum* and *S. spontaneum* bulks. This is an indication that the RAPD markers, $OPA11_{640}$, $OPA17_{438}$, and $OPB11_{464}$, are all cross-specific markers.

4.4 Discussion

The segregation ratio of polymorphic fragments is an indication of the dosage of the genetic locus. A mutation in the priming site of one homologue results in a 1: 1 segregation ratio (Wu *et al.*, 1992), since only 50% of the progeny will inherit the polymorphic locus. The other 50% will inherit the null allele. Converting the RAPD fragment to a SCAR may/may not result in the amplification of the null allele depending on the nature of the mutation that caused the polymorphism.

The detection of RAPD polymorphisms due to a deletion mutation or total absence of the priming site should not result in loss of polymorphism with conversion to SCARs. This type of polymorphism is usually detected in cases where introgression of genetic loci occurs from a different genetic background e.g. a wild relative (Maisonneuve *et al.*, 1994, Qi *et al.*, 1996). Absence of the priming site in the null allele is due to an introgression destroying the priming site. Alternatively, the new genetic locus has a priming site, which is not present in the new genetic background. Deletion mutations in the priming site have the same effect.

Polymorphisms detected due to insertion at the priming site will not be detected in the null allele with SCARs since their primers are designed from the sequence of the positive allele. Insertions internally may also prevent amplification if the fragments inserted are too long thus making the distance too large to support amplification. In all these cases, the conversion to a SCAR marker may serve as a technically better marker for usage in MAS.

A point mutation or a mismatch at the priming site should result in the amplification of a fragment of the same size in all the individuals in the SCAR analysis. This means that

converting RAPD markers to SCARs will result in the loss of polymorphism. In all the three SCARs investigated in this study, the polymorphism was lost indicating that the RAPD polymorphism had been due to either a mismatch or a point mutation in the priming site of the homologous genetic locus.

The conversion of RAPD markers to SCARs serves another purpose in genetic analyses, that of distinguishing other copies of the locus (Paran and Michelmore, 1993, Adam-Blondon *et al.*, 1994, Ohmori *et al.*, 1996). In a diploid situation where there are only two copies, conversion to SCARs may reveal the homologous locus. In a RAPD profile, the homologous loci are indistinguishable from the other unrelated loci. Since SCAR primers are locus specific, the homologous locus that is differently sized because of an insertion or deletion can be resolved. In a diploid situation the presence of a second locus indicates heterozygosity (the second locus should be specific to the other parent).

The resolution of co-migration bands $OPA17_{430}$ and $OPB11_{440}$ as separate but homologous loci, in all individuals and varieties tested, is an indication that SCARs may distinguish zygosity in sugarcane. The fragments $OPA17_{438}$ and OPB11464 are probably derived from the other homologues in the genome. The limiting condition is that the detection of heterozygosity is only possible if the locus has an insertion or deletion internally. Since sugarcane is a high polyploid (Sreenivasan *et al.*, 1987), the non-segregating co-migration fragments are presumed to represent the other copies of the locus. In both these cases, the co-segregating fragments were smaller indicating that the polymorphic RAPD loci were a result of an insertion in the locus.

The degree of detection of co-dominant RAPD markers in crop species has been very low (Schulz *et al.*, 1994, Davis *et al.*, 1995). This indicates that the majority of mutations responsible for RAPD polymorphisms are sequence related. The failure to detect co-dominance with a higher degree is due to the difficulty in the fulfillment of the condition required for their detection. In diploids, an insertion or deletion internal of the genetic locus is required such that the presence of one fragment is always accompanied by absence of the other differently sized but homologous fragment (complete repulsion).

In sugarcane the chances of detecting co-dominant RAPD markers is lessened by the high number of homologous loci estimated between 8x to 12x (Sreenivasan *et al.*, 1987, D'Hont *et al.*, 1994, D'Hont *et al.*, 1996). An insertion in one homologue will result in its segregation in single dose fashion in the progeny. However, the other homologous loci will not segregate, as they will be in multiple dosage. A condition for the detection of repulsion phase linkages in sugarcane is bivalent pairing between homologues, indicative of genetic differences at the molecular level or unequal homology between different homologues (Da Silva and Sobral, 1995). D'Hont *et al.* (1994) and Grivet *et al.* (1996) have observed preferential pairing in sugarcane, especially with respect to chromosomes derived from the *S. spontaneum* progenitor.

In the present study the segregation of the RAPD and SCAR markers can be attributed to two levels. In the first level, the insertion mutations resulted in the segregation of markers $OPA17_{438}$ and $OPB11_{464}$ in a single dose fashion, since only 50% of the progeny had these markers. The co-segregating fragments, $OPA17_{430}$ and $OPB11_{440}$, representing the other homologues did not segregate because of higher dosage. If this is the only level present, the

conversion of RAPD markers to SCARs should have resulted in the amplification of both the original RAPD fragment and the co-segregating fragments.

However the relationship should have been preserved i.e. the RAPD polymorphism retained since it is still present in only 50% of the progeny. The hypothesis of a point mutation or a mismatch in the priming site with respect to the other unknown homologous loci can be discarded on the basis of the amplification of both length variants in the RAPD and SCAR analyses, and the Southern blot analyses results.

The loss of the RAPD polymorphism due to the SCAR PCR amplification of the null allele indicates the presence of the same genetic locus in the other parent. Since mutations that are detected as segregating polymorphisms occur in the parents, and are transmitted to progeny in a fixed ratio, it means their detection as polymorphic in the progeny is dependent on the non-occurrence of a corresponding mutation in the other parent. If there is no corresponding or similar point mutation in the same locus of the other parent then they will be detected as polymorphic in the RAPD analysis.

The RAPD segregation analysis results of both $OPA17_{438}$ and $OPB11_{464}$ taken on their own indicate that the above conditions are valid i.e. a mutation causing length variants in one parent, and the progeny segregating in a single dose fashion for the polymorphic fragment. The null allele is actually due to the absence of this mutation in the other parent. However the amplification of the null allele by the SCAR primers indicates the polymorphic length variant locus is present in the other parent.

The conclusions that can be drawn from the amplification of the null allele is that the two mutations are independent. The first type of mutation is only detected with the RAPD analysis of this population and therefore can be presumed to be a cross specific mutation. It is most probably a negative mutation i.e. it prevents the amplification of the genetic locus in the other parent. This is based on the assumption that the polymorphic length variant is derived from the monomorphic loci. In evolutionary terms, it is the more recent one. The results of the analysis with $SA11_{640}$ indicate that the RAPD polymorphism originally detected belongs to this class of markers.

The second mutation resulting in length variation is stably inherited, and can distinguish the different homologues. This is confirmed by the fact that both fragments were amplified in all progenies, parents and commercial varieties. The segregation of the higher length variants in a single dose ratio was therefore due to 1) the fortuitous mutations in the homologous allele of one parent and 2) segregation of the length variant.

The single dose segregation of the length variants in the two cases studied is indicative of preferential pairing. The proof of this conclusion is that the genetic loci, $OPA17_{438}$ and $OPB11_{464}$, are single dose in both parents, otherwise the null alleles would have been amplified. The rationale for this observation is that since the polymorphism is due to a negative mutation, the presence of higher dose alleles of the null parent would have resulted in amplification, and therefore monomorphism.

The implication of these observations is that single dose RAPD markers, used in conjunction with SCARs may be used to identify chromosomes where preferential pairing is observed. This is obviously dependent on a second mutation resulting in a null allele in the RAPD

analysis. The detection of these relationships with such a limited sample is probably due to the different genomes in sugarcane. However, it should be stated that the genetic loci studied were present in both *S. spontaneum* and *S. officinarum* clones investigated, so that the origin of the fragments could not be attributed solely to either of them.

Furthermore, the power of the RAPD assay to detect polymorphisms due to point mutations and mismatches is an important contributing factor to the detection of these relationships. The detection of this relationship has important implications for the application of RAPD-based markers (including SCARs) in a breeding programme.

Cross utility of RAPD-based markers across populations and species

Cross utility of RAPD markers across populations has to be considered according to two frameworks. The first one is the utility of the marker for species-wide genetic map integration. In this case, RAPD polymorphisms that are detected as a result of mutations in the actual priming site are of little use since those mutations may be cross specific. The results of this study with all three SCARs indicate that the conversion of such RAPD marker will result in their loss of polymorphism, which means that they cannot be used for tagging important genes. Da Silva (pers comm [1]) also reported a loss of polymorphim, when RLFP markers were converted to locus specific PCR markers.

Where the RAPD marker is detected, because of an insertion, deletion or transposition event at the 3'end of the priming site and internally, applicability across populations should be expected. In fact, such a marker may be scorable as a co-dominant marker depending on the length of the mutation i.e. is the longer genetic locus PCR amplifiable?

[1] JAG da Silva, Centro de Tecnologia Copersucar, C.P. 162, Piracicaba SP 13400, Brasil

The second framework is the origin of the mutation in the priming site, this means that a mutation that is stably inherited across generations is most likely to be cross-utilised in different populations. However, its utility in the breeding programme will depend on its recombination fraction with the locus of interest because loose linkage may mean it cannot be detected in other crosses. The result showing the stability of RAPD marker $OPC16_{889}$ across different cultivars is a justification for its retention.

Urrea *et al.* (1996) also detected co-dominant RAPD marker linked to the *bgm-1* gene, conferring resistance to golden mosaic virus in common bean that were applicable across all the germplasm tested. This marker was not only gene pool non-specific but also could be amplified across several laboratories, using similar but also distinct DNA extraction and PCR protocols.

However, they noted that when using lower quality DNA obtained from mini-prep extraction protocols, a Stoffel fragment or similarly engineered DNA polymerase might be critical to the reaction. In this case the justification for converting a RAPD marker to a SCAR marker has to be weighed against its recombination fraction with respect to the locus it is tagging.

Cross utilisation of a RAPD marker in common bean was also reported by Haley *et al.* (1994), where a RAPD tightly marker linked to the *I* gene (potyvirus resistance) was found to be applicable when tested in 40 different crosses. The cross utilisation of RAPD markers in different populations has also been shown with the $OPO6_{950}$ linked to the rust resistance gene in sunflower. Huckett and Botha (1995) also reported several RAPD fragments that could be amplified across different crosses in the genealogy of sugarcane, one of, which is the putative marker $OPC16_{889}$ (for fibre %cane) that has been reported in this study. This marker was

reported as $OPC16_{904}$, the slight difference in size can be attributed to the different methods of measuring the size. In sugarcane the gene pool is restricted so that it is possible that a RAPD marker may be applicable across populations.

The detection of preferential pairing in sugarcane, and the low levels of the *S.spontaneum* genome (5-10%) reported by D'Hont *et al.* (1994), provide the only case where conversion of RAPD markers to SCARs may be useful. In these cases, the dosage of the genetic locus is the limiting condition since multiple dosage loci will not be detected whether in RAPD or SCAR form.

Certainly, if a single dose genetic locus of any genetic origin, tagging a trait or genetic locus of interest is detected with the RAPD assay it can be converted to a SCAR marker for the purposes of MAS. Recently, a RAPD marker putatively linked to rust resistance in a commercial sugarcane variety has been converted to a SCAR marker, and preliminary results indicate that length variants of the genetic marker can be resolved (J. Barnes, pers comm)[1]. Other specific PCR-based markers are finding practical application in confirming true crosses between *Erianthus* and *Saccharum* species in sugarcane breeding programmes (D'Hont *et al.*, 1995, Harvey *et al.*, 1997). The stability of the marker and the dosage, conditioned by preferential pairing, will determine the efficacy of such markers and their widespread application in MAS.

[1] J. Barnes, SASEX, Biotechnology Department, Private Bag X02, Mount Edgecombe, 4300

Chapter 5

Searching for RAPD marker - trait associations

5.1 Introduction

Since the advent of molecular markers which have made it possible to dissect quantitatively inherited traits, simultaneous statistical models have been developed and refined to identify, locate and resolve the individual genetic loci affecting quantitative traits. In general these models assume that both the marker and trait data are recorded on the same individuals drawn from segregating populations of either doubled haploid, recombinant inbred, backcross derived, F_2, or F_3 progenies (Fisch *et al.*, 1996). These models present restrictions in terms of application to breeding programs since breeding populations are usually advanced lines (Fisch *et al.*, 1996). Furthermore, the initial choice of mapping populations is motivated by the need to maximise polymorphisms (Paterson and Wing, 1993). It is generally accepted that populations that are used for map construction are not necessarily the most suitable ones for QTL dissection, however, practical considerations necessitate that models for QTL dissection should be adapted to correct for this paradox.

In predicting the size of QTLs, it has been shown that F_2s have a lesser power of resolution than doubled haploid or recombinant inbred populations. This is due to the higher number of marker classes in F_2s (three vs two) which confound the estimation of QTLs (Rebaï *et al.*, 1994). In addition, linkage relationships between QTLs and marker loci may also bias the estimate of QTLs (Lander and Botstein, 1989). In highly complex species like sugarcane, the

task becomes even more intractable because the classes of genotypes are numerous (Da Silva, 1993).

The transmission genetics of sugarcane (2n + n), in the crosses between *S. officinarum* and *S. spontaneum*, with the resultant erosion of the *S. spontaneum* component genome provides an opportunity to investigate the extent of linkage disequilibrium in modern commercial varieties. *S. spontaneum* clones are characterised by very low sugar and high fibre content, while *S. officinarum* have high sugar content and high fibre (Babu, 1979). This implies that some of the genetic loci involved in the control of these characters may carry alternate alleles, with opposite effects, in modern sugarcane varieties.

This observation coupled with the reported low levels of recombination between *S. officinarum* and *S. spontaneum* derived chromosomes in modern varieties (D'Hont *et al.*, 1996), indicate that the possibility may exist for tagging genetic loci controlling these characters with molecular markers. This observation is based on the assumption that the low levels of recombination between the chromosomes of the two species are an indication of sufficient sequence differences, thus making them amenable to RAPD analysis.

The present study was therefore a preliminary attempt to investigate the possibility of finding molecular markers linked to these quantitative traits in sugarcane. The genealogy of the cross AA157 indicates that there has only been one introgression of *S. spontaneum* (Appendix A). In addition, the two parents of cross, N14 and 79F2011, share the same genetic lineage, which

implies that the phenotypic segregation of the progeny from low to high extreme of the trait distributions, may be indicative of strong linkage disequilibria.

The investigation was extended to include the effect of different environments on the search for molecular markers linked to traits. This was made possible by the availability of phenotypic data from three seasons (1992, 1993 and 1994), at different ages of growth (12 months, eight months and nine months respectively).

5.2 Materials and methods

5.2.1 Plant material

Plant material and the sampling of the phenotypes were as described previously (Chapter 3). The whole population (150 individuals) was assayed for the 1992 season. Three replicates (450 individuals) were assayed in the 1993 season. In the 1994 season only one replicate of the 80 individuals, whose marker segregation was analysed for the BSA, were assayed. The phenotypic data for 1994 is from a ratoon crop, which introduces the element of repeatability of traits.

5.2.2 Quantitative traits

In addition to fibre %cane, the traits of brix %cane, pol %cane and estimated recoverable sucrose (ers %cane) whose data was available for the 1992, 1993 and 1994 seasons were chosen for the analysis. The cane was harvested at different ages for all three years, 12 months for 1992, 8 months for 1993, and 9 months for the 1994 season. Since there was only one

location (Pongola), which means that three environments (years and/or age of harvest) were assayed (Mather and Jinks, 1974).

Correlation of traits was estimated using the Pearson product moments (Sigma Stats 2.0, 1996). The significance threshold was set at $p < 0.001$, unless indicated. These were based on the phenotypic data of the same 80 individuals. Correlations were calculated between years for a trait to estimate the effect of years/and or age of harvest in a single trait. They were also calculated between the traits for each year to estimate the degree of association between traits. No partitioning, to estimate the genetic and environmental components was carried out. The correlations between traits were done simply as a measure of association, on the assumption that clonal variation due to the environment was constant in each trait.

Clonal repeatability estimates were calculated for each trait by performing an ANOVA on the 1993 season phenotypic data, with replications and clones considered as random effects. Clonal repeatability was estimated from the equation:

$$\frac{F-1}{F}$$ (Falconer and Mackay, 1997)

where F is the phenotypic variance ratio. The estimation was based on all the individuals (450) in the replicates.

5.2.3 Molecular markers

A set of 22 RAPD markers, nineteen of which were single dose fragments, generated from the bulk segregant analysis was used (Chapter 3). Two RAPD markers were linked. The three RAPD markers whose segregation ratio was skewed were also included in the analysis.

5.2.4 Marker-trait analysis

Fifty-three progeny of the cross AA157 were used for identification of RAPD markers linked to quantitative trait loci. The 53 individuals were selected on the basis of 'full marker data' i.e. no missing data for each RAPD marker used in the study. The investigations of the associations between markers and traits were carried out using single point analysis (Soller *et al.*, 1976) with multiple regression analyses. Single-marker analysis has been suggested in sugarcane because of the meiotic instabilities in elite crosses (Da Silva, 1993). The trait data for the 1993 season was the mean of the three replications.

The multiple regression analyses were performed in two steps. Firstly, backwards stepwise regression, with all the markers, was utilized to find the best model explaining phenotypic variation. The two linked markers $OPA15_{667}$ and $OPB16_{618}$ were used separately to avoid the problems of collinearity (Bridges and Sobral, 1996). Collinearity in regression arises as a result of measuring essentially the same quantity under different names (Mosteller and Tukey, 1977). Since the location of QTLs in relation to the two linked markers, $OPA15_{667}$ and $OPB16_{618}$, is not known this may create problems in regression analysis.

The markers that were in the best model were then used in the second step to estimate the percentage trait variation ascribed as a function of the regression coefficient, R^2. The second step involved performing linear or multiple regression analyses of the form:

$$y = b_0 + b_1x + b_2x + \ldots + b_nx_n$$

where y is the dependent phenotypic value and x is the independent marker variable, and b_0 the constant. Markers are added sequentially to the equation for the multiple marker analyses model.

5.3 Results

5.3.1 Heritability estimates for different traits

The clonal repeatabilities were estimated from the F-values in the 1993 season, and these were taken to indicate heritabilities (Table 5.1). The degree of genetic determination (DGD) for the three traits of fibre %cane (0.85), pol %cane (0.93), and ers %cane (0.95) were high.

Table 5.1 Analysis of variance for the four traits of fibre %cane, brix %cane, pol %cane and estimated recoverable sucrose (ers) %cane for the 1993 season.

Source of variation	Degrees of freedom	Mean squares			
		Fibre %cane	Brix %cane	Pol %cane	ers %cane
Groups	1	0.01	32.96	60.65	77.18
Reps in grp1	2	0.14	1.05	5.99	9.74
Reps in grp2	2	8.99	0.45	6.95	13.40
Indv. In grp1	74	6.07	1.28	1.60	1.94
Indv. In grp2	74	7.41	2.11	2.59	3.00
Indv. In 1+2	148	6.74	1.70	2.10	2.47
Residual	296	1.34	0.40	0.48	0.71
Total	449	3.14	0.90	1.20	1.56
F-value		6.72	1.12	14.55	18.91
(F – 1)/F		0.85	0.11	0.93	0.95

Reps = replications, grp = group, and indv = individuals.

Brix %cane had the lowest degree of genetic determination value (0.11).

5.3.2 Correlation of the same trait in different environments (years)

The investigation of correlation between traits was done specifically for two reasons. Firstly it was to determine if there were any significant differences in a particular trait from one season to the next. This was done to ascertain if the different seasons had any effect on the traits, and whether their correlations could then be used to make generalised inferences about environmental effects, specifically the year and age of cane variables in the G x E equation (Simmonds, 1981).

5.3.2.1 Brix %cane

The measurements for brix %cane were poorly correlated across the three years (Table 5.2). The correlations between the 1992 and 1994 seasons were the lowest (0.19). The correlations between 1994 and the other two years were significant only when the stringency was relaxed, by lowering the significance levels to 80% and 90% (Table 5.2).

Table 5.2 Phenotypic correlations of brix %cane on a year to year basis.

Year	1992	1993	1994
1992	-	0.375	0.19[b]
1993	0.375	-	0.24[a]
1994	0.19[b]	0.24[a]	-

a - only significant at 90% probability level, and b - denotes 80% significance level

The lowest correlations between the two years most distant (1992 and 1994) reflects that the year variable is an important factor in the determination of the brix %cane phenotype in this population. However, since correlations between 1993 and 1994 were not highly significant

(90%), other factors may be involved. A more plausible explanation for the low correlation between brix %cane measurements in 1994 and measurements in the two other seasons is the low broad sense heritability (DGD = 0.11). The low DGD indicates the presence of environmentally sensitive QTLs or many QTLs with small effects, which may affect trait expression in the ratoon crop. This observation is supported by the significant correlations in the plant crop years (1992 and 1993).

5.3.2.2 Pol %cane

The phenotypic correlations in pol %cane are from low to intermediate (Table 5.3). The highest correlations were for 1994 vs 1993 pol %cane data (0.58), and the lowest correlation coefficients were for 1992 vs 1994 (0.23).

Table 5.3 Phenotypic correlations of pol %cane on a year to year basis.

Year	1992	1993	1994
1992	-	0.34	0.23[a]
1993	0.34	-	0.58
1994	0.23[a]	0.58	-

a - denotes significant at 90% probability level

The high correlation for 1993 vs 1994 seasons is an indication that both age of harvest and year variables were important determinants of the pol %cane phenotype. The correlations of 1992 vs 1992 were intermediate, confirming the year variable. The lack of correlation in 1992 vs 1994, which are both distant in years and age of harvest is a confirmation of this observation.

5.3.2.3 Ers %cane

Age of harvest was found to be important in ers %cane. The lowest phenotypic correlations were in 1992 and 1994 (Table 5.4). Large differences in correlations between the data for 1992 vs 1993 (0.33), and 1993 vs 1994 (0.63) were observed. The implication of this observation is that the correlation of 1992 vs 1993 are mainly due to the closeness in years, since the age of cane is different. The almost doubling of the correlation coefficient for 1993 vs 1994 (0.63) compared to 1992 vs 1993 (0.33) can best be explained on the basis of the closeness of the age of cane in 1993 and 1994 (eight and nine months respectively).

Table 5.4 Phenotypic correlations of ers %cane on a year to year basis.

Year	1992	1993	1994
1992	-	0.33	0.29
1993	0.33	-	0.63
1994	0.29	0.63	-

The correlations for ers %cane in different environments are similar to those observed for pol %cane, which is an indication of the genetic relationship between these two traits.

The major difference in correlations for all three sucrose related traits were observed for 1992 vs 1994 data, where both brix % cane and pol % cane were only correlated at lower significant levels, while ers %cane had an intermediate value (0.29). This may be due to the fact that the high degree of genetic determination (0.95) for ers % cane is indicative of stable QTLs across both the environment and crop and ratoon stages. This observation is a generalisation on the whole trait, not individual genetic loci.

5.3.2.4 Fibre %cane

The lowest correlations in the fibre trait were observed between 1992 and 1994, only when the significance threshold had been dropped to 90% (Table 5.5). This lack of correlation is indicative of the importance of the year variable.

Table 5.5 Phenotypic correlations of fibre %cane on a year to year basis.

Year	1992	1993	1994
1992	-	0.57	0.24[a]
1993	0.57	-	0.54
1994	0.24[a]	0.54	-

a denotes 90% significance level

The small differences between the 1992 vs 1993 (0.57) and 1993 vs 1994 (0.54) data indicates that the age of cane is less of a contributing variable than years in the phenotype of fibre %cane.

5.3.3 Correlations between traits in a single season (one environment)

The investigation of correlation between traits in a single season was done in only one environment to minimise the number of factors that would confound the analysis. The environment was assumed to be constant, so that the correlations are an estimate of the genetic correlations.

The phenotypic correlations between traits in a single season, where clonal variation due to the environment is presumed to be constant, was done to ascertain the relationships at the genetic

level. This information could then be used to investigate further the RAPD marker-trait associations. The rationale for this is that genetic correlations are caused by pleiotropic effects where one gene affects more than one trait in either direction or close linkages between genes affecting different traits (Falconer and Mackay, 1997). Phenotypic correlations have been shown to closely approximate genetic correlations for several traits in sugarcane (James and Falgout, 1969, James, 1971, Reimers *et al.*, 1982, Wu *et al.*, 1983, Tai and Miller, 1989).

5.3.3.1 Phenotypic correlations in 1992

Brix %cane, pol %cane and ers %cane were highly correlated in the 1992 season as the results indicate (Table 5.6). Fibre was negatively correlated to all the other three traits.

Table 5.6 Correlations for the traits for the 1992 season when cane was harvested at 12 months old.

Trait	Fibre %cane	Brix %cane	Pol %cane
Fibre %cane	-	-0.45	-0.35
Brix %cane	-	-	0.95
Pol %cane	-0.35	0.95	-
Ers %cane	-0.38	0.91	0.99

The negative correlation of fibre %cane and the three sucrose related traits is possibly as a result of carbon partitioning, since in broad terms fibre %cane represents one of the components of total insolubles, and the three other traits are components of solubles in cane.

5.3.3.2 Phenotypic correlations in 1993

The phenotypic correlations of brix %cane, pol %cane, and ers %cane were very high in the 1993 season (Table 5.7).

Table 5.7 The correlations for the traits for the 1993 season when cane was harvested at 8 months old.

Trait	Fibre %cane	Brix %cane	Pol %cane
Fibre %cane	-	Ns	Ns
Brix %cane	Ns	-	0.95
Pol %cane	Ns	0.95	-
Ers %cane	Ns	0.91	0.99

NS = not significant.

Fibre was not correlated to any of the sucrose-related traits, which is a deviation from the previous season. Since fibre was found to be less responsive to the age of harvest, the age of harvest of the other three traits may be responsible for this observation. The test of this hypothesis should be the measure of the similarity of the observations in the 1994 season with the data for 1993, since the age of harvest is almost similar.

Furthermore, since the four traits being measured are actually components, among others, of solubles and insolubles; the relationship between them may be confounded in young cane. The different rates of sucrose accumulation observed in sugarcane varieties may play a role (Babu, 1979, Inman-Bamber, 1996), since the progeny of cross AA157, are essentially equivalent to varieties since they have different genotypes.

5.3.3.3 Phenotypic correlations in 1994

The results of the 1994 season indicate that brix %cane and pol %cane were significantly correlated at 0.84 (Table 5.8). Ers %cane was also highly correlated to both traits as in the previous season. Fibre is again not correlated to any of the other traits confirming the observation of the importance of the age of harvest for the other three traits.

Table 5.8 The correlations for the traits for the 1994 season when cane was harvested at 9 months old.

Trait	Fibre %cane	Brix %cane	Pol %cane
Fibre %cane	-	Ns	Ns
Brix %cane	Ns	-	0.84
Pol %cane	Ns	0.84	-
Ers %cane	Ns	0.72	0.96

Ns denotes not significant

The results of the correlations between traits in a single environment taken together indicate that the traits of brix %cane, pol %cane, and ers %cane are highly correlated. Their high correlation is an indication that the same genetic loci or very close linkages may control them. This is further confirmed by the observation that the high correlations were constant in all the environments, despite the fact that within-trait correlations differed substantially in the different environments.

5.3.4 RAPD markers

Twenty-two RAPD markers were investigated for their association with the four traits, in the three seasons (Table 5.9).

Table 5.9 Segregation of RAPD markers, and their χ^2 values.

RAPD marker	Present	Absent	Missing	χ^2 *
$A12_{736}$	36	42	2	0.46
$A15_{667}$	39	38	3	0.013
A16	27	40	12	2.52
$A17_{438}$	39	41	0	0.05
$A19_{1239}$	36	43	1	0.62
$A20_{1218}$	41	36	3	0.32
$A20_{641}$	43	33	4	1.32
$B1_{778}$	34	44	2	1.28
$B11_{464}$	45	32	3	2.19
B12	42	31	7	1.66
$B13_{1532}$	34	45	1	1.53
$B13_{500}$	51	28	1	6.69[b]
$B13_{681}$	44	33	3	1.57
$B15_{517}$	31	48	1	3.66
$B16_{618}$	33	36	11	0.13
$C4_{321}$	31	41	8	1.39
$C7_{798}$	35	44	1	1.03
$C7_{869}$	28	51	1	6.70[b]
$C11_{352}$	43	30	7	2.31
$C16_{889}$	37	41	2	0.205
$C17_{1822}$	29	51	0	6.05[b]
$C17_{567}$	37	35	8	0.05

* - less than 3.84 is significant at 95% probability level for single dose segregation (1:1),

b - the segregation of these markers is skewed.

In addition to the 19 single dose RAPD markers, the three RAPD markers whose segregation was skewed were also included in the analysis. Skewness of marker segregation is not expected to affect the marker-trait linkage analysis (Stuber *et al.*, 1992).

5.3.4.1 RAPD marker – Phenotypic trait associations

A total of 18 RAPD markers were found linked to the four traits investigated in the three seasons. The effect of the RAPD markers was mainly additive. The quantitative variation they ascribed ranged from 7.6% fibre %cane ascribed by marker $OPB11_{464}$ in 1992 to 29.6 fibre %cane ascribed by three markers in 1993 (Table 5.10).

For brix %cane the quantitative variation ascribed by markers ranged from 9.6% ascribed by one marker in 1993 to 16.3% ascribed in 1994. The quantitative variation ascribed by the markers for ers %cane ranged from 12% in 1993 to 15.3% in 1994; while two markers explained 17.2% of the pol %cane variation in 1992 and three markers ascribed 25.4% variation in 1994. The sum of the percentage trait variation ascribed by each individual marker was equal to the total. The exception was for pol %cane in 1994, where the sum of the individual pol %cane (21.3%) was less than that of the multiple regression model (25.4%). This result may be due to epistatic interactions of the markers.

Only four markers were found common across different environments, and three of them were linked to the fibre %cane trait. The fourth marker was linked to ers%cane. No marker was common in all three environments. Two markers ($OPA17_{438}$ and $OPB11_{464}$) were linked to more than one trait.

Table 5. 10 Summary of RAPD-marker trait associations over the three seasons.

Trait	Marker	1992	1993	1994
			R^2	
Brix %cane	$OPC7_{831}$	8.3	-	-
	$OPB1_{778}$	5.8	-	-
	$OPB11_{464}$	-	9.6	-
	$OPA15_{667}$	-	-	10.3
	$OPA17_{438}$	-	-	6.0
Pol %cane	$OPA17_{438}$	8.5	-	-
	$OPA15_{667}$	8.7	-	-
	$OPB11_{464}$	-	-	14.6
	$OPB13_{500}$	-	-	4.6
	$OPC17_{1822}$	-	-	2.1
Ers %cane	$OPB11_{464}$	13.5	5.9	-
	$OPA20_{641}$	-	6.1	-
	$OPA17_{438}$	-	-	8.9
	$OPA15_{667}$	-	-	6.3
Fibre %cane	$OPC16_{889}$	7.6	6.0	-
	$OPA17_{438}$	-	17.6	3.1
	$OPB16_{618}$	-	6.6	3.1
	$OPB1_{778}$	-	-	3.8

The brix %cane results indicate that markers found putatively linked to brix %cane differ from year to year, with no common markers. There are two possible explanations for the detection

of different markers in each year. One is that the linkages in any one-year may have been spurious. The second reason may be that since the brix %cane phenotype has a low degree of genetic determination (0.11), the QTLs controlling its expression may be environmentally sensitive. Paterson *et al.* (1991) detected markers linked to QTLs in tomato, which are specific to particular environments.

Two markers, $OPA17_{438}$ and $OPA15_{667}$, were found to be linked to pol %cane in 1992. They ascribed a total of 17.3% pol %cane phenotypic variation. No association with pol %cane was detected in 1993. Three markers ascribing a total of 25.4% pol %cane phenotypic variation in 1994 were detected. The sum total of the individual pol %cane variation ascribed by markers individually was 21.3%, which indicates that there may be interactions between the markers. No common markers were detected for the different environments.

One marker $OPB11_{464}$ was associated with the ers %cane phenotype in two seasons. There was a reduction of more than 50% in the level of ers %cane ascribed by marker $OPB11_{464}$ in the 1993 season indicating the influence of the environment on the genetic locus.

Common markers were found associated with fibre phenotype in different years with a higher degree than any of the other traits. The common markers are $OPC16_{889}$ for the 1992 and 1993 seasons, $OPB16_{618}$ and $OPA17_{438}$ for the 1993 and 1994 seasons.

The observable trend in the common sets of markers detected in different environments mimics the phenotypic correlations between environments. For example, the phenotypic

correlations for fibre were stronger for the 1992/1993 and 1993/1994 data than for the 1992/1994 season and so are the detected linkages of the fibre phenotypic and common markers. No markers were detected for the 1992 and 1994 seasons.

5.3.4.4 Common markers

Based on the strong phenotypic correlations of brix %cane, pol %cane, and ers %cane in any one season, common markers were expected to be observed in the three traits. The 1992 data indicates that no common markers were detected between these traits. In 1993 marker $OPB11_{464}$ was associated with both brix %cane and ers %cane (Table 5.11).

Table 5.11 Common markers for different traits

Year	RAPD marker	Traits
1993	$OPB11_{464}$	Brix %cane, ers %cane
1994	$OPA17_{438}$	Brix %cane, pol %cane, ers %cane, fibre %cane

Marker $OPA17_{438}$ was associated with all four traits in the 1994 season, including fibre %cane, which was not phenotypically correlated to any of the other three traits.

The failure to detect common markers in 1992 and the low numbers detected in 1993 and 1994 despite the high phenotypic correlations is a result of the small sample of markers. This restricted sampling of a larger proportion of the genome. However, detection of common

markers in the different environments, despite limitations on sample size, is an indication of their strong linkage to the traits.

5.4 Discussion

Heritabilities are specific to environments and populations used. They can vary for different populations in the same environment, since each population has different genotypes (Falconer and Mackay, 1997). Broad sense heritabilities can be used to predict the genetic gains from one clonal selection to the next (Falconer and Mackay, 1997). In the present investigation, clonal repeatabilities were used as an estimation of the degree of genetic determination (DGD) of the four traits. This estimate is important in molecular marker analysis because it can give an indication *a priori* of the chances of success when searching for marker–trait associations. In the present study fibre %cane, and the two sucrose related traits of pol %cane and ers %cane had high DGD values (0.85, 0.93, and 0.95 respectively). These high values are possibly a consequence of two different factors.

Fibre %cane represents the first factor, which is strong linkage disequilibrium. There has been a limited introgression of *S. spontaneum* in the genealogy of this cross, followed by successive cycles of what essentially amounts to inbreeding (Appendix A). The two parents of the cross, are also genetically closely related as indicated by only 10.5% RAPD variation detected (Table 3.4).

A large proportion of the genetic variation in cross AA157 may be due to the *S. spontaneum* component of the genome. A large percentage of genetic variation in *Saccharum* has

previously been shown to be due to the *S. spontaneum* component (Glaszmann *et al.*, 1989, Eksomtramage *et al.*, 1992, D'Hont *et al.*, 1993, Al-Janabi *et al.*, 1994a, Lu *et al.* 1994a,b, Sobral *et al.*, 1994). This factor coupled with the 2n + n transmission in *Saccharum* and the resultant erosion of the *S. spontaneum* component in the genome of modern sugarcane cultivars (Gill and Grassl, 1984), implies that the phenotypic traits under the genetic control of this component are likely to have at least moderate to high heritabilities.

Since high fibre %cane in sugarcane is characteristic of the *S. spontaneum*, the expectation is that the probability of finding markers linked to its trait loci are substantially higher than the sucrose related traits. The detection of a high number of markers linked to fibre %cane (three vs one for the other traits combined) in different environments seems to confirm this hypothesis. This high level of detection cannot be explained solely on the basis of the sensitivity of BSA. This is because the RAPD markers used in this investigation are presumed to be random with respect to their distribution in the genome (Williams *et al.*, 1990).

Genetic variation, in the RAPD analysis context, is an indication of DNA sequence variation. By extension this means opposite alleles for a genetic locus. It precludes variability due to permutations of different trait alleles, which may be an important determinant of the trait distribution. The two sucrose related traits of pol %cane and ers %cane represent the genetic variation due mainly to the different combinations. Since most of these alleles are possibly derived from the *S. officinarum*, the intense selection implies that the gene frequency is biased towards favourable alleles (Figures 3.1 and 3.2). This may explain the detection of only one marker for sucrose related traits, $OPB11_{464}$ linked to ers %cane, in more than one environment.

This is despite the high DGD values for both pol % cane and ers %cane, and the overall detection of equal numbers of putative markers.

Low broad sense heritability estimates may indicate that the trait has a large environmental variance, which may indicate many QTLs of small effects (Mather and Jinks, 1974). It has been established that QTLs of small effects are more likely to escape detection (Wang and Paterson, 1994). Brix %cane had low values of degree of genetic determination (0.11) in this trial. Brix %cane is not directly selected for in the breeding programme, and therefore its gene frequencies are not as skewed towards high levels as the other two sucrose related traits.

The possibility therefore exists that a substantial number of alleles with opposite effects for the trait are present. This may be the reason why the number of markers linked to brix %cane is the same as that of the three highly heritable traits. However, the low heritability of the trait is evident in the non-detection of common markers in different environments. The sensitivity of QTLs to different environments has been reported in other studies (Paterson *et al.*, 1991, Bubeck *et al.*, 1993, Tinker *et al.*, 1996, Brummer *et al.*, 1997).

The heritabilities could not be estimated directly for the two seasons of 1992 and 1993, as a consequence of the nature of the data i.e. lack of replicated data. The alternative was to use the set of environments presented by the data to measure correlations between traits as a generalised measure of genetic determination of the traits. The traits with the highest correlation values were chosen to search for molecular markers linked to QTLs since they

provided an additional internal control or cross validation to infer genetic relationships between putative markers and QTLs.

Correlations are important in genetic analysis because they measure the degree of association (genetic and non-genetic) between traits (Hallauer and Miranda, 1988). If genetic association exists then selection for one trait may cause a change in other correlated traits. This effect is due to pleiotropic effects or linkage disequilibrium. Correlations between characters may be environmental, which is found within genetically uniform individuals (e.g. NILs) or phenotypic which is a combination of both environmental and genetic correlations (Falconer and Mackay, 1997). In this study, the linkage of marker $OPB11_{464}$ to the three traits that were highly correlated is indicative of their genetic basis.

The within-trait correlations in the study were generally low to intermediate. The implications of the low correlations for some of the traits from year to year is that they provide a readily available tool for cross-validation of markers found in one environment, since the different trait data is equivalent to different trials. In addition, since the genetic trait loci are constant, the reduced levels of correlations indicate that the trait loci are influenced by the environment.

The four markers that were detected across environments represent linkage either to QTLs that have large effects, or that are not sensitive to the environment. In this study, three of the four environmentally consistent QTLs were linked to fibre %cane. A high proportion of QTLs for yield traits in maize has been reported to be environmentally stable (Stuber *et al.*, 1992, Veldboom and Lee, 1996).

The widely used method of interval mapping to search for markers linked to QTLs is only applicable to species or populations where reasonably saturated genetic maps are available (Lander and Botstein, 1989). In cases where no genetic maps are available, the single marker at a time method is the only available tool. Several statistical approaches can be used to search for marker trait associations including t-tests between marker means and ANOVAs to establish linkages between the markers and traits. To estimate the quantitative variation due to the QTL, regression methods are usually utilized (Bridges and Sobral, 1996). The choice of method used can impose restrictions on the analysis and introduce biases in the results. Several studies have been carried out on the nature of these biases (section 2.2.5).

In searching for marker-trait linkages, the null hypothesis is that the presence or absence of particular RAPD fragments has no bearing on the trait. Since only twenty two RAPD fragments were analysed for their associations to four traits, and the alpha level was set at <0.001, the size of the Type 1 error i.e. association due to chance was expected to be 1% which is 0.22 spurious linkages per trait. Averaging the number of markers investigated and traits over the three seasons, 264 marker-trait associations were investigated. This means that in the whole analyses approximately three linkages were expected to occur by chance. Eighteen linkages of RAPD markers to traits were detected which is six times the number expected by chance. This indicates that the majority of the linkages are real.

Only single dose fragments were analysed for their association to trait loci to minimise the effects of dosage in the estimation of quantitative loci effects. Higher dosage markers linked to QTLs may result in false positives i.e. plants with the marker but without the genetic trait

loci (Da Silva, 1993). Furthermore, higher dosage markers may confound the estimation of QTLs due to the higher number linkage phases. This is especially crucial in this study since point analysis was used on a random set of markers, and also the linkage phase between markers and trait loci was unknown.

Using single dose RAPD fragments, only two phases are possible 1) marker-QTL in coupling and 2) marker-QTL in repulsion. This imposed restriction reduces the data set that can be used, and thereby the chances of marker-QTL associations that are likely to be detected. However, it also reduces the bias that can be caused by unresolved linkages.

The detection of marker-trait associations with such a limited sample of markers is an indication of the potential for intercrosses to serve as QTL mapping populations in sugarcane. The number of RAPD marker–trait associations and the quantitative variation was quite high considering the sample size. This indicates that intercrosses may be suitable for mapping QTLs in sugarcane provided the population has enough replications to estimate heritabilities. It also means that populations that are already in the breeding trials and have undergone rigorous phenotypic evaluation can be readily available for QTL dissection. Fortunately most of these populations consist of large numbers of clones, since one of the major limitations of the present study was the number of clones. This means that these results are, at best, preliminary in nature.

Finally, in the search for marker-trait associations the effect of clonal repeatability between crop plants and ratoon crops has to be taken into consideration. This is because the results of

the generally low levels of correlations in the traits between 1994 (ratoon crop) and the 1992 and 1993 seasons (plant crops) was closely mimicked by the lack of common markers for 1994 and the other two seasons. This is crucial for eventual application of MAS since sugarcane is grown for a number of ratoons (Milligan *et al.*, 1990).

Chapter 6

General Conclusions

Genetic studies in sugarcane have accelerated over the past five years, raising hopes that the gains made will be applicable to MAS. Several marker systems have been employed, with varying degrees of emphasis, in the study of sugarcane genetics. In earlier studies, isozymes were widely utilized for analysis of genetic diversity (Glaszmann *et al.*, 1989, Eksomtramage *et al.*, 1992). However, the limitations associated with isozyme analysis, coupled with their paucity necessitated the utilisation of more effective marker systems.

RFLPs, which had found successful application in other crop plants were an obvious choice for map construction and genetic analyses (Burnquist, 1991, Da Silva *et al.*, 1993, D'Hont et al., 1993, Lu *et al.*, 1994a and b, Grivet *et al.*, 1996, Besse *et al.*, 1997, Dufour *et al.*, 1997). Due to its high throughput the RAPD technique has also found increasing usage in sugarcane genetic analyses (Al-Janabi *et al.*, 1993, Al-Janabi *et al.*, 1994a,b, Harvey *et al.*, 1994, Msomi and Botha, 1994, Sobral *et al.*, 1994 , Huckett and Botha, 1995, Sills *et al.*, 1995, Harvey and Botha, 1996).

However, the sensitivity of the RAPD technique to variations in reaction conditions may impose restrictions with regard to its widespread usage. This is because these variations may result in unbalanced data. However, this observation is tempered by the fact, in molecular tagging studies such as the present one, the segregation in the progeny of a putative RAPD marker provides a useful control to monitor spurious RAPD PCR artifacts, which can then be discarded.

Since the scoring of RAPD bands is a subjective process, it can lead to biases. For example, in this study the percentage of single dose fragments may have been underestimated due to the proximity or co-migration of bands, which were excluded from further analysis. This limitation can be ameliorated by using a better system of PCR products resolution such as polyacrylamide gels. In addition, the high throughput of the RAPD technique also compensates for the loss due to exclusion of ambiguous fragments.

Some of the peculiar RAPD properties have important implications for tagging studies based on BSA. One example is the bulk contamination genetics of RAPDs, whereby the success or failure of amplification of a locus may be dependent on the locus and/or percentage contribution in the bulks. While in this study control measures were employed by assaying the individuals comprising the bulks, the locus/dosage dependence implies some useful markers may escape detection. However, RAPDs are still possibly the best marker system for BSA precisely for this peculiarity. No detection of markers is possible in contaminated bulks with RFLPs, because of their high fidelity. A contaminating individual in the null bulk will always prevent identification of the null locus, because of the presence of the restriction site. This is crucial in investigations where individuals comprising the bulk are not included in the initial stages of the BSA.

The suitability of any cross in sugarcane for BSA has to be considered from two perspectives. The first is the possible hybrid genome of modern sugarcane varieties. Recent findings that both *S. robustum* and *S. officinarum* are possibly segmental allopolyploids (Andersen and Fairbanks, 1994, Al-Janabi *et al.*, 1994b); and coupled with the fact that even though *S. spontaneum* is an autopolyploid, its chromosomes pair preferentially in the genome of modern sugarcane varieties (D'Hont *et al.*, 1996, Grivet *et al.*, 1996).

The behaviour of SCAR markers provides direct evidence of preferential pairing in cross AA157. Preferential pairing is indicative of differences in sequence and/or origin of homologues (Da Silva and Sobral, 1996). This means that the chances of detecting single dose fragments are enhanced, because at anaphase the two homologues will separate and end up in different haploid meiocytes. These may also carry trait loci variants, thus increasing the likelihood of finding markers linked to trait loci.

In this study all the bulk specific fragments segregated in a single dose ratio. This is a consequence of using the BSA to search for markers linked to traits, in a polyploid like sugarcane. Only markers segregating in a single dose fashion can be detected. Markers in higher dosage cannot be detected, and these may be important in dosage conditioned traits.

The results of the BSA (Chapter 3), taken together with the random search for marker-trait associations (Chapter 5) indicate that the efficacy of BSA in sugarcane will probably be limited to traits with high linkage disequilibrium. The genetic loci in linkage disequilibrium should also have sufficiently large effects so that they can be detected in different environments and selection stages (crop plant to ratoon). In addition, estimation of broad sense heritabilities using clonal repeatabilities as a criterion for the suitability of a trait for the BSA is inadequate. The source or reason for the high heritability should be accurately characterised. A distinction should be made between high heritability due to skewed gene frequencies favouring only a particular extreme phenotype such as high sucrose yield, or linkage disequilibrium.

The observation that $OPA17_{438}$ and $OPB11_{464}$ pair preferentially in the parents of the cross AA157, yet are detected in both *S. officinarum* and *S. spontaneum*, may indicate that the fragments are derived from a common ancestor of both progenitors. The mechanism of the mutations of the two fragments provides an insight into the events that are implicated in the evolution of modern sugarcane varieties. The three SCAR markers that were investigated in this study were found present in both *S. officinarum* and *S. spontaneum* bulks. However, the RAPD fragments from which they were derived did not amplify in either *S. officinarum* or *S. spontaneum*, which indicates recent origin of the point mutations.

The conversion of RAPD markers to SCARs is likely to be informed by 1) the need for studying the genetic structure of the *Saccharum*, *vis a vis* 2) the cost of development, which is dependent on their possible application in MAS. In studying the genetic structure of sugarcane, SCARs may play a role in distinguishing zygosity. However, in this regard, they will only be able to play a limited role because of the number of homologues that are involved for any genetic locus. A more immediate application of SCAR technology in sugarcane genetic analysis and MAS, is the role they can play in distinguishing loci with different genetic origin, since these would be easily resolvable. RAPDs can identify these loci, but their sensitivity to point mutations may prevent amplification of the same locus in other individuals. This would have the effect of misclassification since a null result may be misinterpreted as representing absence of the genetic locus.

Another pertinent issue with regard to the conversion of RAPD markers to resolve dosage is the possibility that no new information regarding dosage may be obtained. In a case where the single dose segregation of a RAPD marker is caused by a mismatch or a point mutation in the priming site, the other copies of the genetic locus will be resolved as monomorphic fragments

of the same mobility in the progeny. The initial test would be to perform Southern blot analysis with the polymorphic fragment to reveal related loci.

In a case where the single dose segregation of a polymorphic RAPD fragment is a result of an insertion or deletion mutation it will be possible to differentiate the two differently sized homologs. In this case, SCARs will be able to reveal genetic events like preferential pairing. They may therefore be useful for gene tagging. Southern blot analysis on the RAPDs on its own is not sufficient for this type of analysis, because it does not reveal the nature of the null allele.

As sugarcane maps become saturated, and more efforts are directed at QTL mapping, the issue of suitable populations will become more pertinent. The plasticity of sugarcane provides a formidable challenge in the identification of QTLs linked to traits. Several factors are responsible for this plasticity; the high ploidiness being one of the major reasons which gives rise not only to different possible combinations of QTL alleles, but also different dosages of such permutations. This is manifested in the phenotypic segregation observed for most traits in sugarcane.

Since QTL detection is dependent on the availability of alternate marker and trait locus variants and recombination distance between marker and trait loci, wide crosses involving parents with extreme phenotypes provide the most suitable populations. However, to restate, since trait distribution in a polyploid like sugarcane involves not only alternate trait loci variants but also combinations and dosage factors, the utility of wide crosses involving modern varieties may be limited.

This limitation may be due to the inability to differentiate extreme phenotypes on the basis of linkage disequilibrium, or trait loci combinations and dosage. Notwistanding the level of detection of RAPD markers associated with fibre and sucrose related traits suggests that a fair amount of linkage disequilibrium exists in modern sugarcane varieties. It is instructive that the only unambiguously gene pool specific RAPD marker ($OPC16_{889}$) in this study was traced to a *S. officinarum* origin (Table 4.2).

The loss of association of most of the detected markers to the sucrose-related traits in different environments is due to their loose linkage to trait loci. This is a function of the marker sample size, and therefore genome coverage, in the study. Alternatively, the loss of association between RAPD markers and sucrose-related traits is due to linkage of markers to environmentally sensitive QTLs. This contention may be a demonstration of the nature of genetic control of sucrose related traits. On the other hand, the retention of a higher number of RAPD marker-fibre trait associations is indicative of linkage disequilibrium with respect to this trait. This provides an indication of which traits will be amenable to tagging with molecular markers, either through BSA or traditional map-based QTL analysis; and the level of map saturation required for successful QTL analysis.

References

Adam-Blondon AF, Sevignac M, Bannerot H, Dron M (1994). SCAR, RAPD, and RFLP markers linked to a dominant gene (*Are*) conferring resistance to anthracnose in common bean. *Theor Appl Genet* **88**: 865-870.

Ajmone-Marsan P, Monfredini G, Ludwig WF, Melchinger AE, Franceschini P, Pagnotto G, Motto M (1995). In an elite cross of maize a major quantitative trait locus controls one-fourth of the genetic variation for grain yield. *Theor Appl Genet* **90**:415-424.

Akagi H, Yokozeki Y, Inagaki A, Fujimura,T (1996). Microsatellite DNA markers for rice chromosomes. *Theor Appl Genet* **93**: 1071-1077.

Akkaya MS, Bhagwat AA, Cregan PB (1992). Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* **134**: 1131-1139.

Aldrich PR, Doebley J, Schertz KF, Stec A (1992). Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. *Theor Appl Genet* **85**: 451-460.

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 (1994b). Chromosome assortment in *Saccharum*. *Theor Appl Genet* **89**: 959-963.

Al-Janabi SM, McClelland M, Petersen C, Sobral BWS (1994a). Phylogenetic analysis of organellar DNA sequences in the *Andropogoneae: Saccharinae*. *Theor Appl Genet* **88**: 933-944.

Allard RW (1956). Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* **24**: 235-278.

Allen FL (1994). Usefulness of plant genome mapping to plant breeding. *In*: Gresshoff,PM (ed). Plant Genome Analysis, CRC Press, Boca Raton, pp 11-18.

Andersen WR, Fairbanks DJ (1994). Genetic mapping in sugarcane using single dose RAPD markers. Report to the International Sugarcane Biotechnology Consortium, Plant Genome Congress, San Diego.

Anonymous (1985). Laboratory Manual for South African Sugar Factories. South African Sugar Technologists' Association. Mount Edgecombe.

Anonymous (1988). Selection of biomass. *In*: Annual Progress Report of South African Sugar Association Experiment Station 1988: 220-221.

Anonymous (1993). Instruction Manual: pCR-Script™ SK(+) cloning kit. Stratagene, La Jolla, California.

Anonymous (1995a). Instruction Manual : Gel Extraction Kit. Qiagen. Chatsworth, California.

Anonymous (1995b). Instruction Manual : Nucleobond AX. Manherev-Nagel, Germany.

Anonymous (1996a). Instruction Manual: AFLP™ Plant Mapping Kit. Perkin Elmer.

Anonymous (1996b). Instruction Manual : CDP-Star™. Boehringer Mannheim, Germany.

Arnheim N, Strange C, Erlich H (1985). Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: Studies of the *HLA* class II loci. *Proc. Natl. Acad. Sci. USA* **88**: 6970-6974.

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992). Map-based cloning of a gene controlling *Omega-3* fatty acid desaturation in *Arabidopsis*. *Science* **258**: 1353-1354.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1995). Current Protocols in Molecular Biology **I**. John Wiley & Sons Inc., USA.

Ayliffe MA, Lawrence GJ, Ellis JG, Pryor AJ (1994). Heteroduplex molecules formed between allelic sequences cause nonparental RAPD bands. *Nucleic Acids Res* **22**: 1632-1636.

Babu CN (1979). Sugarcane. Allied Publishers Private Limited. New Dehli.

Batcha MBGR, Palanichamy K (1978). Studies on the inheritance of ligular process in the interspecific cross *S.officinarum* (var. Green Sport) x *S. spontaneum* (CBE). *Proc Int Soc Sugar Cane Technol* **16**: 227-231.

Becker J, Heun M (1995). Mapping of digested and undigested random amplified microsatellite polymorphisms in barley. *Genome* **38**: 991-998.

Beckmann JS, Soller M (1983). Restriction fragment length polymorphisms in genetic improvement: Methodologies, mapping and costs. *Theor Appl Genet* **67**: 35-43.

Beer SC, Siripoonwiwat W, O'Donoughue LS, Souza E, Matthews D, Sorrels ME (1997). Associations between molecular markers and quantitative traits in an oat germplasm pool: Can we infer linkages? *Journal of Quantitative Trait Loci*, vol 3 (1).

Bender W, Spierer P, Hogness DS (1983). Chromosomal walking and jumping to isolate DNA from *Ace* and *rosy* loci and the bithorax complex in *Drosophila melanogaster*. *J Mol Biol* **168**: 17-33.

Berding N, Roach BT (1987). Germplasm collection, maintenance, and use. *In*: Heinz,DJ (ed) Developments in Crop Science II: Sugarcane Improvement Through Breeding. Elsevier Press, Amsterdam, pp 143-210.

Berthomieu P, Meyer C (1991). Direct amplification of plant genomic DNA from leaf and root pieces using PCR. *Plant Mol Biol* **17**: 555-557.

Besse P, McIntyre CL, Berding N (1997). Characterisation of *Erianthus* sect. *Ripidium* and *Saccharum* germplasm (Andropogoneae – Saccharinae) using RFLP markers. *Euphytica* **93**: 293-292.

Bhat SR, Gill BS (1985). The implication of 2n egg gametes in nobilisation and breeding of sugarcane. *Euphytica* **34**: 377-384.

Blose MJ (1992). Estimation of genetic parameters of agronomic traits in selected genotypes of the South African sugarcane breeding population. M Sc thesis, University of Natal, Pietermaritzburg.

Bogani P, Simoni A, Lio P, Scialpi A, Buiatti M (1996). Genome flux in tomato cell clones cultured in vitro in different physiological equilibria.II. A RAPD analysis of variability. *Genome* **39**: 846-853.

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.

Bor NL (1960). The Grasses of Burma, Ceylon , India, and Pakistan. Pergammon, London, pp 767.

Botstein D, White RL, Skolnick MH, Davis RW (1980). Construction of a genetic map in man using restriction fragment length polymorphisms. *Amer J Hum Genet* **32**: 314-331.

Brady JL, Scott NS, Thomas MR (1996). DNA typing of hops (*Humulus lupulus*) through application of RAPD and microsatellite marker sequences converted to sequence tagged sites (STS). *Euphytica* **91**: 277-284.

Brandes EW (1929). Into primeval Papua New Guinea by seaplane. *Nat Geogr Mag* **56(3):** 253-332.

Brandes EW (1958). Origin, classification, and characteristics. *In*: Artschwager,E, Brandes,EW, (eds). Sugarcane (*Saccharum officinarum* L.). *US Dep Agric Agri Handbk* **122**: 1-35, 260-262.

Brandes EW, Sartoris GB, Grassl CO (1939). Assembling and evaluating wild forms of sugarcane and related plants. *Proc Int Soc Sugar Cane Technol* **6**: 128-154.

Bremer G (1923). A cytological investigation of some species and species hybrids within the genus *Saccharum*. *Genetica* **5**: 97-148, 273-326.

Bremer G (1961). Problems in breeding and cytology of sugar cane. *Euphytica* **10**: 59-78.

Bridges Jr WC, Sobral BWS (1996). Validation strategies for QTL mapping. *In*: Sobral,BWS (ed). The Impact of Plant Molecular Genetics. Birkhauser, Boston, pp 39-48.

Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC (1997). Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Sci* **37**: 370-378.

Bubeck DM, Goodman MM, Beavis WD, Grant D (1993). Quantitave loci controlling resistance to grey leaf spot in maize. *Crop Sci* **33**: 838-847.

Burner DM, Legendre BL (1993). Chromosome transmission and meiotic stability of sugarcane (*Saccharum* spp.) hybrid derivatives. *Crop Sci* **33**: 600-606.

Burnquist WL (1991). Development and application of restriction fragment length polmorphism technology in sugarcane (*Saccharum* spp.) breeding. PhD thesis, Cornell University, Ithaca.

Burow MD, Simpson CE, Paterson AH, Starr JL (1996). Identification of peanut (*Arachis hypogaea* L.) RAPD markers diagnostic of root-knot nematode (*Meloidogyne arenaria* (Neal) Chitwood) resistance. *Mol Breeding* **2**: 369-379.

Burstin J, de Vienne D, Dubreuil P, Damerval C (1994). Molecular markers and protein quantities as genetic descriptors in maize. I. Genetic diversity among 21 inbred lines. *Theor Appl Genet* **89**: 943-950.

Byrne PF, McMullen MD, Snook ME, Musket TA, Theuri JM, Widstrom NW, Wiseman BR, Coe EH (1996). Quantitative trait loci and metabolic pathways: Genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc. Natl. Acad. Sci. USA* **93**: 8820-8825.

Caetano-Anolles G, Bassam GJ, Gresshof PM (1991). DNA amplification fingerprinting: A strategy for genome analysis. *Plant Mol. Biol. Reporter* **4**: 294-307.

Cervera TC, Gusmao J, Steenackers M, Van Gysel M, Van Montagu M, Boerjan W (1996). Application of AFLP™-based molecular markers to breeding of *Populus* spp. *Plant Growth Regulation* **20**: 47-52.

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD (1989). RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. *Theor Appl Genet* **78**: 495-504.

Charcosset A, Gallais A (1996). Estimation of the contribution of quantitative trait loci (QTL) to the variance of a quantitative trait by means of genetic markers. *Theor Appl Genet* **93**: 1193-1201.

Charters YM, Robertson A, Wilkinson MJ, Ramsay G (1996). PCR analysis of oilseed rape cultivars *Brassica napus* L ssp. *oleifera*) using 5'-anchored simple sequence repeat (SSR) primers. *Theor Appl Genet* **92**: 442-447.

Cheung WY, Chao S, Gale MD (1991). Long-range physical mapping of the α-amylase-1 (α-Amy-1) loci on homoeologous group 6 chromosomes of wheat. *Mol Gen Genet* **229**: 373-379.

Chyi Y.-S, Hoenecke ME, Sernyk JL (1992). A genetic linkage map of restriction fragment length polymorphism loci for *Brassica rapa* (syn. *campestris*). *Genome* **35**: 746-757.

Clancy JA, Jitkov VA, Han F, Ullrich SE (1996). Barley tissue as direct template for PCR: a practical breeding tool. *Mol Breeding* **2**: 181-183.

Coe E, Hancock D, Kowalski S, Polacco M (1995). Gene list and working maps. *Maize Genet Coop Newslt* **69**: 191-267.

Daniels J (1973). Origin of *S. officinarum* and *S. robustum*. *Sugarcane Breed Newslett* **32**: 4-13.

Daniels J, Paton N, Smith P, Williams CA (1980). Further studies on leaf flavonoids as evolutionary indicators in *Saccharum officinarum* L. *Proc Int Soc Sugar Cane Technol* **17**: 1317-1335.

Daniels J, Roach BT (1987). Taxonomy and evolution. *In*: Heinz,DJ (ed). Developments in Crop Science II: Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 7-84.

Darvasi A, Soller M (1992). Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor Appl Genet* **85**: 353-359.

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**: 943-951.

Da Silva JAG (1993). A methodology for genome mapping of autopolyploids and its application to sugarcane (*Saccharum spp.*). PhD thesis. Cornell University, Ithaca.

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

Da Silva JAG, Sobral BWS (1996). Genetics of polyploids. *In*: Sobral,BWS (ed). The Impact of Plant Molecular Genetics. Birkhauser, Boston, pp 3-37.

Da Silva JAG, Sorrels ME, Burnquist W, 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. *Theor Appl Genet* **92**: 1059-1064.

Davis TM, Yu H, Haigis KM, McGowan PJ (1995). Template mixing: a method of enhancing detection in interpretation of codominant markers. *Theor Appl Genet* **91**: 582-588.

De Oliveira AC, Richter T, Bennetzen JL (1996). Regional and racial specificities in sorghum germplasm asssessed with DNA markers. *Genome* **39**: 579-587.

Dedryver F, Jubier M-F, Thouvenin J, Goyeau H (1996). Molecular markers linked to the leaf rust resistance gene *Lr24* in different wheat cultivars. *Genome* **39**: 830-835.

Dellaporta SL, Wood J, Hicks JB (1983). A plant DNA minipreparation: version II. *Plant Mol Biol Rep* **1**: 10-21.

Demeke T, Laroche A, Gaudet DA (1996). A DNA marker for the Bt-10 common bunt resistance gene in wheat. *Genome* **39**: 51-55.

Devey ME, Delfino-Mix A, Kinloch BB, Neale DB (1995). Random amplified polymorphic DNA markers linked to a gene for resistance to white pine blister rust in sugar pine. *Proc. Natl. Acad. Sci. USA* **92**: 2066-2070.

Devos KM, Gale MD (1992). The use of random amplified polymorphic DNA in wheat. *Theor Appl Genet* **84**: 567-572.

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. *Mol Gen Genet* **250**: 405-413.

D'Hont A, Lu YH, Feldmann P, Glaszmann JC (1993). Cytoplasmic diversity in sugar cane revealed by heterologous probes. *Sugar Cane* **1**: 12-15.

D'Hont A, Lu YH, De Leon DG, Grivet L, Feldmann P, LanaudC, Glaszmann JC (1994). A molecular approach to unraveling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. *Genome* **37(2)**: 222-230.

D'Hont A, Rao PS, Feldmann P, Grivet L, Islam-Faridi N, Taylor P, Glaszmann JC (1995). Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum* X *Erianthus arundinaceus*, with molecular markers and DNA in situ hybridisation. *Theor Appl Genet* **91**: 320-326 .

Dib C, Faure' S, Fizames C, Samson D, Drouot N, Vignal A, Millaseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach (1996). A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* **380**: 152-154.

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ, Evans CA, DeAngelis MM, Levinson DM, Kruglyak L, Goodman N, Copeland NG, Jenkins NA, Hawkins TL, Stein L, Page DC, Lander ES (1996). A comprehensive genetic map of the mouse genome. *Nature* **380**: 149-152.

Doerge RW, Rebaï A (1996). Significance thresholds for QTL interval mapping tests. *Heredity* **76**: 459-464.

Doerge RW, Zeng Z-B, Weir BS (1997). Statistical issues in the search for genes affecting quantitative traits in populations. *Stat Sci* **12**: 195-219.

Dos Santos JB, Nienhuis J, Skroch P, Tivang J, Slocum MK (1994). Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica olarecea* L. genotypes. *Theor Appl Genet* **87**: 909-915.

Dufour P, Deu M, Grivet L, D'Hont A, Paulet F, Bouet A, Lanaud C, Glaszmann JC, Hamon P (1997). Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. *Theor Appl Genet* **94**: 409-418.

Dupuis J (1994). Statistical problems associated with mapping complex and quantitative traits from genomic mismatch scanning data. PhD thesis, Stanford University, California.

Dvorak J, Chen KC (1984). Distribution of non-structural variation between wheat cultivars along chromosome arm 6Bp: evidence from the linkage map and physical map of the arm. *Genetics* **106**: 325-333.

Echt CS, Erdahl LA, McCoy TJ (1992). Genetic segregation of random amplified polymorphic DNA in diploid cultivated alfalfa. *Genome* **35**: 84-87.

Echt CS, May-Marquardt P, Hseih M, Zahorchak R (1996). Characterization of microsatellite markers in eastern white pine. *Genome* **39**: 1102-1108.

Edwards M (1992). Use of molecular markers in the evaluation and introgression of genetic diversity for quantitative traits. *Field Crops Res* **29**: 241-260.

Edwards M, Johnson L (1994). RFLPs for rapid recurrent selection. In: Analysis of Molecular Marker Data. Am Soc Hort Sci and Crop Sci Soc Am, Madison, WI, pp 33-40.

Edwards MD, Page NJ (1994). Evaluation of marker-assisted selection through computer simulation. *Theor Appl Genet* **88**: 376-382.

Eksomtramage T, Paulet F, Noyer JL, Feldmann P, Glaszmann JC (1992). Utility of isozyme in sugar cane breeding. *Sugar Cane* **3**: 14-29.

Falconer DS, Mackay TFC (1997). Introduction to quantitative genetics (4th edition). Longman, London, pp100-204.

Faris JD, Anderson JA, Francl LJ, Jordahl JG (1997). RFLP mapping of resistance to chlorosis induction by *Pyrenophora tritici-repentis* in wheat. *Theor Appl Genet* **94**: 98-103.

Feingold E, Brown PO, Siegmund D (1993). Gaussian models for genetic linkage analysis using complete high-resolution map of identity by descent. *Am J Hum Genet* **53**: 234-251.

Feldmann P (1984). Analyse du polymorphisme enzymatique de la canne á sucre (*Saccharum* sp.): utilisation pour la recherche de variabilité aprés culture in vitro. Thése, Université de Paris-sud, Orsay, Paris.

Fisch RD, Ragot M, Gay G (1996). A generalization of the mixture model in the mapping of quantitative trait loci for progeny from a biparental cross of inbred lines. *Genetics* **143**:571-577.

Fitch MM, Moore PH (1983). Haploid production from anther culture of *Saccharum spontaneum* L.. *Z. Pflanzenphysiol Bd S* 197-206.

Foisset N, Delourme R, Barret P, Hubert N, Landry BS, Renard M (1996). Molecular-mapping analysis in *Brassica napus* using isozyme, RAPD and RFLP markers on a doubled-haploid progeny. *Theor Appl Genet* **93**: 1017-1025.

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

Ganal MW, Young ND, Tanksley SD (1989). Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the *Tm-2a* region of chromosome 9 in tomato. *Mol Gen Genet* **215**: 395-400.

Gebhardt C, Ritter E, Barone A, Debener T, Walkemeier B, Schactshabel U, Kaufmann H, Thompson RD, Bonierbale MW, Ganal MW, Tanksley SD, Salamini F (1991). RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theor Appl Genet* **83**: 49-57.

Genstat (1993). Lawes Agricultural Trust , Rothamstead Experimental Station, U.K.

Gill BS, Grassl CO (1986). Pathways of genetic transfer in intergeneric hybrids of sugarcane. *Sugar Cane* **2**: 2-7.

Gill KS, Gill BS (1994). Mapping in the realm of polyploidy: the wheat model. *BioEssays* **16**: 841-846.

Gimelfarb A, Lande R (1994). Simulation of marker assisted selection in hybrid populations. *Genet Res Camb* **63**: 39-47.

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD (1991). Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. *Nucl Acids Res* **19**: 6553-6558.

Glaszmann JC, Fautret A, Noyer JL, Feldmann P, Lanaud C (1989). Biochemical markers in sugarcane. *Theor Appl Genet* **78**: 537-543.

Göcmen B, Jermstad KD, Neale DB, Kaya Z (1996). Development of random amplified polymorphic DNA markers for genetic mapping in Pacific yew (*Taxus brevifolia*). *Can J For Res* **26**: 497-503.

Grassl CO (1946). *Saccharum robustum* and other wild relatives of "noble" sugar canes. *J Arnold Arbor Harv Univ* **27**: 234-252.

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, Glaszman JC (1996). RFLP mapping in cultivated sugarcane (*Saccharum* spp): Genome organisation in a highly polyploid and aneuploid interspecific hybrid. *Genetics* **142 (3)**: 987-1000.

Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchel-Olds T, Sherman B, Vujcic S, Williams C, Neale D (1994). Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* **138**: 1293-1300.

Haldane JBS (1919). The combination of linkage values and the calculation of distance between the loci of linked factors. *J Genet* **8**: 299-309.

Haley SD, Afanador L, Kelly JD (1994). Selection for monogenic pest resistance traits with coupling and repulsion phase RAPD markers. *Cell Biol Mol Genet Crop Sci* **34**: 1061-1066.

Haley CS, Knott SA (1992). A simple regression method for mapping quantitative traits in line crosses using flanking markers. *Heredity* **69**: 315-324.

Hallauer AR, Miranda JB (1988). Quantitative genetics in maize breeding. Iowa State University Press, Ames, Iowa, pp 52, 64, 150-153.

Hallden C, Hansen M, Nilsson N.-O, Hjerdin A, Sall T (1996). Competition as a source of errors in RAPD analysis. *Theor Appl Genet* **93**: 1185-1192.

Hämäläinen JH, Watanabe KN, Valkonen JPT, Arihara A, Plaisted RL, Pehu E, Miller L, Slack SA (1997). Mapping and marker-assisted selection for a gene for extreme resistance to potato virus Y. *Theor Appl Genet* **94**: 192-197.

Hamby RK, Zimmer EA (1988). Ribososmal RNA sequences for inferring phylogeny within the grass family (*Poaceae*). *Plant Syst Evol* **160**: 29-37.

Harlan JR, de Wet JMJ (1972). A simplified classification of cultivated sorghum. *Crop Sci* **12**: 172-176.

Hart GE (1983). Genetics, and evolution of multilocus isozymes in hexaploid wheat. *In*: Isozymes: Current Topics in Biological Research. Alan R. Liss, New York, p365.

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, D'Hont A, Alix K, Huckett BI (1997). Use of PCR-based markers for identification of *Erianthus* genetic material in putative intergeneric hybrids (*Saccharum* x *Erianthus*). (in press).

Harvey M, Huckett BI, Botha FC (1994). Use of the polymerase chain reaction and random amplification of polymorphic DNAs (RAPDs) for the deterrmination of genetic distances between 21 sugarcane varieties. *Proc. S.A. Sugarc Technol. Assn* **68**:36-40.

Heinze B, Geburek TH (1995). Searching for DNA markers linked to leaf colour in copper beech *Fagus sylvatica* L. var. atropunicea. *Silvae Genetica* **44**: 5-6.

Heun M, Murphy JP, Philips TD (1994). A comparison of RAPD and isozyme analyses for determining the genetic relationships among *Avena sterilis* L. accessions. *Theor Appl Genet* **87**: 689-696.

Hofmeyr JDJ (1939). Sex-linked inheritance in *Carica papaya* L. *S Afr J Sci* **35**: 300-304.

Hogarth DM (1987). Genetics of sugarcane. *In*: Heinz,DJ (ed). Developments in Crop Science II – Sugarcane Improvement through Breeding. Elsevier, Amsterdam, pp 255-271.

Honeycutt RJ, Sobral BWS, Keim P, Irvine JE (1992). A rapid DNA extraction method for sugarcane and its relatives. *Plant Mol Biol Rep* **10**: 66-72.

Hormaza JI, Dollo L, Polito VS (1994). Identification of a RAPD marker linked to sex determination in *Pistacia vera* using bulked segregant analysis. *Theor Appl Genet* **89**: 9-13.

Hu J, Quiros CF (1991). Identification of broccoli and cauliflower cultivars with RAPD markers. *Plant Cell Reports* **10**: 505-511.

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. *Proc natl Acad Sci USA* **87**: 4251-4255.

Inman-Bamber NG (1996). Factors affecting dry matter and sucrose content of sugarcane. *Proc S Afr Sug Technol Ass* **70**: 37.

Iqbal MJ, Rayburn AL (1994). Stability of RAPD markers for dtermining cultivar specific DNA profiles in rye (*Secale cereale* L.). *Euphytica* **75**: 215-220.

James NI (1971). Yield components in random and selected sugarcane populations. *Crop Sci* **11**: 906-908.

James NI, Falgout RN (1969). Association of five characters in progenies of four sugarcane crosses. *Crop Sci* **9**: 88-91.

Jansen RC (1992). A general mixture model for mapping quantitative trait loci by using molecular markers. A general mixture model for mapping quantitative trait loci by using molecular markers. *Theor Appl Genet* **85**: 252-260.

Jansen RC (1993). Interval mapping of multiple quantitative trait loci. *Genetics* **135**: 205-211.

Jansen RC (1996). A general Monte Carlo method for mapping multiple quantitative trait loci. *Genetics* **142**: 305-311.

Jansen RC, Stam P (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447-1455.

Jarvis P, Lister C, Szabo V, Dean C (1994). Integration of CAPs markers into the RFLP map generated using recombinant lines of *Arabidopsis thaliana*. *Plant Mol Biol* **24**: 685-687.

Jeswiet J (1925). Beschrijving der Soorten van het Suikerriet. 11de Bijdrage. Bijdrage tot de Systematiek van het Geslacht *Saccharum*. *Arch Suikerind Ned Indie, Meded Proefsta Java Suiker Ind* **33(3)**: 391-404.

Johnston SA, den Nijs TPM, PeloquinSJ, Hanneman Jr. RE (1980). The significance of genic balance to endosperm development in interspecific crosses. *Theor Appl Genet* **57**: 5-9.

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R,

Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997). Reproducibility testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories. *Mol Breeding* **3**: 381-390.

Jourdren C, Barret P, Horvais R, Delourme R, Renard M (1996a). Identification of RAPD markers linked to linolenic acid genes in rapeseed. *Euphytica* **90**: 351-357.

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

Kandasami PA, Alexander KC, Santhakumariamma PN, Radhakrishnan CM, Palanichamy K, Ramana Rao TC (1980). Inheritance of smut resistance in sugarcane. *Indian J. Agric. Sci.* **50**: 659-663.

Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L, Ferreira A, Webb DM (1997). A high-density soybean genetic map based on AFLP markers. *Crop Sci* **37**: 537-543.

Kesseli R, Paran I, Michelmore RW (1992). Efficient mapping of specifically targeted genomic regions and the tagging of these regions with reliable PCR-based markers. *In*: Applications of RAPD Technology to Plant Bacteria. Am Soc Hort Sci, Alexandria, VA, pp 31-36.

Kesseli R, Witsenboer H, Stanghellini M, Vandermark G, Michelmore R (1993). Recessive resistance to *Plasmopara lactucae-radicis* maps by bulked segregant analysis to a cluster of dominant disease resistance genes in lettuce. *Mol Plant-Microbes Interactions*l **6 (6)**: 722-728.

Klimyuk VI, Carroll BJ, Thomas CM, Jones JDG (1993). Alkali treatment for rapid preparation of plant material for reliable PCR analysis. *Plant J* **3**: 493-494.

Knapp SJ (1991). Using molecular markers to map multiple quantitative trait loci: Models for backcross, recombinant inbred, and doubled haploid progeny. *Theor Appl Genet* **81**: 333-338.

Kosambi DD (1944). The estimation of map distances from recombination values. *Ann Eugenics* **12**: 172-175.

Kowalski SP, Lan TH, Feldmann KA, Paterson AH (1994). Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveal islands of conserved organisation. *Genetics* **138**: 499-510.

Kubisiak TI, Nelson CD, Nance WL, Stine M (1995). RAPD linkage mapping in a longleaf pine x slash pine F-1 family. *Theor Appl Genet* **90**: 119-1127

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin Y-H, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang Z-X, Momma T, Umehara Y, Yano M, Sasaki T, Minobe

Y (1994). A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genetics* **8**: 365-372.

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.

Lagercrantz U, Ellegren H, Andersson L (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucl Acids Res* **21**: 1111-1115.

Lamboy WF (1994a). Computing genetic similarity coefficients from RAPD data: the effects of PCR artifacts. *PCR Methods Appl* **4**: 31-37.

Lamboy WF (1994b). Computing genetic similarity coefficients from RAPD data: correcting for the effects of PCR artifacts caused by variations in experimental conditions. *PCR Methods Appl* **4**: 38-43.

Lanaud C, Risterucci AM, Ngoran AKJ, Clement D, Flament MH, Laurent V, Falque M (1995). A genetic linkage map of *Theobroma cacao* L. *Theor Appl Genet* **91**: 987-993.

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.

Langridge U, Schwall M, Langridge P (1991). Squashes of plant tissue as substrate for PCR. *Nuc Acids Res* **19**: 6954.

Lassner MW, Peterson P, Yoder JI (1989). Simultaneous amplification of multiple DNA fragments by polymerase chain reaction in the analysis of transgenic plants and their progeny. *Plant Mol Biol Rep* **7**: 116-128.

Leberg PL (1992). Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution* **46**: 477-494.

Lee M (1995). DNA markers and plant breeding programs. *Advances in Agronomy* **55**: 265-344.

Lee SH, Bailey MA, Mian MAR, Shipe ER, Ashley DA, Parrot WA, Hussey RS, Boerma HR (1996). Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. *Theor Appl Genet* **92**: 516-523.

Lehner A, Campbell MA, Wheeler NC, Pöykkö T, Glössl J, Kreike J, Neale DB (1995). Identification of a RAPD marker linked to the pendula gene in Norway spruce (*Picea abies* (L.) Karst. F. *pendula*). *Theor Appl Genet* **91**: 1092-1094.

Lennox CG (1939). Sugarcane collection in New Guinea during 1937. *Proc Int Soc Sugar Cane Technol* **6**: 171-182.

Leyser HMO, Lincoln CA, Timpte C, Lammer D, Turner J, Estelle M (1993). *Arabidopsis* auxin-resistance gene *AXR1* encodes a protein related to ubiquitin-activating enzyme *E1*. *Nature* **364**: 161-164.

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

Lin J.-Z, Ritland K (1996b). The effects of selective genotyping on estimates of proportion of recombination between linked quantitative trait loci. *Theor Appl Genet* **93**: 1261-1266.

Lin YR, Schertz KF, Paterson AH (1995). Comparative analysis of QTLs affecting plant height and maturity across the P*oaceae*, in reference to an interspecific sorghum population. *Genetics* **141**: 391-411.

Linc G, Karsai I, Molnar-Lang M, Bedo Z (1996). Comparison of RFLP probes and RAPD primers for studying genetic diversity in barley (*Hordeum vulgaris* L.). *Cereal Research Communications* **24(3)**: 283-290.

Liu S-C, Kowalski SP, Lan T-H, Feldmann KA, Paterson AH (1995). Genome wide high-resolution mapping by recurent intermating using *Arabidopsis thaliana* as a model. *Genetics* **142**??

Livak,KJ, Flood,JA, Marmaro,J, Giusti,W, and Deetz,K (1995). *PCR Methods Appl.* **4**: 357-362.

Lu YH, D'Hont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC (1994b). 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 (1994a). Relationships among ancestral species of sugarcane revealed with RFLP using single-copy maize nuclear probes. *Euphytica* **78**: 7-18.

Lu J, Knox MR, Ambrose MJ, Brown JKM, Ellis THN (1996). Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. *Theor Appl Genet* **93**: 1103-1111.

Maisonneuve B, BellecY, Anderson P, Michelmore RW (1994). Rapid mapping of two genes for resistance to downy mildew from *Lactuca serriola* to existing clusters of resistance genes. *Theor Appl Genet* **89**: 96-104.

Manly KF, Elliot RW (1991). MapManager, a microcomputer program for analysis of data from recombinant inbred strains. *Mammalian Genome* **1**: 123-126.

Mano Y, Takahashi H, Sato K, Takeda K (1996). Mapping genes for callus growth and shoot regeneration in barley (*Hordeum vulgare* L.). *Breeding Sci* **46**: 137-142.

Markert Cl, Moller F (1959). Multiple forms of enzymes: Tissues, ontogenetic, and species specific patterns. *Proc Natl Acad Sci USA* **45**: 753-763.

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. *Proc Natl Acad Sci USA* **88**: 2336-2340.

Martinez O, Curnow RN (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor Appl Genet* **85**: 480-488.

Massal E, Barrau J (1956). Food plants of the South Sea Islands. *South Pac Comm Tech Paper* **94**: 51.

Mather K (1938). Measurement of linkage in heredity. Methuen and Co., London.

Mather DE, Tinker NA, LaBerge DE, Edney M, Jones BL, Rossnagel BG, Legge WG, Briggs KG, Irvine RB, Falk DE, Kasha KJ (1997). Regions of the genome that affect grain and malt quality in a North American two-row barley cross. *Crop Sci* **37**: 544-554.

Mather K, Jinks JL (1974). Biometrical Genetics. Chapman and Hall Ltd, London.

Maughan PJ, Maroof MAS, Buss GR, Huestis GM (1996). Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near isogenic line analysis. *Theor Appl Genet* **93**: 392-401.

McEarchen MJ, Blackburn EH (1994). A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. *Proc Natl Acad Sci USA* **91**: 3453-3457.

McGrath JM, Jancso MM, Pichersky E (1993). Duplicate sequences with a similarity to expressed genes in the genome of *Arabidopsis thaliana*. *Theor Appl Genet* **86**: 880-888.

McGrath JM, Quiros CE (1991). Inheritance of isozyme and RFLP markers in *Brassica campestris* and the comparison with *B. oleracea*. *Theor Appl Genet* **82**: 668-673.

Melchinger AE (1990). Use of molecular markers in breeding for oligogenic disease resistance. *Plant Breed* **104**: 1-9.

Mendell NR, Thode HC, Finch SJ (1991). The likelihood ratio test for the two-component normal mixture problem: powder and sample size analysis. *Biometrics* **47**: 1143-1148.

Menkir A, Goldsbrough P, Ejeta G (1997). RAPD based assessment of genetic diversity in cultivated races of sorghum. *Crop Sci* **37**: 564-569.

Messmer MM, Melchinger AE, Woodman WL, Lee EA, Lamkey KR (1991). Genetic diversity among progenies and elite lines from Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. *Theor Appl Genet* **83**: 97-107.

Michelmore RW, Kesseli RV, Francis DM, Paran I, Fortin MG, Yang CH (1992). Strategies for cloning plant disease resistance genes. *In*: Gurr,S, McPherson,MJ, and Bowles,DJ (eds). Molecular Plant Pathology 2 - A practical approach. Oxford University Press, pp 233-288.

Michelmore RW, Paran I, Kesseli RV (1991). Identification of markers linked to disease resistance genes by bulk segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. *Proc Natl Acad Sci USA* **88**: 9828-9832.

Miklas PN, Stavely JR, Kelly JD (1993). Identification and potential use of a molecular marker for rust resistance in common bean. *Theor Appl Genet* **85**: 745-749.

Milligan SB, Gravois KA, Bischoff KP, Martin FA (1990). Crop effects on broad sense heritabilities and genetic variances of sugarcane yield components. *Crop Sci* **30**: 344-349.

Moore PH, Irvine JE (1991). Genomic mapping of sugarcane and its potential contribution to improvement and to selection of new varieties. *Proc S Afr Sug Technol Ass* **65**: 96-102.

Moore PH, Wu KK (1991). Sugarcane genome mapping: prospects for variety improvement. *Proc Hawaiian Sugar Technol*, pp 34-41.

Morchen M, Cuguen J, Michaelis G, Hanni C, Saumitou-Laprade (1996). Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris L. Theor Appl Genet* **92**: 326-333.

Morden WC, Doebley J, Schertz KF (1989). Allozyme variation in old world races of *Sorghum bicolor* (*Poaceae*). *Am J Bot* **76**: 245-255.

Morden WC, Doebley J, Schertz KF (1990). Allozyme variation among spontaneous species of Sorghum section Sorghum. *Theor Appl Genet* **80**: 296-304.

Moreno-Gonzalez J (1992a). Estimates of marker-associated QTL in Monte Carlo backcross generations using multiple generations. *Theor Appl Genet* **85**: 423-434.

Moreno-Gonzalez J (1992b). Genetic models to estimate additive and non-additive effects of marker-associated QTLs using multiple regression techniques. *Theor Appl Genet* **85**: 435-444.

Moser H, Lee M (1994). RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats. *Theor Appl Genet* **87**: 947-956.

Mosteller F, Tukey JW (1977). Data analysis and regression. Addison-Wesley Publishing Company, London, pp280-296.

Msomi N, Botha FC (1994). Identification of molecular markers linked to fibre using bulk segregant analysis. *Proc South African Sugar Tech Assn* **68**: 41-45.

Mukherjee SK (1954). Revision of the genus *Saccharum* Linn. *Bull Bot Soc Bengal* **8**: 143-148.

Mukherjee SK (1957). Origin and distribution of *Saccharum*. *Bot Gaz* **119**: 55-61.

Muranty H (1996). Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Heredity* **76**: 156-165.

Nair MK, Ratnambal MJ (1967). Pachytene analysis in *Narenga Sclerostachya* hybrid. *Proc Int Soc Sugar Technol* **12**: 875-877.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**: 3321-3323.

Nei M, Li WH (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76**: 5269-5273.

Nevo E, Zohary D, Brown AHD, Haber M (1979). Genetic diversity and environmental associations of wild barley *Hordeum spontaneum* in Israel. *Evolution* **33**: 815-833.

Nodari RO, Tsai SM, Gilbertson RL, Gepts P (1993). Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. *Theor Appl Genet* **85**: 513-520.

Novy RG, Kobak C, Goffreda J, Vorsa N (1994). RAPDs identify varietal misclassification and regional divergence in cranberry (*Vaccinium macrocarpon* Ait.). *Theor Appl Genet* **88**: 1004-1010.

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. *Theor Appl Genet* **92**: 840-849.

O'Brien SJ (1990). Genetic maps: Locus maps of complex genomes. Cold Spring Harbour Laboratory Press, New York.

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

Oropeza M, de Garcia E (1997). The use of molecular markers for identification of sugarcane varieties (Saccharum spp). *Phyton Int J Exp Bot Arg* **61** (1-2): 81-85.

Ortiz A, Renaud R, Calzada I, Ritter E (1997). Analysis of plum cultivars with RAPD markers. *J Hort Sci* **72**: 1-9.

Panje RR, Babu CN (1960). Studies in *Saccharum spontaneum* L. distribution and geographical association of chromosome numbers. *Cytologia* **25**: 152-172.

Paran I, Kesseli R, Michelmore R (1991). Identification of RFLP and RAPD markers linked to downy mildew resistance genes in lettuce using near-isogenic lines. *Genome* **34**: 1021-1027.

Paran I, Michelmore RW (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor Appl Genet* **85**: 985-993.

Paterson AH (1997). Comparative mapping of plant phenotypes. *In*: Janick,J (ed). Plant Breeding Reviews, Vol 14. John Wiley & Sons Inc, pp 13-37.

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander SE, 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 a complete map of restriction fragment length polymorphisms. *Nature* **335**: 721-726.

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995a). Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *Science* **269**: 1714-1718.

Paterson AH, Schertz KF, Lin Y.-R, Liu S.-C, Chang Y.-L (1995b). The weediness of wild plants: Molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense*(L.) Pers. *Proc Natl Acad Sci USA* **92**: 6127-6131.

Paterson AH, Wing RA (1993). Genome mapping in plants. *Current Opinion in Biotechnology* **4**: 142-147.

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, Laroche A, Scoles G, Molnar SJ, Fedak G (1993a). Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods and Applications* **2**: 341-345.

Penner,GA, Chong,J, Levesque-Lemay,M, Molnar,SJ, and Fedak,G (1993b). Identification of a RAPD marker linked to the oat stem rust gene *Pg3*. *Theor Appl Genet* **85**: 702-705.

Perry BC, Mahinur SA, Arvind AB, Lavi U, Jiang R (1994). Length polymorphisms of simple sequence repeat (SSR) DNA as molecular markers in plants. *In*: Gresshof PM (ed). Plant Genome Analysis. CRC Press, Boca Raton, London, pp 47-56.

Pillen K, Steinrücken G, Wricke G, Herrmann RG, Jung C (1993). An extended linkage map of sugar beet (*Beta vulgaris* L.) including nine putative lethal genes and the restorer gene X. *Plant Breeding* **111**: 265-272.

Pineda O, Bonierbale MW, Plaisted RL, Brodie BB, Tanksley SD (1993). Identification of RFLP markers linked to the *H1* gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Genome* **36**: 152-156.

Plaschke J, Ganal MW, Röder MS (1995). Detection of genetic diversity in closely related bread wheat using microsatellite markers. *Theor Appl Genet* **91**: 1001-1007.

Plomion C, Durel CE (1996). Estimation of the average effects of specific alleles detected by the pseudo-testcross QTL mapping strategy. *Genet Sel Evol* **28**: 223-235.

Plomion C, O'Malley DM, Durel CE (1995). Genomic analysis in maritime pine (*Pinus pinaster*). Comparison of two RAPD maps using selfed and open-pollinated seeds of the same individual. *Theor Appl Genet* **90**: 1028-1034.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breeding* **2**: 225-238.

Price S (1963). Cytogenetics of modern sugar canes. *Econ Bot* **17**: 97-106.

Price S (1965). Cytology of *Saccharum robustum* and related sympatric species and natural hybrids. US Dep Agric, Agric Res Serv, *Tech Bull* **1337**: 47.

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.

Quiros CF, This P, Laudie M, Benet A, Chevre A-M, Delseny M (1995). Analysis of a set of RAPD markers by hybridisation and sequencing in *Brassica*: a note of caution. *Plant Cell Reports* **14**: 630-634.

Raghavan TS, Govindwasamy S (1956). Sugarcane as a material for genetic studies. *Proc Int Soc Sug Technol* **9**: 677-694.

Rebaï A, Goffinet B (1993). Power of tests for QTL detection using replicated progenies derived from a diallel cross. *Theor Appl Genet* **86**: 1014-1022.

Rebaï A, Goffinet B, Mangin B (1994). Approximate thresholds of interval mapping tests for QTL detection. *Genetics* **138**: 235-240.

Reimers JF, Hall P, Hogarth DM (1982). The relationship between Fiji disease susceptibility and yield. *Proc Aust Soc Sugar Cane Technol*, pp103-110.

Reiter RS, Williams JGK, Feifmann KA, Rafalski JA, Tingey SV, Scolnick PA (1992). Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc Natl Acad Sci USA* **89**: 1477-1481.

Rick CM, Forbes JF (1975). Allozyme variation in the cultiavted tomato and closely related species. *Bul Torrey Bot Club* **102**: 376-384.

Roach BT (1972). Chromosome numbers in *Saccharum edule*. *Cytologia* **37**: 155-161.

Röder MS, Plaschke J, König SU, Börner A, Sorrels ME, Tanksley SD, Ganal MW (1995). Abundance and variability and chromosomal location of microsatellites in wheat. *Mol Gen Genet* **246**: 327-333.

Ronald PC, Albano B, Tabien R, Abenes L, Wu KS, McCouch S, Tanksley SD (1992). Genetic and physical analysis of the rice bacterial blight disease resistance locus, *Xa21*. *Mol Gen Genet* **236**: 113-120.

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

Sanchez de la Hoz MP, Davila JA, Loarce Y, Ferrer E (1996). Simple sequence repeat primers used in polymerase chain reaction amplifications to study genetic diversity in barley. *Genome* **39**: 112-117.

Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996). A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* **144**: 805-816.

Sax K (1923). The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**: 552-560.

Schachermayr GM, Messmer MM, Feuillet C, Winzeler H, Winzeler M, Keller B (1995). Identification of molecular markers linked to the *Agropyron elongatum*-derived leaf rust resistance gene *Lr24* in wheat. *Theor Appl Genet* **90**: 982-990.

Schachermayr GM, Siedler H, Gale MD, Winzeler H, Winzeler M, Keller B (1994). Identification and localization of molecular markers linked to the *Lr9* leaf rust resistance gene of wheat. *Theor Appl Genet* **88**: 110-115.

Schüller C, Backes G, Fischbeck G, Jahoor A (1992). RFLP markers to identify the alleles on the *Mla* locus conferring powdery mildew resistance in barley. *Theor Appl Genet* **84**: 330-338.

Schulz B, Westphal L, Wricke G (1994). Linkage groups of isozyme, RFLP and RAPD markers in carrot (*Daucus carota* L. ***sativa***). *Euphytica* **74**: 67-76.

Senior ML, Chin ECL, Lee M, Smith JSC, Stuber CW (1996). Simple sequence repeat markers developed from maize sequences found in the GENBANK database: Map construction. *Crop Sci* **36**: 1676-1683.

Senior ML, Heun M (1993). Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. *Genome* **36**: 884-889.

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* **0**: 1-9.

Simmonds NW (1959). Archery in South East Asia and the Pacific. *J Malays Branch R Asiat Soc* **32 (1)**: 67-104.

Simmonds NW (1976). Sugarcanes. *In*: Simmonds,NW (ed). Evolution of crop plants. Longman, London, pp104-108.

Simmonds NW (1981). Genotype (G), Environment (E) and GE components of crop yields. *Experimental Agriculture* **18**: 355-362.

Siripoonwiwat W, O'Donoughue LS, Wesenberg D, Hoffman DL, Barbosa-Neto JF, Sorrels ME (1997). Chromosomal regions associated with quantitative traits in oat. *Journal of Quantitative Trait Loci* **2 (1)**.

Skinner JC (1956). The inheritance of hairiness in sugarcane. *Proc Int Soc Sugar Cane Technol* **9**: 668-677.

Skinner JC, Hogarth DM, Wu KK (1987). Selection methods, criteria, and indices. *In*: Heinz,DJ (ed). Developments in Crop Science II: Sugarcane Improvement Through Breeding. Elsevier, Amsterdam, pp 409-453.

Slocum MK (1989). Analyzing the genomic structure of *Brassica* species and subspecies using RFLP analysis. *In*: Helentjaris,T, and Burr,B (eds). Development and Application of Molecular Markers to Problems in Plant Genetics. Cold Spring Harbor Laboratory Press, New York, pp 73-80.

Slocum MK, Fidgore SS, Kennard WC, Suzuki JY, Osborn TC (1990). Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. *Theor Appl Genet* **80**: 57-64.

Sobral BWS, Braga DPV, LaHood E, Keim P (1994). Phylogenetic analysis of chloroplast restriction-enzyme site mutations in the Saccharinae: Andropogoneae. *Theor Appl Genet* **87**: 843-853.

Sobral BWS, Honeycutt RJ (1994). Genetics, plants, and the polymerase chain reaction. *In*: Mullis,KB, Ferré,F, Gibbs A (eds). The polymerase chain reaction. Birkhauser, Boston.

Sobral WS, McClelland M (1993). Identification of molecular markers linked to agronomically important traits in sugarcane. *Theor Appl Genet* **86**: 105-112.

Soller M, Genizi A (1978). The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics* **34**: 47-55.

Soller M, Genizi A, Brody T (1976). On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor Appl Genet* **47**: 35-39.

Soltis DE, SoltisPS (1993). Molecular data and the dynamic nature of polyploidy. *Crit Rev Plant Sci* **12**: 243-273.

Sondur SN, Manshardt RM, Stiles JI (1996). A genetic linkage map of papaya based on randomly amplified polymorphic DNA markers. *Theor Appl Genet* **93**: 547-553.

Sourdille P, Baud S, Leroy P (1996). Detection of linkage between RFLP markers and genes affecting anthocyanin pigmentation in maize (*Zea mays* L.). *Euphytica* **91**: 21-30.

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

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

Sreenivasan TV, Sreenivasan J (1984). Cytology of *Saccharum* complex from New Guinea, Indonesia and India. *Caryologia* **37**: 351-357.

Stam P (1993). Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J* **3**: 739-744.

Staub J, Bacher J, Poetter K (1996a). Sources of potential errors in the application of random amplified polymorphic DNAs in cucumber. *Hortscience* **31(2)**: 262-266.

Staub JE, Serquen FC, Gupta M (1996b). Genetic markers, map construction, and their application in plant breeding. *HortScience* **31(5)**: 729-741.

Stevenson GC (1965). Genetics and Breeding of Sugarcane. Longmans, Green and Co, London.

Stiles JI, Lemme C, Sondur S, Morshidi MB, Manshardt R (1993). Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. *Theor Appl Genet* **85**: 697-701.

Stringham GR (1977). Linkage relationships of four seedling mutants in turnip rape. *J Hered* **68**: 391-394.

Stuber CW (1994). Heterosis in plant breeding. *In*: Janick,J (ed). Plant Breeding Reviews. Wiley, New York.

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992). Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* **132**: 823-839.

Stuber CW, Sisco PH (1991). Marker-facilitated transfer of QTL alleles between elite inbred lines and responses in hybrids. *Annu Corn Sorghum Res Conf Proc* **46**: 104-113.

Suzuki DT, Griffiths AJF, Miller JH, Lewontin RC (1989). Linkage 1-Basic eukaryotic chromosome mapping. *In*: An Introduction to Genetic Analysis. W.H. Freeman and Company, New York.

Szewc-McFadden AK, Kresovich S, Bliek SM, Mitchell SE, McFerson JR (1996). Identification of polymorphic, conserved simple sequence repeats (SSR) in cultivated *Brassica* species. *Theor Appl Genet* **93**: 534-538.

Tai PYP, Miller JD (1989). Family perfomance at early stages of selection and frequency of superior clones from crosses among Canal Point cultivars of sugarcane. *Amer Soc Sugar Cane Technol Proc* **9**: 62-70.

Talbert LE, Bruckner PL, Smith LY, Sears R, Martin TJ (1996). Development of PCR markers linked to resistance to wheat streak mosaic virus in wheat. *Theor Appl Genet* **93**: 463-467.

Tanhuanpää PK, Vilki JP, Vilki HJ (1996). A linkage map of spring turnip rape based on RFLP and RAPD markers. *Agric Food Sci Finland* **5**: 209-217.

Tanksley SD, Bernatzky R, Lapitan NL, Prince JP (1988b). Conservation of gene order repertoire but not gene order in pepper and tomato. *Proc Natl Acad Sci USA* **85**: 6419-6423.

Tanksley SD, Fulton TM, McCouch SR (1993). In: O'Brien,SJ (ed). Genetic Maps - Book 6 Plants. Cold Spring Harbor Larboratory Press, New York, pp 6.61-6.74.

Tanksley SD, Ganal W, Martin GB (1995). Chromosome landing: a paradigm for map-based cloning in plants with larger genomes. *Trends Genet* **11**: 63-68.

Tanksley SD, Miller J, Paterson A, Bernatzky R (1988a). Molecular mapping of plant chromosomes. *In*: Gustafson,JP, and Appels,R (eds). Chromosome structure and function: Impact of new concepts. Plenum Press, New York.

Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl Acids Res* **17**: 6463-6471.

Thormann CE, Osborn TC (1992). Use of RAPD and RFLP markers for germplasm evaluation. *In*: Applications of RAPD technology to plant breeding. Proceedings, Joint Plant Breeding Symposia Series, 1992, Minneapolis, Minnesota, pp 9-11.

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

Tinker NA, Mather DE, Rossnagel BG, Kasha KJ, Kleinhofs A, Hayes PM, Falk DE, Ferguson T, Shugar LP, Legge WG, Irvine RB, Choo TM, Briggs KG, Ullrich SE, Francowiak JD, Blake TK, Graf RJ, Dofing SM, Maroof MAS, Scoles GJ, Hoffman D, Dahleen LS, Kilian A, Chen F, Biyashev RM, Kurdna DA, Steffenson BJ (1996). Regions of the genome that affect agronomic perfomance in two-row barley. *Crop Sci* **36**: 1053-1062.

Tivang JG, Nienhuis J, Smith OS (1994). Estimation of sampling variation of molecular-marker data using the bootstrap procedure. *Appl Theor Genet* **89**: 259-264.

Torres AM, Weeden NF, Martin A (1993). Linkage among isozyme, RFLP and RAPD markers in *Vicia faba*. *Theor Appl Genet* **85**: 937-945.

Tsumura Y, Yoshimura K, Tomaru N, Ohba K (1995). Molecular phylogeny of conifers using RFLPs analysis of PCR-amplified specific chloroplast genes. *Theor Appl Genet* **91**: 1222-1236.

Tulsieram LK, Glaubitz JC, Kiss G, Carlson JE (1992). Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. *Bio/Technology Vol. 10*

Tuskan GA, Francis KE, Russ SL, Romme WH, Turner MG (1996). RAPD markers reveal diversity within and among clonal and seedling stands of aspen in Yellowstone National Park, U.S.A. *Can J For Res* **26**: 2088-2098.

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

Urrea CA, Miklas PN, Beaver JS, Riley RH (1996). A co-dominant randomly amplified polymorphic DNA (RAPD) marker useful for indirect selection of bean golden mosaic virus resistance in common bean. *J Amer Soc Hort Sci* **121 (6)**: 1035-1039.

Vallejos CE, Sakiyama NS, Chase CD (1992). A molecular marker-based linkage map of *Phaseolus vulgaris* L. *Genetics* **131**: 733-740.

Van Ooijen JW (1994). DrawMap: A computer program for drawing genetic linkage maps. *Heredity* **85**: 66

Veldboom LR, Lee M (1996). Genetic mapping of quantitative trait loci in maize in stress and nonstress environments: I. Grain yield and yield components. *Crop Sci.* **36**: 1310-1319.

Verhaegen D, Plomion C (1996). Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. *Genome* **39**: 1051-1061.

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

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. *Nucl Acids Res* **23**: 4407-4414.

Waldron JC, Glasziou KT (1972). Isoenzymes as a method of varietal identification in sugarcane. *Proc Int Soc Sugar Cane Technol* **14**: 249-256

Wang G.-L, Paterson AH (1994). Assessment of DNA pooling strategies for mapping of QTLs. *Theor Appl Genet* **88**: 355-361.

Wang G-L, Wing R, Paterson AH (1993). PCR amplification from single seeds, facilitating DNA marker-assisted breeding. *Nucl Acids Res* **21**: 2527.

Warburton ML, Becerra-Velasquez VL, Goffreda JC, Bliss FA (1996). Utility of RAPD markers in identifying genetic linkages to genes of economic interest in peach. *Theor Appl Genet* **93**: 920-925.

Weber J, May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* **44**: 388-396.

Weeden NF, Doyle JJ, Lavin M (1989). Distribution and evolution of a glucosephosphate isomerase duplication in *Leguminosae*. *Evolution* **43**: 1637.

Weeden NF, Timmerman GM, Lu,J (1994). Identifying and mapping genes of economic importance. *Euphytica* **73**: 191-198.

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

Welsh J, McClelland M (1992). PCR-amplified length polymorphisms in tRNA intergenic spacers for categorizing staphylococci. *Mol Microbiol* **6**: 1673-1680.

Whitkus R, Doebley J, Lee M (1992). Comparative genome mapping of sorghum and maize. *Genetics* **132**: 1119-1130.

Williams KJ, Fisher JM, Langridge P (1996). Development of a PCR-based allele-specific assay from an RFLP probe linked to resistance to cereal cyst nematode in wheat. *Genome* **39**: 798-801.

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

Williams JGK, Reiter RS, Young RM, Scolnik PA (1993). Genetic mapping of mutations using phenotypic pools and mapped RAPD markers. *Nucl Acids Res* **21**: 2697-2702.

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

Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD (1992). The detection and estimation of linkage in polyploids using single dose restriction fragments. *Theor Appl Genet* **83**:279-285.

Wu H.-K, Chen T.-C, Chung M.-C (1996). Analysis of 5' region of glutelin genes from wild rye species. *Bot Bull Acad Sinica* **37**: 41-49.

Wu KK, Heinz DJ, Meyer HK (1983). Heritability of sugarcane smut resistance and correlation between smut grade and yield components. *Crop Sci* **23**: 54-56.

Wu K, Jones R, Danneberger L, Scolnik PA (1994). Detection of microsatellite polymorphisms without cloning. *Nucl Acids Res* **22**: 3257-3258.

Wu KS, Tanksley SD (1993). Abundance, polymorphism, and genetic mapping of microsatellite loci in rice. *Mol Gen Genet* **241**: 225-235.

Wu DY, Ugozzoli L, Pal BK, Wallace RB (1989). Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. *Proc Natl Acad Sci USA* **86**: 2757-2760.

Xu S (1995). A comment on the simple regression method for interval mapping. *Genetics* **141**: 1657-1659.

Yamagishi M (1995). Detection of section-specific random amplified polymorphic DNA (RAPD) markers in *Lilium*. *Theor Appl Genet* **91**: 830-835.

Yang W, de Oliveira AC, Godwin I, Schertz K, Bennetzen JL (1996). Comparison of DNA marker technologies in characterising plant genome diversity: Variability in Chinese sorghums. *Crop Sci.* **36**:1669-1676.

Young ND (1993). Applications of DNA genetic markers to the study of plant growth and development. *Plant Growth Regulation* **12**: 229-236.

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

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.

Zabeau M, Roberts RJ (1979). The role of restriction endonucleases in molecular genetics. *In*: Taylor,JH (ed) Molecular genetics, vol III. Academic Press, London, pp 1-63.

Zeng ZB (1994). Precision mapping of quantitative trait loci. *Genetics* **136**: 1457-1468.

Zhang YX, Gentzbittel L, Vear F, Nicolas P (1995). Assessment of inter- and intra-inbred line variability in sunflower (*Helianthus annuus*) by RFLPs. *Genome* **38**: 1040-1048.

Zietkiewcz E, Rafalski A, Labuda D (1994). Genome fingerprinting by simple sequence repeats (SSR)-anchored PCR amplification. *Genomics* **20**: 176-183.

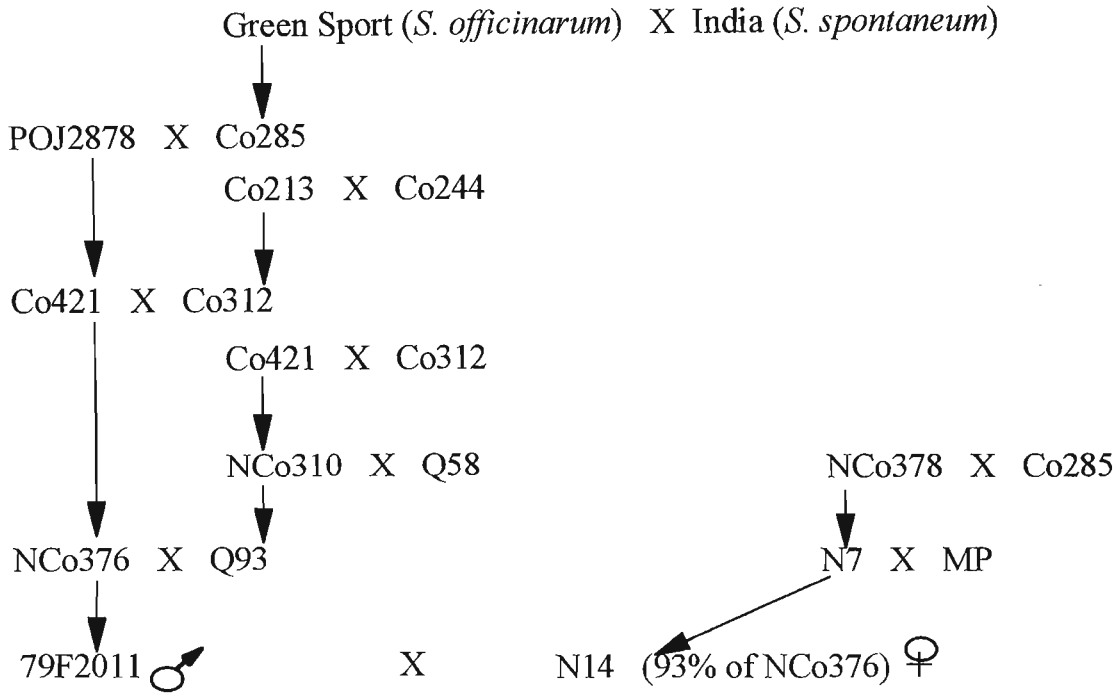
Appendix A

The parents of the six populations.

1. Cross AA40: N18 x CP57/614
2. Cross AA157: 79F2011 x N14
3. Cross AA1292: NCo376 x 75F2297
4. Cross AA1312: CP56/59 x Co1001
5. Cross Z927: MZC74/275 x 77F0790
6. Cross Z586: 81W0050 x N14

The genealogy of cross AA157

(From SASEX Plant Breeding records)[1]



Another source of *S. spontaneum* genome in the genealogy of cross AA157 is POJ2878, which has Java (*S. spontaneum*) as one of its progenitors. However, the total complement of the *S.*

[1]Plant Breeding Department, SASEX, Mount Edgecombe, 4300

spontaneum is low because of the 2n + n transmission (Simmonds, 1976). The NCo varieties are derived from Coimbatore (India), but raised in Natal (South Africa). The Q varieties are from Queensland (Australia), while the Co varieties are from Coimbatore (India) (Stevenson, 1965, Berding and Roach, 1987). The MP (= melting pot) refers to an unknown mixture of varieties used as parents. NCo 376 and NCo 310 are used as standard varieties in the SASEX breeding programme, such that high fibre or sucrose is relative to these two varieties (Blose, 1992).

Appendix B

Methods of trait analysis (Anon., 1985)

Brix % cane = mass of all soluble matter as a percentage of fresh mass of cane (brix content or refractometer solids)

= (b / 100) [100 x (C + W)/ C – 1.25F] where,

b = brix % extract

= refractometer reading as adjusted according to calibration tables

C = mass of cane sample

W = mass of cane added to the digester bowl

F = fibre % cane as described below

Sample mass is standardised to 1 kg and water quantity to 2 kg :

Brix % cane = b (3 – 0.0125F). These standard quantities are used in the following formulae.

Fibre % cane = mass of water insoluble dry matter as percentage of mass of cane (fibre content).

= [100C – C x M – (C + W) b] + C (1 – 0.0125b) where

M = moisture % cane

Pol % cane = mass of sucrose as percentage of mass of fresh cane

= 100 x (mass of pol in cane) + mass of cane.

After standardisation the formula becomes,

Pol % cane = p (3 – 0.0125F) where,

p = pol % extract

= (normal mass x saccharimeter reading) + (mass in g of 100 cm^3 of

solution) where

normal mass = 26.000 g when the saccharimeter is fitted with the International Sugar Scale

F = fibre % cane

Purity = (ratio of brix % cane to pol % cane) x 100

Ers % cane = mass of estimated recoverable sugar as percentage of mass of fresh cane (sugar content).

Appendix C

Derivation of the equation for contamination levels in RAPD bulks

Example A: Total number of loci amplified (T) = 10

Total number of individuals in the bulks (N) = 5

Aberrant locus (A) = 1

Total number of loci in the bulks = N(T-A) + A

At time = 0, there is x amount of primer :

At time = 1 (after one round of amplification), there is x – 45 amount of primer, 90 copies of all the other loci, and two copies of the aberrant locus.

At time = 2 : x – 90 primer, 180 loci, and 4 aberrant loci

At time = 3 : x – 180 primer, 360 loci, and 8 aberrant loci.

$$\% \text{ Contamination at } t = 0 : \quad \frac{1}{(5 \times 9) + 1} \times 100\%$$

$$= \quad 1/46 \times 100\%$$

$$= \quad 2.2\%$$

$$\% \text{ Contamination at } t = 1 : \quad \frac{4}{180 + 4} \times 100\%$$

$$= \quad 4/184 \times 100\%$$

$$= \quad 2.2\%$$

At time = 2, 3,...k, the proportion of aberrant locus to other loci remains the same the same but the primer concentrations dwindle.

Total number of loci (T) at time = 0 refers to the RAPD bands observed in the gel, each of which may be in higher dosage at the start of the reaction, so that it is always equal to or above the number of observed bands.

Rationale : If at the beginning of the RAPD reaction there is an x amount of primers, there is equal competition for all 10 loci. This is based on the assumption that there is 100% homology between the primer sequences at both the 3' and 5' end binding sites. In addition, all loci are in equal dosage at the beginning of the RAPD reaction. Other factors are taken to be equal. At the beginning of the RAPD reaction, the primer will bind to all 10 loci. However the actual number or proportion of the other non-bulk specific loci is effectively multiplied by the number of individuals in the bulk. The proportion of the contamination by the aberrant locus is therefore one divided by the total number of individuals in the bulks (N) multiplied by the total number of all the other loci (T-A) plus the contaminating locus (A) itself.

Effectively the contamination levels given refer to the initial state. At the second round of amplification, there are two copies of the contaminant against 45 (x2) copies of the other loci. Since the PCR amplification is an exponential process, the proportion of aberrant loci against all the other loci should remain the same. However the competition factor due to the dwindling amounts of primer concentrations, and the increase in amplified loci means that eventually the aberrant locus is likely to be diluted out.

Consequences

A. Factors like steric hindrance will favour the loci that are in higher proportion, and this coupled with the primer competition implies that eventually the amplification product of the rare allele will be diluted out.

B. In cases where the dosage of the aberrant individual is higher, and other factors like less than 100% homology of primer binding sites of the other competing loci, the competition factor in the subsequent rounds of amplification may be in favour of the aberrant individual resulting in its amplification. Locus specificity in amplification therefore is probably a result of the change in the contamination percentage during the subsequent rounds of amplification in favour of the aberrant locus.

Appendix D

Determination of clonal repeatabilities for the four traits (1993 data in three replicates)

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
1	1	9.6	9.2	5.7	3.4
2	1	12.3	7.7	4.2	1.9
3	1	9.9	8.1	4.4	2.1
4	1	11.3	8.0	4.2	1.8
5	1	10.3	7.2	3.3	0.8
6	1	13.0	7.7	4.0	1.5
7	1	10.2	7.2	3.9	1.8
8	1	11.5	8.0	4.4	1.9
9	1	11.0	9.4	6.0	3.7
10	1	14.2	7.9	5.0	2.8
11	1	11.2	7.4	3.9	1.5
12	1	11.7	8.0	4.9	2.7
13	1	11.6	7.7	4.4	2.2
14	1	11.0	8.6	4.9	2.4
15	1	7.7	7.6	4.3	2.3
16	1	11.7	8.0	4.5	2.2
17	1	11.0	7.4	3.9	1.6
18	1	11.1	8.3	4.8	2.5
19	1	11.4	8.3	4.7	2.3
20	1	11.4	7.4	4.2	1.9
21	1	8.5	7.5	3.8	1.5
22	1	9.9	7.5	4.3	2.2
23	1	11.2	8.9	4.4	1.6
24	1	11.0	8.3	3.4	0.4
25	1	11.0	9.2	5.8	3.5
26	1	10.3	9.2	5.8	3.5
27	1	11.7	7.4	4.6	2.6
28	1	10.9	8.6	5.3	3.0
29	1	10.1	7.2	3.5	1.1
30	1	11.1	7.2	3.6	1.3
31	1	8.5	8.1	3.8	1.2
32	1	11.5	8.3	4.0	1.3
33	1	12.0	7.4	3.0	0.2
34	1	8.1	7.5	3.6	1.3
35	1	8.3	7.0	2.6	0.1
36	1	10.8	8.0	4.7	2.4
37	1	11.4	7.1	3.6	1.2
38	1	11.2	7.7	4.2	1.8

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
39	1	11.7	8.8	5.5	3.2
40	1	10.1	8.9	5.9	3.8
41	1	8.3	8.7	5.0	2.7
42	1	9.6	8.6	5.4	3.3
43	1	10.0	8.0	4.2	1.8
44	1	11.8	6.8	3.7	1.5
45	1	11.9	8.3	5.1	2.9
46	1	10.2	8.3	4.7	2.4
47	1	9.1	7.5	3.8	1.5
48	1	11.0	7.2	3.7	1.4
49	1	11.7	7.7	4.2	1.9
50	1	12.8	7.4	3.4	0.7
51	1	7.7	6.1	2.5	0.4
52	1	9.6	8.1	4.3	1.9
53	1	9.5	7.2	4.2	2.2
54	1	11.8	7.7	4.4	2.2
55	1	10.3	7.2	4.1	2.0
56	1	9.1	8.7	5.5	3.5
57	1	15.2	8.1	4.7	2.2
58	1	9.6	6.6	3.1	0.9
59	1	12.2	9.1	5.5	3.1
60	1	12.1	5.7	2.1	-0.4
61	1	12.8	7.4	3.5	0.9
62	1	12.0	7.7	3.9	1.4
63	1	10.1	8.3	4.7	2.3
64	1	14.8	7.0	4.3	2.1
65	1	10.6	8.9	5.3	2.9
66	1	12.5	9.1	5.6	3.3
67	1	13.4	9.1	4.9	2.1
68	1	13.9	7.3	4.3	2.0
69	1	8.2	10.1	6.4	4.1
70	1	13.1	7.7	4.3	2.0
71	1	14.2	7.9	4.4	1.9
72	1	11.4	8.6	4.9	2.5
73	1	9.6	9.2	5.5	3.1
74	1	13.7	7.9	4.7	2.4
75	1	9.0	8.1	4.9	2.8
1	2	11.2	7.1	3.4	0.9
2	2	14.1	6.2	3.1	0.8
3	2	9.7	7.8	3.3	0.5
4	2	12.7	7.1	3.9	1.6
5	2	8.0	5.8	1.9	-0.4

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
6	2	13.6	7.1	2.9	0.0
7	2	8.5	8.1	4.2	1.9
8	2	11.0	6.9	4.4	2.6
9	2	11.9	8.0	3.2	0.2
10	2	14.2	6.8	4.5	2.6
11	2	11.0	7.7	3.8	1.3
12	2	10.1	7.5	3.8	1.5
13	2	11.6	7.1	4.0	1.8
14	2	12.1	8.3	4.6	2.1
15	2	9.6	6.6	3.2	1.0
16	2	10.1	8.0	2.7	-0.4
17	2	11.4	6.9	3.4	1.1
18	2	11.0	6.3	2.9	0.7
19	2	10.2	6.6	2.8	0.4
20	2	9.2	8.1	4.1	1.6
21	2	7.8	7.0	3.1	0.8
22	2	12.8	7.1	3.5	1.0
23	2	9.7	9.2	4.0	1.0
24	2	10.7	7.2	2.4	-0.6
25	2	10.7	8.3	4.6	2.2
26	2	10.1	8.6	4.4	1.7
27	2	12.0	8.0	4.1	1.5
28	2	10.8	8.3	4.6	2.2
29	2	9.3	6.9	2.6	-0.1
30	2	9.0	6.9	2.5	-0.1
31	2	10.5	6.9	2.5	-0.3
32	2	10.8	9.2	5.0	2.3
33	2	11.5	8.9	5.0	2.5
34	2	8.9	8.4	4.6	2.3
35	2	10.4	7.2	2.4	-0.5
36	2	12.5	8.8	4.5	1.6
37	2	10.3	9.2	4.4	1.4
38	2	11.6	8.0	4.0	1.4
39	2	9.3	9.2	5.4	3.0
40	2	9.7	8.1	3.9	1.4
41	2	8.9	8.4	4.3	1.9
42	2	10.8	8.9	5.3	3.0
43	2	9.6	8.1	3.9	1.4
44	2	11.1	7.4	4.0	1.7
45	2	12.5	9.1	5.6	3.1
46	2	9.2	7.8	3.4	0.7
47	2	10.8	6.6	2.9	0.5

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
48	2	11.3	7.7	3.3	0.5
49	2	13.8	7.9	4.5	2.1
50	2	14.3	7.9	4.1	1.4
51	2	9.1	6.6	2.2	-0.5
52	2	10.7	7.2	3.8	1.6
53	2	10.3	7.2	4.0	1.9
54	2	14.2	7.6	4.1	1.6
55	2	11.3	7.7	3.9	1.4
56	2	8.0	8.1	4.1	1.7
57	2	13.3	8.5	5.5	3.3
58	2	9.2	6.9	2.5	-0.1
59	2	11.1	8.0	4.1	1.6
60	2	13.4	5.9	2.5	0.1
61	2	13.5	7.6	4.1	1.6
62	2	14.7	7.6	3.9	1.2
63	2	8.7	7.5	3.1	0.4
64	2	13.3	7.9	3.7	0.9
65	2	10.1	8.9	5.2	2.8
66	2	9.5	7.5	3.0	0.3
67	2	10.1	8.0	3.9	1.3
68	2	11.8	8.6	4.7	2.2
69	2	10.4	8.6	5.0	2.6
70	2	11.8	7.4	3.7	1.2
71	2	9.9	8.3	4.5	2.1
72	2	12.8	8.2	4.9	2.6
73	2	9.1	8.9	5.5	3.4
74	2	12.7	7.4	3.6	1.0
75	2	8.2	8.1	4.5	2.2
1	3	9.4	8.1	5.0	3.0
2	3	11.8	7.1	3.2	0.6
3	3	10.0	8.3	5.5	3.5
4	3	13.2	7.1	3.6	1.2
5	3	11.0	6.6	3.0	0.7
6	3	14.2	7.1	3.2	0.6
7	3	10.5	7.2	3.3	0.8
8	3	9.9	8.6	4.7	2.2
9	3	10.0	9.8	6.0	3.6
10	3	12.7	7.1	3.8	1.4
11	3	13.1	7.9	3.9	1.2
12	3	9.6	8.1	4.0	1.5
13	3	11.6	8.9	5.2	2.8
14	3	12.2	8.0	4.6	2.3

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
15	3	9.2	8.4	4.7	2.4
16	3	12.6	7.4	3.4	0.7
17	3	12.2	6.8	3.5	1.1
18	3	9.4	6.9	2.7	0.2
19	3	9.9	8.3	4.2	1.6
20	3	9.8	8.1	4.2	1.8
21	3	7.8	7.3	4.0	1.9
22	3	11.4	6.9	3.2	0.8
23	3	12.1	6.8	2.5	-0.4
24	3	12.2	7.4	3.6	1.1
25	3	11.1	8.3	4.2	1.6
26	3	11.3	8.6	5.5	3.3
27	3	10.3	9.2	5.2	2.6
28	3	10.8	8.0	4.6	2.4
29	3	9.5	7.2	3.0	0.4
30	3	11.1	7.4	3.6	1.1
31	3	10.7	7.7	3.7	1.1
32	3	12.6	9.1	5.9	3.6
33	3	11.7	7.7	3.9	1.3
34	3	8.7	7.8	4.1	1.7
35	3	9.8	6.6	2.5	0.0
36	3	11.9	8.0	4.3	1.8
37	3	11.5	7.1	3.5	1.2
38	3	10.8	8.3	4.8	2.5
39	3	12.0	6.8	3.0	0.5
40	3	9.4	7.8	3.8	1.3
41	3	8.6	8.7	4.9	2.5
42	3	10.5	9.2	5.6	3.2
43	3	9.9	7.8	4.3	2.1
44	3	9.4	7.2	2.9	0.3
45	3	12.4	8.8	4.9	2.3
46	3	10.5	7.5	4.2	2.0
47	3	9.6	7.5	3.2	0.5
48	3	12.4	7.4	2.6	-0.4
49	3	10.4	8.6	4.9	2.6
50	3	12.2	8.0	4.2	1.7
51	3	8.9	5.8	2.0	-0.4
52	3	12.0	7.1	3.5	1.0
53	3	9.4	7.5	3.8	1.4
54	3	11.8	7.4	4.0	1.6
55	3	9.8	8.1	4.4	2.0
56	3	7.2	7.0	2.9	0.6

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
57	3	14.8	8.2	4.5	1.9
58	3	10.6	7.2	2.7	-0.1
59	3	11.8	8.3	4.4	1.8
60	3	11.0	6.0	2.4	0.1
61	3	12.8	7.7	3.6	0.8
62	3	12.3	7.7	3.4	0.6
63	3	9.4	7.2	2.8	0.2
64	3	15.0	7.3	3.9	1.3
65	3	6.6	8.8	4.6	2.2
66	3	10.4	7.7	4.6	2.5
67	3	10.1	9.8	5.9	3.5
68	3	10.6	9.7	6.1	3.8
69	3	10.8	8.0	4.4	2.0
70	3	13.9	7.3	3.8	1.3
71	3	10.9	7.4	3.2	0.6
72	3	12.6	7.4	3.9	1.4
73	3	11.2	9.4	6.5	4.4
74	3	12.3	7.1	3.5	1.1
75	3	10.4	7.2	4.8	3.1
76	1	14.0	7.6	4.4	2.0
77	1	11.7	6.8	3.4	1.1
78	1	14.4	7.3	4.0	1.5
79	1	8.5	8.7	5.2	3.0
80	1	13.8	7.6	4.6	2.3
81	1	9.5	8.6	5.6	3.5
82	1	9.9	6.6	3.1	0.9
83	1	11.9	5.7	1.8	-0.8
84	1	13.5	7.9	4.8	2.6
85	1	12.7	8.2	4.6	2.2
86	1	13.7	7.4	4.1	1.8
87	1	9.7	8.1	4.7	2.6
88	1	12.6	9.1	5.5	3.0
89	1	10.2	8.3	5.1	2.9
90	1	9.3	10.4	6.4	3.9
91	1	10.2	8.3	3.8	1.1
92	1	12.4	8.8	4.7	2.0
93	1	8.3	9.0	5.1	2.8
94	1	15.2	10.1	5.8	2.9
95	1	9.6	8.4	4.0	1.3
96	1	10.1	7.8	3.8	1.4
97	1	13.6	8.8	4.1	1.0
98	1	12.0	7.1	4.1	1.9

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
99	1	8.6	8.7	3.3	0.2
100	1	10.4	9.5	5.3	2.7
101	1	10.8	8.9	4.5	1.8
102	1	11.7	8.6	4.9	2.4
103	1	10.0	7.5	3.1	0.4
104	1	11.1	8.9	3.6	0.4
105	1	10.3	8.9	4.5	1.8
106	1	8.0	6.7	2.8	0.5
107	1	13.9	7.1	2.8	-0.1
108	1	8.6	7.2	3.7	1.5
109	1	9.7	6.6	3.4	1.4
110	1	12.0	8.0	4.4	1.9
111	1	13.0	9.1	5.9	3.6
112	1	10.2	8.3	5.1	3.0
114	1	10.0	8.1	5.0	2.9
115	1	11.3	7.1	4.0	1.8
116	1	10.2	8.0	4.4	2.0
117	1	7.1	8.7	4.8	2.5
118	1	10.7	10.6	8.2	6.5
119	1	14.3	9.0	6.8	4.9
120	1	8.6	9.5	6.8	5.0
121	1	9.0	9.2	6.1	4.0
122	1	8.6	7.5	3.9	1.7
123	1	10.8	8.3	4.5	2.0
124	1	10.7	8.3	5.0	2.8
125	1	11.4	8.3	4.6	2.2
126	1	13.3	8.2	4.9	2.5
127	1	10.1	10.3	7.5	5.6
128	1	12.1	7.7	4.6	2.4
129	1	14.9	9.3	6.5	4.4
130	1	9.4	7.2	4.0	1.9
131	1	13.2	8.5	5.6	3.4
132	1	11.9	8.6	5.4	3.1
133	1	14.4	8.5	5.3	3.0
134	1	13.3	7.1	3.8	1.5
135	1	9.8	10.4	6.3	3.7
136	1	11.9	7.7	3.7	1.0
137	1	11.0	8.6	4.6	2.0
138	1	12.3	9.1	5.1	2.4
139	1	11.4	8.9	5.1	2.7
140	1	17.2	8.6	5.0	2.2
141	1	13.6	10.5	5.3	2.0

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
142	1	12.1	8.0	3.9	1.2
143	1	14.8	9.0	4.8	1.9
144	1	12.2	9.4	6.2	3.9
145	1	9.3	9.5	5.9	3.6
146	1	9.6	7.8	4.3	2.1
147	1	10.8	10.0	5.9	3.3
148	1	10.7	8.0	4.4	2.0
149	1	9.7	7.5	3.2	0.6
150	1	10.1	8.9	5.8	3.7
76	2	13.0	7.9	4.1	1.5
77	2	8.1	7.5	2.8	0.0
78	2	11.0	6.9	3.6	1.4
79	2	9.3	9.2	6.1	4.0
80	2	10.3	7.5	4.0	1.7
81	2	10.9	8.0	4.4	2.1
83	2	8.6	7.5	3.5	1.1
84	2	11.0	9.2	5.5	3.1
85	2	9.2	8.4	4.7	2.4
86	2	11.2	8.9	5.7	3.5
87	2	8.2	9.0	5.8	3.9
88	2	10.8	10.3	7.4	5.4
89	2	9.9	7.2	4.3	2.4
90	2	10.3	8.9	5.7	3.6
91	2	10.4	7.5	4.2	2.0
92	2	12.8	8.8	5.7	3.4
93	2	9.3	8.4	5.1	3.0
94	2	14.9	8.2	5.5	3.4
95	2	9.9	7.5	3.5	1.0
96	2	8.6	7.5	3.9	1.7
97	2	9.6	9.5	6.4	4.3
98	2	8.8	7.5	4.8	3.0
99	2	10.9	7.4	4.5	2.5
100	2	10.3	8.6	5.2	3.0
101	2	8.6	8.7	5.5	3.5
102	2	8.4	11.0	8.0	6.0
103	2	9.7	7.5	4.3	2.3
104	2	8.7	8.1	4.6	2.4
105	2	9.9	7.8	3.7	1.2
106	2	7.2	6.7	1.9	-0.9
107	2	10.3	7.8	2.5	-0.7
108	2	8.1	7.5	3.0	0.4

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
109	2	10.7	6.9	2.2	-0.7
110	2	10.2	8.0	3.7	1.1
111	2	16.4	7.5	3.3	0.3
112	2	11.4	7.4	3.8	1.3
113	2	10.8	10.3	6.6	4.3
114	2	9.8	8.6	3.6	0.7
115	2	11.0	7.7	3.9	1.5
116	2	11.2	7.4	3.3	0.6
117	2	8.7	7.5	3.3	0.8
118	2	13.2	9.9	5.3	2.2
119	2	16.1	8.1	3.9	1.0
120	2	11.4	8.0	3.1	0.0
121	2	12.0	8.8	4.4	1.5
122	2	10.9	7.4	2.7	-0.3
123	2	9.9	8.3	3.8	1.1
124	2	10.7	8.0	4.0	1.5
125	2	12.4	8.5	4.3	1.6
127	2	12.8	9.7	5.7	3.0
128	2	9.9	8.3	4.1	1.5
129	2	11.7	10.6	6.6	4.0
130	2	10.3	6.6	3.3	1.1
131	2	9.9	10.1	5.2	2.3
132	2	10.2	8.3	4.5	2.0
133	2	12.8	9.1	5.3	2.7
134	2	10.5	7.5	3.7	1.2
135	2	9.6	10.1	5.9	3.3
136	2	10.9	8.0	4.3	1.8
137	2	10.0	7.2	4.1	2.0
138	2	11.3	8.9	5.3	2.9
139	2	11.5	8.9	5.6	3.3
140	2	13.8	7.6	4.5	2.1
141	2	11.3	9.2	5.2	2.7
142	2	9.7	9.2	5.3	2.9
143	2	13.1	8.2	5.1	2.9
144	2	8.7	9.0	6.0	4.1
145	2	10.7	8.9	5.6	3.5
146	2	9.7	6.6	3.6	1.6
147	2	9.0	9.2	5.6	3.3
148	2	10.9	7.4	4.3	2.1
149	2	10.3	6.6	3.2	0.9
150	2	10.0	8.6	6.1	4.3
76	3	11.1	8.3	4.7	2.4

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
77	3	8.8	6.9	4.1	2.2
78	3	12.4	8.5	4.7	2.1
79	3	6.4	11.1	7.5	5.5
80	3	12.2	7.1	4.4	2.4
81	3	12.0	8.6	5.2	2.9
82	3	10.5	7.7	4.7	2.7
83	3	10.9	7.2	4.2	2.1
84	3	14.1	10.7	8.4	6.4
85	3	11.0	7.4	4.6	2.6
86	3	12.7	8.5	6.1	4.2
87	3	11.2	8.3	5.6	3.7
88	3	13.6	9.6	6.9	4.9
89	3	9.6	7.5	4.7	2.9
90	3	11.5	8.3	5.7	3.8
91	3	9.7	8.1	4.9	2.8
92	3	11.0	8.9	5.9	3.8
93	3	8.7	8.4	5.5	3.6
94	3	14.2	8.2	5.2	2.9
95	3	11.4	7.7	4.3	2.0
97	3	10.8	10.3	6.6	4.1
98	3	10.3	7.8	4.9	3.0
99	3	10.4	8.0	4.9	2.8
100	3	10.8	7.7	4.7	2.7
101	3				
102	3				
103	3	12.3	8.0	5.0	2.9
104	3	11.4	8.9	5.8	3.6
105	3	10.2	8.0	4.2	1.8
106	3	8.5	6.9	3.1	0.8
107	3	14.2	8.2	3.7	0.7
108	3	9.4	7.5	4.4	2.3
109	3	9.7	8.1	4.9	2.9
110	3	13.1	7.9	4.6	2.2
111	3	12.0	9.1	5.7	3.3
112	3	11.1	10.0	5.9	3.3
113	3	13.6	9.6	6.6	4.4
114	3	10.5	8.3	5.4	3.4
115	3	11.1	8.3	4.9	2.6
116	3	12.2	7.7	4.2	1.8
117	3	7.3	9.6	6.2	4.1
118	3	12.2	9.4	6.2	3.9
119	3	15.5	7.6	4.4	1.9

Indv. no	Rep No	fib %cane	brix %cane	pol %cane	ers %cane
120	3	6.3	9.1	5.4	3.2
121	3	10.7	6.9	3.4	1.1
122	3	7.9	8.7	3.7	0.8
123	3	12.2	6.8	5.3	3.8
124	3	11.0	7.7	4.5	2.3
125	3	11.6	8.0	4.5	2.2
126	3	13.8	9.3	6.3	4.0
127	3	10.5	9.8	6.7	4.6
128	3	12.4	6.8	3.8	1.6
129	3	13.8	9.0	6.1	3.8
130	3	8.7	8.1	4.8	2.7
131	3	11.4	8.9	5.4	3.1
132	3	9.2	7.8	5.1	3.3
133	3	12.4	8.8	5.3	2.9
134	3	10.4	7.5	3.9	1.6
135	3	8.4	9.8	5.8	3.3
136	3	10.7	7.7	3.9	1.4
137	3	8.2	8.4	4.6	2.3
138	3	10.5	8.3	4.8	2.5
139	3	11.3	7.7	4.3	2.1
140	3	14.5	7.6	4.7	2.4
141	3	12.3	8.3	5.2	3.0
142	3	9.2	9.2	4.8	2.1
143	3	9.4	9.2	5.6	3.3
144	3	9.0	9.2	6.1	4.1
145	3	11.6	10.6	7.2	4.9
146	3	10.4	7.5	4.7	2.8
147	3	7.5	10.5	6.8	4.5
148	3	10.2	8.3	5.0	2.8
149	3	11.0	6.9	4.5	2.7
150	3	10.7	8.9	6.2	4.3

		fib %cane	brix %cane	pol %cane	ers %cane
Means(1-75)					
Rep1		11.01	7.95	4.40	2.06
Rep2		10.92	7.73	3.85	1.35
Rep3		10.95	7.77	4.03	1.59
Grp1.mean		10.96	7.82	4.09	1.66
Max		14.42	9.20	5.83	3.63
Min		8.06	5.89	2.23	-0.17
Means(76-150)					
Rep1		11.32	8.41	4.77	2.36
Rep2		10.63	8.27	4.56	2.15
Rep3		10.90	8.40	5.16	2.96
Grp2.mean		10.95	8.36	4.83	2.49
Max		15.30	10.09	6.68	4.43
Min		7.70	6.77	2.62	-0.03
Gmean		10.96	8.09	4.46	2.08
Analysis of Var.					
	DF				
Grps	1	0.01	32.96	60.65	77.18
Reps.in.grp1	2	0.14	1.05	5.99	9.74
Reps.in.grp2	2	8.99	0.45	6.95	13.40
Indv.in grp1	74	6.07	1.28	1.60	1.94
Indv.in.grp2	74	7.41	2.11	2.59	3.00
indvs.1+2	148	6.74	1.70	2.10	2.47
Residual	296	1.34	0.40	0.48	0.71
Total	449	3.14	0.90	1.20	1.56
F-value		6.72	1.12	14.55	18.91
DGD		0.85	0.11	0.93	0.95
SD(genetic)		1.34	0.66	0.74	0.77
SD(gen.)%mean		12.3	8.1	16.5	36.9
SD(unit)		1.16	0.63	0.69	0.84
CV%		10.6	7.8	15.5	40.5

Curriculum Vitae

1985	Vukuzakhe High School
1987	International Baccalaureate Diploma at Waterford KaMhlaba United World College of Southern Africa
1988-1991	BSc (Honours) at Sussex University (England)
1994	MSc registration
1995 (June)	MSc registration upgraded to a PhD registration
1997	FRD/Prof Schell Award for Plant Biotechnology