

THE IDENTIFICATION OF POTENTIAL GENETIC MARKERS IN SUGARCANE VARIETIES FOR THE PREDICTION OF DISEASE AND PEST RESISTANCE RATINGS

^{1,2}JM BARNES, ¹RS RUTHERFORD AND ^{1,2}FC BOTHA

¹South African Sugar Association Experiment Station P/Bag X02, Mount Edgecombe, 4300

²University of Natal, PO Box 375, Scottsville, Pietermaritzburg, 3209

Abstract

The use of genetic markers that are linked to specific traits in sugarcane has the potential to increase the efficiency of the selection of improved varieties. Conventionally, markers are identified by analysing the segregation of potential markers and traits in the progeny of single crosses. However, this approach is not practical for sugarcane breeding programmes where replicated, well characterised progenies do not exist. The objective of this project was to investigate the potential of using commercial varieties for identifying markers associated with some of the important traits in sugarcane. This approach would be far more effective than dealing with single progenies since the traits of commercial varieties have already been characterised. The DNA of fifty commercial varieties of sugarcane was amplified by RAPD PCR using 41 arbitrary decamer primers. Analysis of the resulting banding profiles, obtained by agarose gel electrophoresis, yielded 54 reliable polymorphic fragments. Multiple regression analysis was used to identify putative markers associated with the traits of resistance to eldana, sugarcane mosaic virus and smut, and the predictive power of the markers was determined. Further analysis of sugarcane varieties using additional polymorphisms has the potential to identify markers linked to important traits. These markers could be used for marker-assisted selection to increase the efficiency of selecting for improved sugarcane genotypes for commercial release.

Keywords: genetic markers, sugarcane varieties, traits, breeding

Introduction

The sugarcane selection programme at the South African Sugar Association Experiment Station (SASEX) takes about 14 years from the time crosses are made between appropriate parent varieties to the commercial release of a variety. This prolonged time period is largely due to the reliance of the selection process on phenotypic characters, which are strongly influenced by the environmental conditions under which the sugarcane was grown. If genetic markers can be found that are linked to specific traits in sugarcane, the breeding and selection programme would have a means of selecting for traits at the genetic level, in the absence

of environmental influence. The accurate, early selection of traits would mean that the unnecessary carriage of undesirable clonal genotypes to successive stages of the selection programme would be minimised.

Sugarcane is a genetically complex crop, and modern varieties are polyploid interspecific hybrids. This complexity has limited the identification of genetic markers linked to particular genes in comparison to diploid crop species. However, in the past few years, the theory and practice of mapping and molecular marker technology has progressed to the stage where molecular marker maps of the ancestral species of sugarcane (Al-Janabi *et al.*, 1994; Da Silva *et al.*, 1995) and commercial sugarcane varieties (D'Hont *et al.*, 1994; Grivet *et al.*, 1996) have been produced with relative ease. This has led to various expectations for marker technology with considerable potential value for the improvement of this crop. One of these is the identification of markers that are tightly linked to traits of interest.

Studies on genetic linkage between markers and particular genes are invariably based on populations, derived from planned crosses, that segregate for the trait of interest. However, due to environmental influences on sugarcane, extensive field trials over several years are required before the traits of such a progeny can be accurately evaluated in this crop. The general lack of well characterised segregating populations in sugarcane means that the identification of markers linked to large numbers of agronomically important traits using conventional linkage analysis is somewhat limited.

The genetic base of modern sugarcane varieties is relatively narrow in view of the fact that early breeders used only a few of the available clones of mainly *Saccharum officinarum* and *Saccharum spontaneum* in the initial crosses, while further variety improvement involved the intercrossing of varieties produced from this process. The narrow genetic background has prompted the suggestion that modern sugarcane varieties can essentially be considered as a large progeny derived from a cross between *S. spontaneum* and *S. officinarum* (Lu *et al.*, 1994a, 1994b). The analysis of a population of modern varieties for genetic markers that are linked to various traits is a far more practical approach than the analysis of a true progeny, since the traits of varieties have already been characterised during the selection programme.

The random amplification of polymorphic DNA (RAPD technique) is a PCR-based technique that uses random decamer primers to amplify polymorphic genomic profiles of a wide variety of species (Welsh and McClelland, 1990; Williams *et al.*, 1990). DNA fragments amplified in this manner can be used as genetic markers, and consequently this approach has been used extensively for the identification of markers linked to traits in many plant species (Martin *et al.*, 1991; Haley *et al.*, 1994; Poulsen *et al.*, 1995; Ohmori *et al.*, 1996). Furthermore, RAPD analysis has been optimised for use in sugarcane (Sobral and Honeycutt, 1993). Since one of the major limitations of sugar production in South Africa is the susceptibility of available sugarcane varieties to pathogens (fungal, bacterial and viral) and pests, markers for disease resistance would be particularly useful. Resistance to the sugarcane borer *Eldana saccharina*, sugarcane mosaic virus (SCMV), and the fungal pathogen smut (*Ustilago scitamineae*) were therefore selected for this study.

The objective of this study was to investigate the potential of identifying genetic markers associated with particular traits in sugarcane, using a collection of 50 varieties rather than a progeny derived from a specific cross. The RAPD technique was used to amplify DNA polymorphisms in the varieties, and multiple regression analysis was used to establish any relationships between the polymorphisms and the traits, and to determine the predictive potential of the putative markers.

Materials and methods

Plant material and trait data

Fifty sugarcane varieties were analysed in this study. These are either varieties that have been commercially released, or varieties that were considered unsuitable for commercial cultivation but are nevertheless used as parents in the sugarcane breeding programme at SASEX. The eldana, SCMV and smut resistance ratings are readily available for these varieties, and these range from 1 to 9, where 1 signifies a level of resistance such that signs of infection or damage are either very limited or absent altogether, and 9 signifies extreme susceptibility with signs of severe infection or damage. All plants were harvested from the germplasm collection grown at SASEX.

DNA extraction

DNA was extracted from the sugarcane using a protocol modified from Honeycutt *et al.* (1992). It has been found that the DNA from at least three separate plants for each variety must be pooled to eliminate individual-specific differences in the banding profiles (Harvey *et al.*, 1994). As a result, leaf roll tissue from three individual plants was pooled (6-10 g in total) and homogenised for two minutes (Ultra-Turrax T-25) in 40 mL ice-cold homogenisation buffer (50 mM Tris(hydroxymethyl)-aminomethane (Tris), pH 8,0, 5 mM ethylene diamine tetra-acetic acid (EDTA), 0,5 mM spermidine, 1% (w/v) polyethylene glycol (8 000), 0,1% (v/v) 2-mercaptoethanol, 0,35 M sucrose). The homogenate was filtered through four layers

of mutton cloth and the filtrate centrifuged at 5 000 g for 20 minutes in a rotor precooled to 4°C. The supernatant was discarded and the cell pellet resuspended in 10 mL cold wash buffer (50 mM Tris, pH 8,0, 25 mM EDTA, 0,5 mM spermidine, 0,1% (v/v) 2-mercaptoethanol, 0,35 M sucrose) and placed on ice. To this suspension were added, sequentially, NaCl to a final concentration of 0,7 M, sodium dodecyl sulphate (SDS) to a final concentration of 0,7% (w/v), and cetyltrimethylammonium bromide (CTAB) to a final concentration of 0,9% (w/v), with gentle agitation after each addition. The solutions were incubated at 60°C for 30-45 minutes and then allowed to cool at room temperature for 15 minutes. Proteins were removed by gentle mixing with an equal volume of chloroform:isoamyl alcohol (24:1), centrifugation of the emulsion at 3 500 g for 10 minutes at 4°C, collection of the aqueous phase and repetition of the extraction. After the final collection of the aqueous phase, DNA was precipitated by the addition of an equal volume of isopropanol, spooled out of solution with a glass hook, drained, and dissolved in 1 mL TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8,0).

DNA assessment

The DNA was quantified and its quality assessed spectrophotometrically (260/280 nm). In addition, aliquots of the DNA samples were electrophoresed on 1% (w/v) agarose gels at 5,6 V/cm. The gels were stained with ethidium bromide (1 µg/mL), destained and the DNA concentrations confirmed by comparison of the intensity of the band staining with that of known standards (uncut Lambda DNA, Boehringer Mannheim). Dilutions of the concentrated stocks were prepared for use in the PCR reactions (3 ng/µL), while the main stock solutions were stored at -20°C.

RAPD amplification

The DNA samples were PCR amplified in the presence of random decamer primers with a G-C content of 60-70% (Operon Technologies). The final volume of the PCR reaction mixtures was 21 µL, and these contained 10 mM Tris-HCl (pH 8,3), 10 mM KCl, 4 mM MgCl₂, 1,4 µg acetylated bovine serum albumin, 0,2 µM primer, 0,1 mM of each of dATP, dTTP, dCTP, dGTP (Boehringer Mannheim), 1,4 U AmpliTaq Stoffel fragment (Perkin-Elmer) and 21 ng template DNA. Reaction mixtures were overlaid with 30 µL mineral oil prior to thermal cycling. PCR amplification was carried out in a Hybaid OmniGene Temperature Cycler using the following cycling profile: one cycle of three minutes at 94°C, one minute at 35°C and two minutes at 72°C with a 2,4°C/sec ramp; 40 cycles of one minute at 94°C, one minute at 35°C and two minutes at 72°C with a 2,4°C/sec ramp; and one cycle of seven minutes at 72°C.

Amplification fragments were separated on 2% (w/v) agarose gels run at 5,6 V/cm in 0,5 X TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8,0). Gels were stained in ethidium bromide (1 µg/mL) for 45 minutes followed by destaining for 30 minutes, both with constant shaking. Amplification profiles

were viewed under ultraviolet light. Only clear, distinct polymorphic bands were considered for analysis and the gels were scored visually by noting the presence (1) or absence (0) of these bands across all the varieties.

Data analysis

Associations between the RAPD polymorphisms and the resistance ratings for the three traits were established using the multiple regression approach. Each trait was treated as the dependent variable and the various RAPD polymorphisms as independent variables, and the analysis was based on the following model:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_px_p$$

which related the variation in the dependent variable (Y , representing the resistance ratings) to a function of the set of independent variables (x_1 to x_p , representing the polymorphisms). The b_1 to b_p terms are the regression coefficients, calculated by regression analysis, that specify the relationships between Y and x . The equation intercept is a and the number of polymorphisms is p . Stepwise regression was performed on the polymorphism data and the most significant model for each trait, determined according to the r^2 and F -ratio values of the regression, was selected. To test the predictive potential of these models, they were used in a 'leave out one at a time' approach where multiple regression analysis was applied to 49 of the 50 varieties at a time, with the prediction of the fiftieth. This allowed for the fitting of 50 models for each trait, and is the equivalent of using the chosen polymorphisms in a regression model calibrated from the 50 varieties to predict 50 unknown varieties.

Results

RAPD output

A total of 41 random primers were screened for their ability to detect reliable polymorphisms in the sugarcane genomic DNA. The results are summarised in Table 1. RAPD amplification of the DNA samples with each primer was done at least twice to test the reproducibility of the amplified loci, and this resulted in

eight of the primers being excluded from further analysis due to the generation of inconsistent banding profiles. The remaining 33 primers amplified a total of 382 loci with an average of 11,6 loci per primer. Only clear, distinct polymorphisms were considered; bands that appeared fuzzy and bands that were obscured by the co-migration of bands of similar size were disregarded. Using this approach, 54 reproducible polymorphic loci were detected from 26 of the primers (average of 2,1 polymorphisms per primer). These fragments ranged in size from 169 to 1 645 base pairs.

Table 1
Output of the RAPD amplification of sugarcane genomic DNA.

Total number of primers screened	41
Number of primers producing reliable profiles	33
Total number of loci amplified from the reliable primers	382
Average number of loci per primer	11,6
Number of primers producing scorable polymorphisms	26
Number of polymorphic loci scored ^a	54

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

Associations between RAPD polymorphisms and the traits

Based on the highest regression F -ratios as a measure of the significance of an r^2 , stepwise models in which three or four polymorphisms were used to predict the traits appeared optimal (Table 2). These models accounted for 31,6 to 40,1% of the observed phenotypic variation within the 50 varieties.

Predictive potential of the putative markers

The 'leave out one at a time' analysis resulted in the prediction of the ratings of the 50 varieties for each of the three traits. This allowed the accuracy of the selected regression models to be

Table 2
Predictive models for the three traits, determined from multiple regression analysis of the data from the 50 varieties.

	Eldana resistance			SCMV resistance			Smut resistance		
		<i>T</i> -test	<i>p</i>		<i>T</i> -test	<i>p</i>		<i>T</i> -test	<i>p</i>
RAPD fragments	OF14801 ^a	2,43	0,02	OD04806	3,47	0,002	OD14926	3,07	0,01
	OA081249	1,83	0,10	OD111085	3,64	0,001	OF06618	2,38	0,05
	OD03984	2,36	0,05	OF151112	2,26	0,050	OF01653	2,66	0,02
	OD141316	2,15	0,05	OD171109	3,61	0,001			
r^2	0,347			0,401			0,316		
<i>F</i> -value	5,972 (45 d.f.)			7,525 (45 d.f.)			7,094 (46 d.f.)		

^aRAPD fragments were named as follows: 'O' is obtained from Operon Technologies, 'A' to 'F' specifies the primer series, '14' indicates the primer number in the particular series, and the subscript indicates the size (bp) of the fragment.
 p = level of resistance.

determined by comparison of the predictions with the actual ratings. To facilitate this, the resistance ratings, which range from 1 to 9, were divided into various categories. Ratings of 1 to 3 were considered as 'resistant', 4 to 6 as 'intermediate', and 7 to 9 as 'susceptible'. As an alternative, two categories were considered whereby ratings of 1 to 4 constituted the resistant group and ratings of 5 to 9 constituted the susceptible group. When three categories were considered for eldana resistance, the four selected polymorphisms resulted in 30% of the varieties being predicted in their correct groups; 70% of the varieties were predicted in adjacent groups, i.e. predicted resistant when actually intermediate (and *vice versa*) or susceptible when actually intermediate (and *vice versa*) (Table 3). None of the varieties were predicted as resistant when actually susceptible (and *vice versa*). When two resistant categories were considered, 76% of the varieties were correctly predicted (Table 3).

Similar results were obtained for the traits of SCMV resistance (Table 4) and smut resistance (Table 5).

Table 3
Accuracy of the predicted eldana resistance ratings.

Varieties predicted...	Three categories ^a	Two categories ^b
in correct groups	30%	76%
in adjacent groups ^c	70%	–
in opposite groups ^d	0%	24%

Table 4
Accuracy of the predicted SCMV resistance ratings.

Varieties predicted...	Three categories ^a	Two categories ^b
in correct groups	52%	74%
in adjacent groups ^c	46%	–
in opposite groups ^d	2%	26%

Table 5
Accuracy of the predicted smut resistance ratings.

Varieties predicted...	Three categories ^a	Two categories ^b
in correct groups	30%	70%
in adjacent groups ^c	68%	–
in opposite groups ^d	2%	30%

^a Resistant (1-3), intermediate (4-6), susceptible (7-9)

^b Resistant (1-4), susceptible (5-9)

^c Resistant/intermediate, intermediate/susceptible

^d Resistant/susceptible

Discussion

This investigation used multiple regression analysis to identify RAPD polymorphisms that appeared to be associated with three resistance traits in sugarcane varieties. The selected polymorphisms were then used to predict the resistance ratings of each of the varieties in turn. In order for a predictive model to be used in a selection programme, the classification of unknown varieties into the various categories should be as accurate as possible. Both the r^2 of the model (the proportion of variability explained by the model) and the number of groups used to classify the individuals, will affect the accuracy of the classification by a predictive model (Schenk and Westerhaus, 1993).

The best predictive model appeared to be that for SCMV resistance. With an r^2 of 0.401 (Table 2), 52% of the varieties were correctly classified when three groups of 'resistant', 'intermediate' and 'susceptible' were considered (Table 4). Only 2% of the varieties were predicted in opposite groups (i.e. predicted resistant when susceptible, and *vice versa*). If such markers were to be used in a sugarcane selection programme, and it was decided that only the varieties predicted as susceptible were to be discarded, a very small percentage of the resistant varieties would be erroneously discarded. If marker-assisted selection (MAS) were to be carried out in the early stages of the selection programme, it could be argued that the few susceptible varieties that get carried through to successive stages, due to prediction as intermediate or resistant, would get discarded subsequently using the conventional techniques of artificial inoculation and field trials.

The success of MAS for the production of improved genotypes depends exclusively on the extent of linkage between the markers and the genetic loci controlling the trait. Studies on genetic linkage are invariably based on populations derived from planned crosses. However, a recent study by Virk *et al.* (1996) demonstrated that diverse rice germplasm (i.e. accessions or varieties) could be used to identify associations between various quantitative traits and RAPD molecular markers. These authors argue that there is no reason why the principles that are applied to a segregating population cannot be applied to a collection of landraces or varieties, assuming that similar associations are observed between marker loci and the various allelomorphous forms of QTLs, and that the basis of these is genetic linkage. Taking into account the narrow genetic background of sugarcane varieties, the low number of meioses since the original inter-specific crosses, and the apparent low rate of recombination in sugarcane (D'Hont *et al.*, 1994), most of the chromosomes initially contributed by the progenitors may have conserved the initial arrangements between markers and useful genes (Lu *et al.*, 1994a). Linkage can only be proven by analysing populations derived from single crosses between two appropriate parents. To overcome the lack of well characterised sugarcane populations, the validity of markers such as those identified in this study can be tested by using them to screen an additional selection of known varieties and determining whether their phenotypes are accurately predicted.

To obtain markers with maximal value to a selection programme, it would be recommended to screen as many new RAPD primers as possible to obtain additional polymorphisms. Multiple regression analysis of those polymorphisms could then be carried out to identify polymorphisms that could supplement, or even replace, the potential markers identified in this study. This is likely to result in a much stronger association between the polymorphisms and the traits, indicated by higher r^2 values than those obtained in this study.

If the particular trait being studied is qualitative, a genetic locus should be found that, when used in isolation, will accurately predict the phenotype. However, the complex nature of the sugarcane genome means that simple Mendelian inheritance has limited applicability in this crop (Hogarth, 1987). The accurate prediction of quantitative traits will require a few, at best, or even several markers. Since the selection of improved sugarcane genotypes requires the screening of thousands of clones, a large number of markers will be impractical and expensive. A primary goal would therefore be to find loci that have a major effect on the phenotype, thereby describing the maximum amount of variation with a minimum number of markers.

The requirements of the sugarcane selection process need to be assessed in order for appropriate markers to be identified. Questions to address include whether varieties with intermediate levels of resistance will be acceptable, and therefore whether it would be more appropriate to consider two categories of resistance (resistant and susceptible) rather than three categories (resistant, intermediate and susceptible). An important consideration is the cost of using markers for selection. Only if the expected return of using MAS instead of conventional selection exceeds the relative cost of using markers, will MAS be a feasible option. Such considerations would have to be assessed separately for each trait that requires the use of markers for selection.

Acknowledgements

The authors wish to thank staff of the Plant Breeding Department at SASEX for assistance in the collection of sugarcane varieties, and staff of the Plant Pathology and Entomology Departments (SASEX) for information on disease and eldana resistance.

REFERENCES

- Al-Janabi, SM, Honeycutt, RJ and Sobral, BWS (1994). Chromosome assortment in *Saccharum*. *Theoretical and Applied Genetics* 89: 959-963.
- Da Silva, JAG, Honeycutt, RJ, Burnquist, W, Al-Janabi, SM, Sorrells, ME, Tanksley, SD and Sobral, BWS (1995). *Saccharum spontaneum* L. 'SES 208' genetic linkage map combining RFLP- and PCR-based markers. *Molecular Breeding* 1: 165-179.
- D'Hont, A, Lu, Y-H, González de León, D, Grivet, L, Feldmann, P, Lanaud, C and Glaszmann, JC (1994). A molecular approach to unravelling the genetics of sugarcane, a complex polyploid of the Andropogoneae tribe. *Genome* 37: 222-230.
- Grivet, L, D'Hont, A, Roques, D, Feldmann, P, Lanaud, C and Glaszmann, JC (1996). RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploid and aneuploid interspecific hybrid. *Genetics* 142: 987-1000.
- Haley, SD, Afanador, LK, Miklas, PN, Stavely, JR and Kelly, JD (1994). Heterogeneous inbred populations are useful as sources of near-isogenic lines for RAPD marker localization. *Theoretical and Applied Genetics* 88: 337-342.
- Harvey, M, Hockett, BI and Botha, FC (1994). Use of the polymerase chain reaction (PCR) and random amplification of polymorphic DNAs (RAPDs) for the determination of genetic distances between 21 sugarcane varieties. *Proc S Afr Sug Technol Ass* 68: 36-40.
- Hogarth, DM (1987). Genetics of Sugarcane. In: DJ Heinz (ed.) *Developments in Crop Science 11: Sugarcane Improvement Through Breeding*. Elsevier, Amsterdam. pp 255-271.
- Honeycutt, RJ, Sobral, BWS, Keim, P and Irvine, JE (1992). A rapid extraction method for sugarcane and its relatives. *Plant Molecular Biology Reporter* 10: 66-72.
- Lu, YH, D'Hont, A, Paulet, F, Grivet, L, Arnaud, M and Glaszmann, JC (1994a). Molecular diversity and genome structure in modern sugarcane varieties. *Euphytica* 78: 217-226.
- Lu, YH, D'Hont, A, Walker, DIT, Rao, PS, Feldmann, P and Glaszmann, JC (1994b). Relationships among ancestral species of sugarcane revealed with RFLP using single copy maize nuclear probes. *Euphytica* 78: 7-18.
- Martin, GB, Williams, JGK and Tanksley, SD (1991). Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proceedings of the National Academy of Sciences USA* 88: 2336-2340.
- Ohmori, T, Murata, M and Motoyoshi, F (1996). Molecular characterization of RAPD and SCAR markers linked to the *Tm-1* locus in tomato. *Theoretical and Applied Genetics* 92: 151-156.
- Poulsen, DME, Henry, RJ, Johnston, RP, Irwin, JAG and Rees, RG (1995). The use of bulk segregant analysis to identify a RAPD marker linked to leaf rust resistance in barley. *Theoretical and Applied Genetics* 91: 270-273.
- Schenk, JS and Westerhaus, MO (1993). *Analysis of Agriculture and Food Products by Near Infrared Reflectance Spectroscopy*. Monograph, Infrasoft International, Port Matilda, PA.
- Sobral, BWS and Honeycutt, RJ (1993). High output genetic mapping of polyploids using PCR-generated markers. *Theoretical and Applied Genetics* 86: 105-112.
- Virk, PS, Ford-Lloyd, BV, Jackson, MT, Pooni, HS, Clemeno, TP and Newbury, HJ (1996). Predicting quantitative variation within rice germplasm using molecular markers. *Heredity* 76: 296-304.
- Welsh, J and McClelland, M (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18: 7213-7218.
- Williams, JGK, Kubelik, AR, Livak, KJ, Rafalski, JA and Tingey, SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18: 6531-6535.