IDENTIFICATION OF MOLECULAR MARKERS LINKED TO FIBRE USING BULK SEGREGANT ANALYSIS

N MSOMI[1,2] AND FC BOTHA[1,2]

[1] South African Sugar Association Experiment Station, Mount Edgecombe
[2] University of Natal

Abstract

The potential use of bulk segregant analysis (BSA) to establish genetic markers for a trait in sugarcane was investigated. Four bulked DNA samples were prepared from a population, AA157, segregating for fibre. The bulks comprised DNA from five and ten individuals taken from the extremes for low and high fibre content. To screen for differences, the bulked DNA samples were amplified using the polymerase chain reaction and arbitrary decamer primers. A total of 749 fragments were amplified in the bulks, eight of which were polymorphic. DNA extracts from 80 individuals from the population were then amplified with the primers which had generated the polymorphisms in the bulked DNA. Six of the polymorphic DNA fragments showed a 1:1 segregation ratio, indicating that they are single dose fragments. The polymorphic fragments were isolated, reamplified and cloned into the pCR-Script SK(+) phagemid for sequencing and subsequent conversion into sequence characterised amplified regions (SCARs).

Introduction

At present, the SASEX sugarcane breeding programme takes an average of 14 years before a new commercial variety can be released. This is due largely to the reliance on morphological markers for selection. These markers are strongly influenced by the environment, which masks the true genetic potential of the plants (Moore and Irvine, 1991). Another contributing factor is the lack of known Mendelian inherited traits in sugarcane (Hogarth, 1987). The selection programme would be greatly improved if molecular markers linked to any trait of interest could be found, since these exist independently of environmental or ontogenic influences. However, the absence of near isogenic lines and the high ploidy levels in sugarcane make the task of finding molecular markers difficult.

Michelmore *et al.* (1991) have described a method that allows quick establishment of genetic markers for any trait of interest, provided well defined segregating populations are available. The parents from which the progeny originate have to be genetically different at the locus of interest. The method involves bulking of the DNA from individuals representing the extremes of the segregating population, and is referred to as a bulk segregant analysis (BSA). The rationale for BSA is that there will be linkage equilibrium for all the genetic regions in the bulks except the target region, which will be in linkage disequilibrium. Therefore the bulked DNA will be genetically dissimilar for the loci controlling the trait but similar for all other genetic regions. This means that any differences that are detected as polymorphisms between the bulks will be linked to the target genetic region. The number of individuals making up a bulk should not be too few as this would increase the chances of scoring false positives (Michelmore *et al.*, 1991). Neither should the bulks be too big, as this would increase the probability of finding individuals with a double crossover at this region, thus reducing the chances of finding a genetic marker (Giovannoni *et al.*, 1991).

Differences in the bulked DNA samples can be analysed by random amplification of polymorphic DNAs (RAPDs) with arbitrary decamer primers (Williams *et al.*, 1990) through the polymerase chain reaction (PCR). Once the RAPD markers have been detected, linkage to the trait can be confirmed by analysing the population, or a random representative sample thereof, from which the bulks were derived (Michelmore *et al.*, 1991). However, RAPD markers tend to be cross specific, as single point mutations in the priming site can completely prevent amplification of the relevant locus. This difficulty can be overcome only if the short primer (10 nucleotides) can be lengthened and made more specific (20-25 nucleotides). This requires cloning and subsequent sequencing of the RAPD fragments, an approach which results in the production of sequence characterised amplified regions (SCARs) (Michelmore *et al.*, 1992).

In this paper the potential use of BSA to establish molecular markers in sugarcane was investigated in a population consisting of 150 F1 individuals from a commercial sugarcane cross. A two point linkage analysis was done with the software program MAPMAKER (Lander *et al.*, 1987) to establish whether any of the markers belonged to the same linkage groups. Preliminary linkage analysis of putative markers to the fibre trait, as well as conversion of these to SCARs, are reported.

Materials and methods

Selection and sampling of plant material

Six independent F1 populations, consisting of 150 individuals each, were analysed for their segregation of brix, fibre, pol and purity. Quantification of these traits was done by millroom analysis, where at least 12 stalks of each clone were collected from the field, crushed and then analysed by various methods (Anon., 1987). The trait showing the best normal distribution was chosen for investigation. The population with the best segregation for the trait was then planted in three replicates. Two measurements of each individual were made a month apart, so that individuals could be ranked according to the mean of the six readings. The highest and the lowest ten individuals were then identified and sampled for the BSA.

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 tissue was homogenised in 40 ml homogenisation buffer (50 mM Trishydroxymethylaminomethane hydrochloride (Tris-HCl), pH 8,0, 5 mM ethylenediaminetetraacetic acid (EDTA), 0,5 mM spermidine, 1% (m/v) polyethylene glycol (8 000), 0,1% (v/v) 2-mercaptoethanol and 0,35 M sucrose) for two minutes, and filtered through two layers of mutton cloth. The filtrate was centrifuged at 5 000 x 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 25 mM EDTA and no polyethylene glycol) and placed on ice. To this solution the

following were added sequentially: 5 M NaCl to a final concentration of 0,7 M, 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 removed by centrifugation at 3 500 x g for 10 minutes at 4°C. The aqueous phase was collected and the protein extraction repeated. The aqueous phase was again collected and the DNA precipitated by addition of an equal volume of isopropanol. Precipitated DNA was then spooled out, dried and resuspended in 1,0 ml of TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA). The DNA was quantified with a Beckman DU 7500 spectrophotometer at 260 nm, and the relative purity assessed by the 260/280 nm ratio. The DNA was then analysed on a 1% (m/v) agarose gel to confirm its concentration and quality. DNA samples were then stored at -20°C.

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 3 μg/l. Bulks of five and 10 individuals from the extremes of the population were then made to a final concentration of 1 μg/l or 3 μg/l. Sixty random decamer primers from Operon Technologies, 20 each of the A, B, and C series, were respectively diluted in water to a final concentration of 6 μM and stored at −20°C.

Random amplification of polymorphic DNAs (RAPDs)

PCR was done as previously described (Al-Janabi *et al.*, 1993). The final reaction volume of 22 μl contained 2 units of AmpliTaq Stoffel fragment, 0,22 μM primer, 0,1 mM of each deoxynucleoside triphosphate (dNTP), 4,0 mM $MgCl_2$, 10 mM KCl, 10 mM Tris-HCl, pH 8,3, 2,2 μg acetylated bovine serum albumin and 7 or 21 ng of template DNA. Each reaction mixture was overlaid with 50 μl mineral oil. Thermal cycling was done in a Hybaid OmniGene thermal cycler, using the following thermal profile: 94° C/3 minutes, 35° C/1 minute and 72° C/2 minutes with a ramp of 2,4° C/sec; then 40 cycles of 94° C/1 minute, 35° C/1 minute and 72° C/2 minutes with a 2,4° C/sec ramp. The final elongation step was at 72° C for 7 minutes.

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 × TBE buffer (45 mM Tris base, 44 mM boric acid and 1 mM EDTA, pH 8). The gel was stained with ethidium bromide at a concentration of 1 μg/ml TBE buffer for 30 minutes with constant shaking, and destained for 40 minutes in fresh TBE buffer. The gel was visualised under ultraviolet light and photographed.

Amplification and cloning of polymorphic DNA

Polymorphic fragments were excised from gels with a Pasteur pipette. They were re-amplified using the PCR protocol, except that the annealing temperature was raised to 38°C

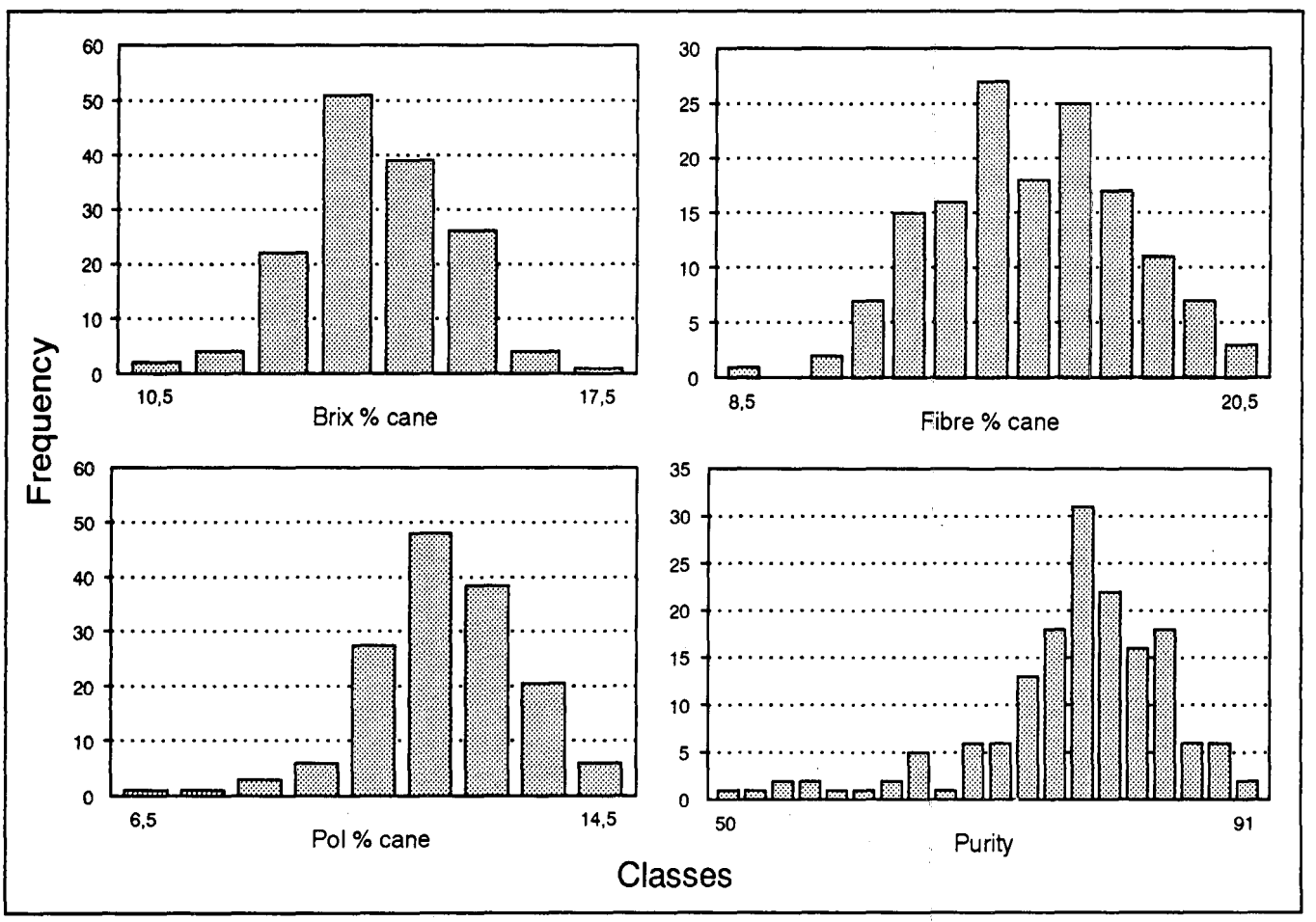

FIGURE 1 The frequency distribution curves for brix, fibre, pol and purity in the F1 population AA157. The range of values for each trait were: brix 14,4 ± 4,45, fibre 15,6 ± 6, pol 12,2 ± 4,15 and purity 84,9 ± 20,35. The phenotypic values for each trait were grouped into classes.

and the magnesium chloride concentration decreased to between 1 and 1,5 mM. The isolated reamplified products were treated with the Klenow fragment to ensure blunt ends, after which they were ligated into the Srf 1 restriction site of the pCR-Script SK(+) phagemid (Anon., 1993). Epicurian *Escherichia coli* cells were transformed and transformants selected (Anon., 1993). White colonies were then grown in liquid LB medium with ampicillin and the plasmid DNA isolated according to the protocol of Sambrook *et al.* (1989). PCR was carried out on the plasmid DNA using the T7 and M13 reverse primers, which have priming sites on either side of the multi-cloning site (MCS). The PCR products were analysed on a 2% (m/v) agarose gel.

Linkage analyses

To determine the segregation ratio of the markers a Chi-square test was done with data from 80 random individuals of population AA157. These data were also used to test for linkage of the markers with the software program MAPMAKER (Lander *et al.*, 1987). Finally, the markers were tested for linkage to fibre by analysing absence or presence of the markers in the top and bottom 20% of individuals in the population.

Results

Population segregation analysis

The frequency distribution curves for all four traits under consideration (brix, fibre, pol and purity) were similar between the six F1 populations. An example of the segregation pattern for traits of population AA157 is presented in Figure 1. In all six populations the distribution curves of pol and purity were skewed. Fibre and brix showed the closest to normal distribution patterns. However, fibre showed the largest range between the low and high extreme individuals. It was therefore decided to use fibre as the target trait and to conduct the BSA on population AA157, which had the largest range among the six populations.

Amplification output and selection of markers

A total of 749 loci were amplified in the AA157 bulks with the 60 random decamer primers. This represents an average of 12 loci amplified per primer (Table 1). There were 79 polymorphic fragments detected between the parents, a ratio of 1,27 polymorphisms per primer. Fragments were scored as polymorphic between the high and low fibre groups based on two criteria. Firstly, fragments were scored solely on the basis of presence or absence in the bulks; for example, OPC11.1 was present in the two low fibre bulks and absent in both high fibre bulks (Table 2). This was done despite the fact that this fragment was present in two individuals of the opposite bulk of five. Five polymorphisms were scored in this manner (0,08 polymorphisms/primer). Secondly, fragments were scored on the basis of their presence in individuals used to construct the bulks, the criterion being that a fragment should be present in four or more individuals on one extreme of the trait, while one aberrant individual could be tolerated in the individuals of the opposite extreme. This criterion was based on the observation that, in some cases, the presence of one aberrant individual resulted in the band being present in the opposite bulk (for example OPB16.1) (Table 2). Three more polymorphisms were scored by using this criterion (0,05 polymorphisms/primer).

Table 1

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

Primers used	60
Loci amplified	749
Polymorphisms detected	8
Polymorphisms/primer	0,13
Parental polymorphisms/primer	1,27
High fibre specific loci	4
Low fibre specific loci	4

Table 2

Distribution of the polymorphisms which were identified as putative markers for fibre in the bulks, in six individuals from which the bulks were derived and in the parents of the cross AA157. The six clones are those ranked 1-5 and 10 for the trait. The fragments were scored as present (+) or absent (−)

Clone	OPA17.1 438 bp	OPB1.1 778 bp	OPB9.1 1832 bp	OPB11.1 464 bp	OPB16.1 618 bp	OPC4.1 321 bp	OPC11.1 352 bp	OPC16.1 819 bp
Low Fibre								
106	+	−	+	+	+	−	+	−
117	+	−	+	+	+	−	+	−
56	−	−	+	+	+	−	+	−
51	+	−	+	−	+	−	+	−
41	+	+	+	+	+	−	+	+
21	+	+	+	+	+	+	+	+
bulk (5)	+	−	+	+	+	−	+	−
bulk (10)	+	−	+	+	+	+	+	+
High Fibre								
140	−	−	−	−	−	+	−	−
119	−	+	−	+	−	+	−	+
94	−	+	−	−	+	−	+	+
57	−	+	−	−	−	+	−	+
64	−	+	+	−	−	+	+	+
74	+	+	+	−	−	−	−	+
Bulk (5)	−	+	−	−	+	+	−	+
Bulk (10)	−	+	+	+	+	+	−	+
N14	−	+	−	+	−	−	−	−
79F2011	+	−	+	−	+	+	+	+

Six of the polymorphic fragments were traced to the male parent, 79F2011, and the remaining two to the female parent, N14 (Table 2).

Marker segregation and linkage

So far, six of the eight RAPD markers identified have been studied for their segregation in 80 individuals of population AA157. Results in Table 3 indicate that all six markers segregate in a 1:1 ratio. This means that they can be mapped into linkage groups with the two point or a multi-point analysis. However, when the analysis was done with the logarithm of odds (lod) score at three and the maximum genetic distance at 50 centimorgans (cM), none of the markers could be linked. When the lod score was reduced to 2,38, linkage was detected but the genetic distance was still 63 cM between markers OPB16.1 and OPC11.1.

Table 3

Segregation of the putative fibre markers within the 80 individuals from the F1 population AA157.

Marker	Polymorphism		Missing data	Parental origin
	Present	Absent		
OPB1.1	34	44	2	N14
OPA17.1	39	41	0	79F2011
OPB16.1	38	36	6	79F2011
OPC11.1	43	30	7	79F2011
OPC4.1	31	41	8	79F2011
OPC16.1	37	41	3	79F2011

To investigate the potential linkage of the polymorphic fragments to fibre, the top and bottom 20% of individuals for the trait were analysed for the presence or absence of each of the putative markers. Strong correlation between the markers and the trait was found, with 75 to 87% of the individuals having the predicted fragment composition.

Conversion of RAPD markers to SCARs

RAPD markers OPB11.1 (464 bp), OPC4.1 (321 bp) and OPC11.1 (352 bp) have been successfully cloned. The PCR products of the cloned markers, when analysed by agarose gel electrophoresis (Figure 2), were found to be about 200 bp longer than the original fragments, as expected from pCR-Script structure, thus confirming that the right size fragments had been inserted.

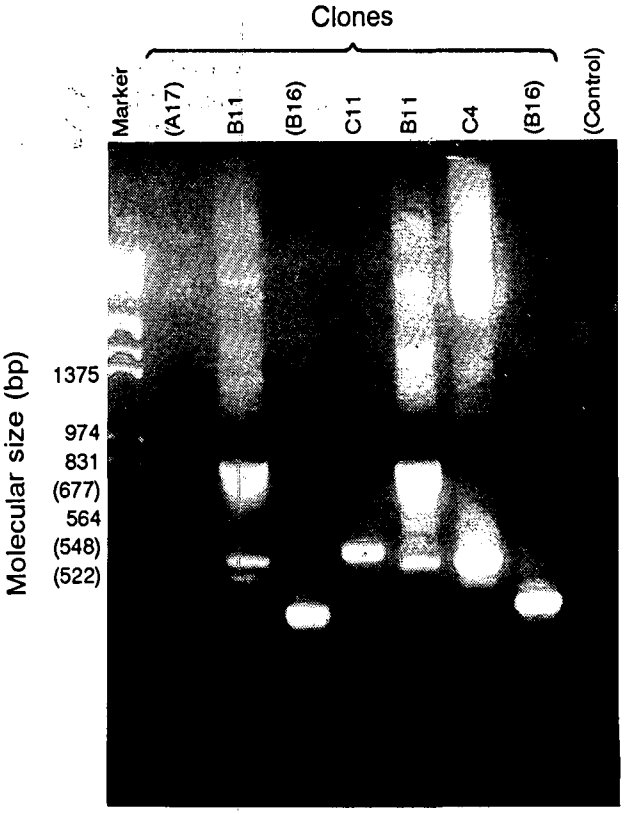

FIGURE 2 Electrophoretic separation on a 2% (m/v) agarose gel of the PCR amplification products from putative transformant clones with the T7 and M13 reverse primers. The clones in brackets indicate either a wrong size insert as in B16 or total failure to transform as in A17.

Discussion

The results obtained in this study indicate that the amplification strength of loci can differ substantially. This is evident from the observation that the inclusion of one individual with a polymorphic fragment in a bulk resulted in the locus being amplified (OPB16.1), whereas for others the inclusion of two aberrant individuals in the bulk did not result in amplification (OPC11.1). For this reason it was decided to adopt a strategy which not only takes into account amplification within the bulks, but also between the individuals from which the DNA samples were pooled. Although only four of the individuals had the fragment, the scoring of fragments as polymorphic meant that the size of the bulks was effectively reduced, and increased the possibility of scoring false positives. However, since the markers have to be confirmed for their linkage to the trait across the whole population, false positives can be identified and removed.

The RAPD results indicated that the genetic variation between the parents was only 10,5%. Ten per cent of these genetic differences (total detected polymorphisms) were potentially associated with fibre, either high or low. The reduction in polymorphisms in the bulks indicate that the technique used in this investigation effectively randomised the genetic background, hence the detected polymorphisms should be potentially linked to the targeted genetic region. The possibility cannot be ruled out, however, that linkage may be to a genetic region controlling a trait that co-segregates with fibre content. For example, in sugarcane it is known that fibre is strongly correlated to sucrose with a value of -0,76 (Hogarth, 1987).

Preliminary results, on the 20% individuals with the highest and lowest fibre contents in the population, indicate a strong possibility (>78%) of linkage of these markers to fibre. However, the next step will be to establish the statistical significance of these putative markers on the whole population to confirm their linkage to fibre. This will allow calculation of how much fibre variation can be ascribed to each of the putative markers.

All the fibre-associated polymorphisms segregated in a 1:1 ratio in the progeny, which is indicative of single dose markers (Wu *et al.*, 1992). At this stage it is unclear whether most polymorphisms detected in a crop like sugarcane, where a minimum genetic variation exists in the germplasm, will represent single dose fragments. However, it is already evident from this study that at least 10% of polymorphisms in the parents of modern commercial crosses have the potential to be mapped. This observation is significant, as many re-

searchers have questioned the possibility of constructing the genetic map of a commercial sugarcane variety.

No significant linkage in either coupling or repulsion was detected between the possible markers for the fibre trait. The absence of linkage between the markers is not unexpected since fibre is a polygenic trait. The markers therefore may be associated with the different loci controlling fibre expression. As is the case with all markers for quantitative trait loci, the most useful ones, from a plant breeding perspective, will be those that are tightly linked to one or a few major genetic regions controlling the trait. However, should fibre expression be controlled by several minor regions, the markers would ascribe more or less the same small percentage of fibre variation. Such markers will be of little value for marker assisted breeding.

RAPD markers are dominant by nature; they are either scored as present or absent (Tingey and del Tufo, 1993). They cannot be used to differentiate between homozygous and heterozygous loci and are thus less informative than restriction length polymorphism (RFLP) markers (Reiter *et al.*, 1992). In addition, as single point mutations could nullify the potential use of a RAPD marker, they could be cross specific (Sobral and McClelland, 1993). This would limit their potential use and prevent wide scale testing of linkage in different crosses. All these limitations to RAPD markers can be circumvented, if the markers are converted to SCARs (Michelmore *et al.*, 1992). By sequencing into the amplified polymorphic region, longer and more specific primers can be designed. Not only will such primers be insensitive to most point mutations but they will differentiate between homozygous and heterozygous loci.

Good progress has already been made towards converting the identified putative markers for fibre to SCARs. The initial phase of cloning the relevant polymorphic fragments has been completed. However, sequencing of these will be delayed until linkage with the trait has been confirmed.

Acknowledgements

We wish to thank the staff of the Plant Breeding and Biometry departments at the Experiment Station for their assistance in the selection of material for this investigation, and Arnold van der Merwe at the Pongola field station for his help in sampling and collection.

REFERENCES

Anon. (1987). Laboratory Manual for South African Sugar Factories. South African Sugar Technologists' Association.

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

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

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

Hogarth, DM (1987). Genetics of sugarcane. In: *Developments in Crop Science 2: sugarcane improvement through breeding.* (Ed) DJ Heinz, Elsevier. pp 255-267.

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

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

Michelmore, RW, Paran, I and 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.

Michelmore, RW, Kesseli, RV, Francis, DM, Paran, I, Fortin, MG and Yang, CH (1992). Strategies for cloning plant disease resistance genes. In: *Molecular Plant Pathology 2: a practical approach.* Oxford, Oxford University Press.

Moore, PH and 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.

Reiter, RS, Williams, JGK, Feldman, KA, Rafalski, JA, Tingey, SV and Scolnik, 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.

Sambrook, J, Fritsch, EF and Maniatis, T (1989). *Molecular Cloning: A laboratory manual.* Cold Spring Harbour. pp 1.25-1.28.

Sobral, WS and McClelland, M (1993). Identification of molecular markers linked to agronomically important traits in sugarcane. *Theor Appl Genet* 86: 105-112.

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

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

Wu, KK, Burnquist, W, Sorrells, ME, Tew, TL, Moore, PH and Tanksley, SD (1992). The detection and estimation of linkage in polyploids using single dose restriction fragments. *Theor Appl Genet* 83: 279-285.