Theor Appl Genet (2006) 114:155–164
DOI 10.1007/s00122-006-0418-8

ORIGINAL PAPER

Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure

Xianming Wei · Phillip A. Jackson · C. Lynne McIntyre · Karen S. Aitken · Barry Croft

Received: 12 March 2006 / Accepted: 23 September 2006 / Published online: 18 October 2006

Abstract Association between markers and sugarcane diseases were investigated in a collection of 154 sugarcane clones, consisting of important ancestors or parents, and cultivars. 1,068 polymorphic AFLP and 141 SRR markers were scored across all clones. Data on the four most important diseases in the Australian sugarcane industry were obtained; these diseases being pachymetra root rot (*Pachymetra chaunorhiza* B.J. Croft & M.W. Dick), leaf scald (*Xanthomonas albilineans* Dowson), Fiji leaf gall (*Fiji disease virus*), and smut (*Ustilago scitaminea* H. & P. Sydow). By a simple regression analysis, association between markers and diseases could be readily detected. However, many of these associations were due to the effects of embedded population structure and random effects. After taking population structure into account, we found that 59% of the phenotypic variation in smut resistance ratings could be accounted for by 11 markers, 32% of variation for leaf scald and pachymetra root rot rating by 4 markers, and 26% of Fiji leaf gall by 5 markers. The results suggest that marker–trait associations can be readily detected in populations generated from modern sugarcane breeding programs. This may be due to special features of past sugarcane breeding programs leading to persistent linkage disequilibrium in modern parental populations.

Communicated by A. Charcosset.

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0418-8 and is accessible for authorized users.

X. Wei
BSES Limited, PMB57, Mackay Mail Centre, Mackay, QLD 4741, Australia

P. A. Jackson (✉)
Davies Laboratory, CSIRO Plant Industry, University Drive, Townsville, QLD 4814, Australia
e-mail: Phillip.Jackson@csiro.au

C. L. McIntyre · K. S. Aitken
Queensland Bioscience Precinct, CSIRO Plant Industry, 306 Carmody Rd, St. Lucia, QLD 4067, Australia

B. Croft
BSES Limited, 90 Old Cove Rd, Woodford, QLD 4514, Australia

Introduction

Sugarcane produces over 60% of the world's sugar and is being increasingly used for renewable energy via ethanol production and electricity generation. Until early in the twentieth century, cultivated sugarcane cultivars consisted mainly of *Saccharum officinarum* clones, collected from Papua New Guinea and Indonesia. Around 1920, breeders in India and Indonesia initiated interspecific breeding programs, which used clones from the species *S. spontaneum* with *S. officinarum* (Daniels and Roach 1987). *S. spontaneum* is a wild species characterised by a number of traits which contrast with *S. officinarum* such as low sucrose, high fibre levels, strong ratoon growth after harvesting, and adaptation to a range of environmental stresses. The initial interspecific hybrids were repeatedly crossed back to *S. officinarum* clones or other hybrids in order to recover sufficiently high sugar content, in a process termed ''nobilisation'' by sugarcane breeders (Bremer

1961). Only a small number of *S. officinarum* and *S. spontaneum* clones were used in these early interspecific crosses (Arcenaux 1967). Most modern cultivars and parental material in modern sugarcane breeding programs trace back to a relatively small number of clones used in the initial interspecific hybridisations, and are less than eight cycles of intercrossing and selection from these original clones. Propagation between these cycles has been exclusively vegetative.

The complexity of the sugarcane genome is probably greater than any other important crop (D'Hont and Glaszmann 2001; Grivet and Arruda 2001). An important feature of sugarcane is the high level of polyploidy. *S. officinarum* clones have $2n = 80$ chromosomes, with a basic number of $x = 10$ indicating these are octoploids, while *S. spontaneum* has a $2n$ number ranging from 40 to 128 with a basic number of $x = 8$ indicating a ploidy series between 5 and 16 (D'Hont et al. 1998). Modern cultivars generally have between 100 and 130 chromosomes (Grivet and Arruda 2001). In situ hybridisation studies have suggested that modern cultivars comprise about 80% of their chromosomes inherited entirely from *S. officinarum*, about 10% entirely from *S. spontaneum* and about 10% are the result of recombination from these two ancestral species (D'Hont et al. 1996).

To date, specially derived mapping populations (e.g. F2, RIL or DH in inbred crops, F1 or BC in outcrossing species) have been key tools for QTL mapping in plants including sugarcane. More recently, methodology for QTL mapping under the names of association mapping and linkage disequilibrium mapping has become popular in human genetics (Cardon and Bell 2001) and is being applied in plants (e.g. Remington et al. 2001; Thornsberry et al. 2001; Gebhardt et al. 2004; Breseghello and Sorrells 2006; Camus-Kulandaivu. 2006). These approaches use naturally occurring populations derived from many different parents to find marker–trait associations. The potential advantages and disadvantages with association mapping in diverse populations, compared with cross-specific populations have been discussed in detail previously (e.g. Pritchard et al. 2000; Breseghello and Sorrells 2006). The main advantages lie in the potential relevance of the results across populations representative of those routinely generated from many parents and crosses in breeding programs. In addition, already available phenotypic databases routinely generated in breeding programs can often be exploited, allowing large populations and genotypes to be sampled at no extra cost (e.g. Parisseaux and Bernado 2004). The major disadvantage lies in potential complications and difficulties in identifying robust marker–trait associations.

An important complication to consider is that of population structure (Cardon and Palmer 2003). The presence of groups of genotypes within the population with an unequal contribution from different ancestors can result in marker–trait associations not due to physical linkage between markers and causal alleles (Knowler et al. 1988). A range of statistical methodologies have been developed that attempt to detect and separate effects of underlying structure (e.g. Pritchard et al. 2000; Yu et al. 2006; Parisseaux and Bernardo 2004). Many populations in plant breeding programs have some level of underlying structure. This issue may be particularly relevant in some sugarcane breeding program populations where grouping based on differing contributions of ancestral clones may be strong because of the limited number of recombination cycles since initial use of the progenitor clones.

In association mapping, the distance over which linkage disequilibrium persists in a population is a key factor to consider, since this will determine the density of markers needed to find associations between QTL and markers. The breeding history of sugarcane, consisting of a strong foundation bottleneck, followed by a small number of cycles of intercrossing and vegetative propagation suggests that linkage disequilibrium should be extensive and that a high density of markers may not be needed in sugarcane populations to detect marker–trait associations. This hypothesis is supported by a study by Janoo et al. (1999) who found considerable linkage disequilibrium among markers due to physical linkage retained in a population of 59 modern sugarcane cultivars.

The aim of the work reported here was to investigate whether association mapping in sugarcane could be a useful approach in identifying marker–trait associations for marker-assisted selection in sugarcane breeding programs. This included an assessment of whether many marker–trait associations found could be attributed to population structure effects, because such associations would be of limited value in selection in future generations. The focus in this study was on resistance to the diseases sugarcane smut (*Ustilago scitaminea* H. & P. Sydow), pachymetra root rot (*Pachymetra chaunorhiza* B.J. Croft & M.W. Dick), leaf scald (*Xanthomonas albilineans* Dowson), and Fiji leaf gall (*Fiji disease virus*). Sugarcane smut is a fungal disease that can cause complete crop loss in susceptible cultivars (Croft and Braithwaite 2006). Sugarcane smut occurs throughout the world but was only recently discovered (2006) in the main sugarcane production areas on the east coast of Australia. DNA markers

would assist breeding programs in Australia because resistance genes could be identified in clones in the absence of the disease. Pachymetra root rot is caused by a heterokont in the class Oomycetes and is unique to Australia (Croft and Magarey 1989). The fungus attacks the primary roots of sugarcane and causes loss of plant anchorage in the soil and significant yield loss. Leaf scald is caused by a bacteria that infects the xylem cells and can cause wilting and death of infected plants (Rott and Davis 2000). Fiji leaf gall is a viral disease that is spread by planthoppers (Ridley et al. 2006) and has caused serious epidemics in Australia and Fiji. All four diseases are managed primarily with resistant cultivars and are therefore of practical importance in sugarcane breeding programs in Australia (Magarey et al. 2005), with leaf scald, Fiji leaf gall and sugarcane smut being important internationally (Comstock 2000; Rott and Davis 2000).

Materials and methods

Selection of genotypes

A panel of 154 clones (Supplementary Table 1s) were sampled from collections maintained by BSES Ltd and CSR Ltd, Australia. These clones were derived from diverse pedigrees and do not form any predesigned mapping population. In most cases the clones were collected from multiple places to confirm identity. In three cases (for clones named as CO1007, H60-3802, and QN77-792) some differences in DNA marker profile were scored and in these cases the different clones under the same name were retained as duplicates (Supplementary Table 1s).

The clones were selected on the basis that they were important ancestors or parents in the Australian breeding program, or that they were past or present cultivars. Some clones bred and selected from outside Australia were also included (Supplementary Table 1s) since such foreign clones are regularly used in the Australian sugarcane breeding program as a source of potentially different genetic diversity. The foreign clones were chosen arbitrarily from the germplasm collections but biased to include clones retained as proven parents in the Australian program. Some advanced stage selections from the Australian sugarcane breeding program, which may also be potentially used as a source of new experimental parents, were also included. The latter class of clones were included with an emphasis in selecting a greater proportion of clones with resistance to smut susceptibility than would otherwise result from random sampling Australian sugarcane cultivars or parents. Smut disease has not been present in eastern Australia, and about 80% of Australian cultivars are currently susceptible to this disease (Croft and Braithwaite 2006). An emphasis is now being placed on breeding for smut resistance as a preparation for likely introduction of smut in the future. A more balanced set of clones in relation to smut susceptibility versus resistance in this study was sought in order to improve the statistical power of detecting associations between markers and smut resistance.

Marker characterisation

For DNA extraction, young leaves were sampled from progeny in the field, freeze-dried, then ground to a powder and stored at –20°C. Genomic DNA was extracted following the CTAB method described by Hoisington (1992). Both SSR and AFLP markers were generated using the methods described in Aitken et al. (2005). SSR primers were obtained from the Sugarcane Microsatellite Consortium collection (Cordeiro et al. 2000).

All clones were screened with 19 AFLP primer pairs and 14 SSRs. A total of 1,599 polymorphic AFLP markers and 181 polymorphic SSR markers were scored on the 154 sugarcane clones. 1,068 of the AFLP markers and 141 of the SSR markers had a frequency of either <95% or >5% occurrence in the 154 clones and were used for further analysis.

Disease ratings

Ratings for resistance to pachymetra root rot, leaf scald, Fiji leaf gall, and smut were obtained from a database maintained by BSES Ltd. Ratings range from 1 to 9 where 1 is highly resistant and 9 is highly susceptible. Procedures used for obtaining these ratings were described in (Magarey et al. 2005). In brief, resistance levels were measured from experiments involving standardised procedures and artificial inoculation. The trials were conducted at Woodford, Australia (for Fiji leaf gall and leaf scald), Tully, Australia (for pachymetra root rot) and Indonesia (for smut). Trials were conducted between 1992 and 2004, with most clones evaluated in one to three trials. There were ten standard clones with known ratings established from extensive prior testing included in common to all trials for each disease. Data was retained for trials where there was a high correlation ($r > 0.8$) between infection levels observed in the standard clones and the previously established ratings. The regression between infection level and resistance rating in the standard clones in each trial

was used to assign a resistance rating to all other clones in each trial. The average value from all trials (where more than one test was done) was used.

Data analysis

Genotypes were grouped using the program STRUCTURE v2 (Pritchard et al. 2000). This program used the marker data to place genotypes into groups based on similarity on overall marker profiles.

Markers were first examined individually using a *t* test to seek initial evidence for an association between presence or absence of the marker and each of the disease ratings in Genstat® (Lawes Agricultural Trust 2005). Threshold values for the test statistic *t* were determined firstly assuming normal distributions of trait values, and secondly using permutation testing (Churchill and Doerge 1994). In all cases the threshold values using both methods were similar, and probabilities (*P*) of greater *t* values assuming the null hypothesis of no differences in marker classes for the trait values are presented in this paper for the former.

It is recognised that markers identified as being associated with disease ratings in the population used could be spurious due to embedded population structure (Jannink and Walsh 2002) or random variation (type 1 statistical errors). Markers significant at a significance threshold of $P < 0.001$ based on individual comparison-wise tests were therefore examined further.

To test for evidence of contribution of marker–trait associations due to population structure, two models were fitted for the markers above:

$$\text{Disease rating} = \text{group} + \text{marker within group} + \text{residual} \quad (1)$$

$$\text{Disease rating} = \text{group} + \text{marker} + \text{group} \times \text{marker} + \text{residual} \quad (2)$$

where group was determined by STRUCTURE when eight subpopulations were assumed. The analyses were carried out in Genstat®.

The presence of significant marker × group interaction would indicate that the effects of the marker differed depending on the ancestral background. The absence of significant marker effects within groups would suggest that marker–trait associations could be due to population structure effects rather than physical linkage. Conversely, the presence of significant within-groups variance would support the hypothesis that marker–trait association is independent of population structure.

Problems with large numbers of multiple comparisons increasing the type 1 error rate above α in relation testing marker–trait associations across the entire genome are well known (e.g. Churchill and Doerge 1994). Therefore, experiment-wise critical values were also computed for comparison in this study using permutation testing. Experiment-wise thresholds for the *t* test statistic were estimated by permutation as described in (Churchill and Doerge 1994) and markers were tested against these thresholds.

Markers identified using individual comparison tests as remaining significant after taking into account the effects of population structure were combined in multiple regression analyses in Genstat®. Due to the large number of markers involved, Procedure RSELECT in Genstat® was used to determine the best subset, which was judged by adjusted r^2.

Table 1 Number of AFLP or SSR markers significantly associated with four disease ratings at three levels, $P \leq 0.05$, ≤ 0.01, and ≤ 0.001

Diseases	No. of significant markers ($P \leq 0.05$)	No. of significant markers ($P \leq 0.01$)	No. of significant markers ($P \leq 0.001$)	Max % variation explained by a marker
AFLPs				
Fiji leaf gall	140 (80)	43 (21)	9 (5)	15.1
Leaf scald	68 (100)	18 (39)	5 (5)	13.0
Pachymetra root rot	190 (89)	59 (24)	12 (5)	13.2
Smut	200 (108)	82 (44)	20 (19)	22.6
SSRs				
Fiji leaf gall	12 (9)	1 (4)	0 (2)	5.6
Leaf scald	11 (17)	5 (7)	0 (1)	7.3
Pachymetra root rot	18 (17)	11 (7)	3 (0)	9.4
Smut	24 (18)	8 (6)	2 (2)	11.1

Figures in brackets were the number of markers significant after the population structure was taken into account

Table 2 Number of markers reaching significance at three α levels on an experiment-wise basis, for each of the four diseases

Diseases	$P \leq 0.05$	$P \leq 0.01$	$P \leq 0.001$
AFLPs			
Fiji leaf gall	1	1	0
Leaf scald	1	0	0
Pachymetra root rot	2	2	1
Smut	10	8	5
SSRs			
Fiji leaf gall	0	0	0
Leaf scald	0	0	0
Pachymetra root rot	3	3	0
Smut	2	1	1

Results

Association between markers and disease resistance

Under the null hypothesis of no linkage between markers and QTL for the traits examined, approximately 53 and 11 AFLP markers ($0.05 \times 1{,}068$ total markers; $0.01 \times 1{,}068$), and 7 and 1 SSR markers (0.05×141; 0.01×141) would be expected to give t values greater than the $P \leq 0.05$ and $P \leq 0.01$ threshold levels, respectively, by random chance. A larger number of markers produced t values greater than these values were observed (Table 1). Very few markers were significant for more than one disease. The number of markers showing association was greatest for smut, with more markers and greater maximum levels of variation explained for this disease than for the others.

Experiment-wise thresholds were used to test for the presence of a QTL anywhere in the genome while controlling the overall type 1 error rate to be α or less. Thresholds for t values determined by permutation methods on an experiment-wise basis were approximately twice as high as the t values used to determine the significance of the effect of a marker on an individual comparison basis. Consistent with results for the individual marker comparison tests, the number of markers reaching significance was greatest for smut disease (Table 2).

Effects of population structure on associations between markers and disease resistance

The number of groups (K) to examine was determined arbitrarily based on (1) an estimate of a ''goodness of fit'' parameter ln $P(X|K)$, which is an informal pointer provided by STRUCTURE of number of subpopulations (Pritchard et al. 2000) and (2) examination of pedigree information. The parameter ln $P(X|K)$ increases with increasing group number, but with the reaching of a plateau being indicative of underlying number of subpopulations. A plateau is indicated when the difference between two K values is within 5 to 10, but could be 50 for large data sets. In the current study, differences of ln $P(X|K)$ were larger than 200 when K was less than 8, but this difference decreased to 37.4 between $K = 8$ and $K = 9$. This grouping level also showed a close correspondence between the group membership and pedigree structure (Supplementary Table 1s), and eight groups was therefore chosen for analysis of the effects of population structure in this study. Group membership at the three-group level is also shown (Supplementary Table 1s) to illustrate relationships among the eight groups and population structure indicated at a higher level. For both levels of grouping, the clear correspondence between the results from the STRUCTURE program and pedigree data provided confidence in the validity of groupings in

Table 3 Average resistance ratings of eight groups of clones (membership shown in Table 1) for Fiji leaf gall, leaf scald, pachymetra root rot, and smut

	Fiji leaf gall			Leaf scald			Pachymetra root rot			Smut		
Group	No.	Rating	SD	No.	Rating	SD	No.	Rating	SD	No.	Rating	SD
1	26	1.8	1.2	27	2.6	1.9	27	3.9	1.7	27	7.1	1.9
2	7	5.0	2.4	6	1.2	0.4	9	7.4	1.9	3	3.0	1.4
3	8	1.4	0.7	9	2.4	1.4	11	3.4	1.8	11	3.9	2.4
4	18	4.5	2.6	16	3.1	2.2	19	4.8	2.5	13	4.5	2.4
5	8	1.9	1.4	8	2.6	1.8	12	5.1	2.2	12	5.4	3.0
6	25	2.5	2.2	23	2.4	1.9	26	4.6	2.2	25	5.6	2.7
7	13	4.2	2.7	12	2.2	1.8	13	4.6	1.5	12	7.3	1.7
8	24	3.4	2.3	24	2.4	1.8	28	5.9	1.8	26	6.1	2.6
	$P < 0.001$			NS			$P < 0.01$			$P < 0.001$		

The number of members in each group contributing to the mean for each disease, and the standard deviations in each group is shown. Due to not all clones having ratings for every disease, numbers in groups are different for each disease

Table 4 Average disease resistance ratings for *Pachymetra*, leaf scald, and smut for clones with and without marker present for markers significant at $P \leq 0.001$ based on individual comparison-wise tests

Marker	Location in Q165	Mean for clones with marker	Mean for clones without marker	Prob of >*F* for marker within-groups effect	Prob of >*F* for marker × group interaction
AFLP					
Fiji leaf gall					
AAC-CTA41	Not present	5.05 (14)	2.80 (113)	0.046	0.19
ACA-CTA55	Not present	1.72 (26)	3.39 (101)	0.561	0.91
ACA-CTG33	Not present	2.00 (43)	3.56 (85)	0.123	0.72
ACC-CAC78	Not present	4.70 (22)	2.69 (106)	0.072	0.47
ACC-CTA23	HG4 LG3	3.72 (64)	2.34 (64)	0.035	0.14
ACG-CTA27	Multidose	3.51 (88)	2.00 (40)	0.066	0.72
ACG-CTA28	Not present	4.63 (23)	2.68 (105)	0.001	0.23
ACT-CAC35	Not present	4.30 (32)	2.61 (96)	0.235	0.65
AGC-CTG54	HG1 LG42	1.56 (36)	3.61 (92)	<0.001	0.41
Leaf scald					
ACC-CTC106	Not present	4.13 (13)	2.30 (111)	0.002	0.072
ACC-CTC62	Multidose	2.09 (80)	3.22 (44)	<0.001	0.023
ACC-CTG11	Not present	1.72 (47)	2.96 (77)	<0.001	0.085
AGC-CTA44a	Not present	1.68 (51)	3.06 (73)	0.002	0.706
AGG-CTC25	Not present	1.79 (53)	3.02 (71)	<0.001	0.209
Pachymetra root rot					
AAC-CTA24	Not present	6.34 (30)	4.54 (113)	0.011	0.062
AAC-CTA49	Unlinked	5.76 (51)	4.44 (92)	0.230	0.702
AAC-CTA91	multidose	4.55 (103)	5.85 (40)	0.014	0.041
AAC-CTC80	Not present	5.51 (69)	4.33 (75)	0.106	0.305
ACA-CTA11	Not present	3.94 (40)	5.27 (103)	0.052	0.400
ACC-CTA23	HG4 LG3	3.72 (64)	2.34 (64)	0.385	0.667
ACC-CTA27	Not present	7.44 (9)	4.72 (135)	0.363	0.933
ACC-CTC3	Multidose	4.23 (65)	5.44 (79)	0.012	0.147
ACG-CTC31	Not present	4.05 (68)	5.65 (76)	0.123	0.576
ACG-CTG39	Multidose	5.33 (94)	4.08 (50)	0.001	0.493
ACG-CTG47a	HG4 LG13	7.22 (9)	4.74 (135)	0.102	0.924
AGC-CTC3b	Multidose	4.06 (51)	5.35 (93)	0.126	0.481
Smut					
ACA-CTA41	Multidose	6.36 (86)	4.72 (41)	0.081	0.243
ACA-CTA77	HG3 LG4	6.50 (76)	4.84 (51)	0.121	0.439
ACA-CTA91	Not present	3.63 (33)	6.61 (94)	<0.001	0.873
ACA-CTG17a	Not present	4.80 (46)	6.41 (82)	0.006	0.038
ACC-CAC2	HG3 LG51	7.09 (57)	4.82 (71)	<0.001	0.536
ACC-CAC46	Not present	3.77 (15)	6.11 (113)	<0.001	0.024
ACC-CAC58	HG4 LG48	6.49 (75)	4.91 (53)	0.003	0.143
ACC-CTC50	Not present	2.39 (6)	6.00 (122)	<0.001	0.938
ACC-CTC68	HG3 LG63	3.20 (11)	6.08 (117)	0.005	0.049
ACG-CAG4	Not present	6.70 (63)	5.00 (65)	0.163	0.63
ACG-CTA10a	HG4 LG3	6.33 (89)	4.71 (39)	0.049	0.066
ACG-CTA34	HG4 LG3	6.61 (66)	5.01 (62)	<0.001	0.004
ACG-CTG6	HG4 LG25	6.97 (70)	4.47 (58)	<0.001	0.843
AGC-CTA30a	Not present	3.84 (23)	6.27 (105)	<0.001	0.261
AGC-CTA30b	HG2 LG85	6.52 (78)	4.77 (50)	0.065	0.233
AGC-CTA32a	Not present	7.38 (28)	5.40 (100)	0.121	0.693
AGC-CTA37	Unlinked	6.72 (61)	5.02 (67)	0.057	0.525
AGG-CAC12b	HG1 LG17	6.69 (57)	5.15 (71)	0.040	0.091
AGG-CAC1d	Not present	6.80 (49)	5.23 (79)	0.139	0.465
AGG-CTC29	Not present	3.46 (16)	6.17 (112)	0.007	0.661
SSR					
Pachymetra root rot					
ms39–1	HG5 multidose	3.99 (52)	5.41 (92)	0.017	0.040
ms286–3	HG3 not present	6.11 (32)	4.55 (112)	0.218	0.750
ms1608–1	HG3 not present	6.08 (31)	4.57 (113)	0.362	0.819

Table 4 continued

Marker	Location in Q165	Mean for clones with marker	Mean for clones without marker	Prob of >*F* for marker within-groups effect	Prob of >*F* for marker × group interaction
Smut					
ms851–9	HG2 not present	4.63 (45)	6.49 (83)	0.315	0.702
ms286–4	HG3 not present	7.69 (18)	5.53 (110)	0.416	0.947

Number of clones in each category is given in parentheses. The probability of >*F* value from analysis of variance for (1) marker within-groups effect and (2) the marker × group interaction shown for each marker. The location (homology group, HG; linkage group, LG) of each marker mapped in cultivar Q165 (Aitken et al. 2005) is indicated: ''not present'' indicates marker not present in Q165, ''multidose'' refers to a higher dosage marker unable to be mapped

Table 5 Results of multiple regression using AFLPs as predictors

Disease	With all markers		Best subset		
	No.	%	No.	%	Markers
Fiji leaf gall	4	31.9	4	31.9	AAC-CTA41, ACC-CTA23, ACG-CTA28, AGC-CTG54
Lead scald	5	26.2	4	25.8	ACC-CTC62, AGC-CTA44a, ACC-CTC106, AGG-CTC25
Pachymetra root rot	4	31.9	4	31.9	AAC-CTA24, AAC-CTA91, ACC-CTC3, ACG-CTG39
Smut	23	62.9	11	59.0	ACA-CTA91, ACA-CTG17a, ACC-CAC2, ACC-CAC58, ACC-CTC50, ACC-CTC68, ACG-CTA10a, ACG-CTA34, ACG-CTG6, AGC-CTA30a, AGG-CTC29

sugarcane based on application of STRUCTURE to random DNA markers. It also provided confidence in the pedigree information about the clones sampled. This is relevant given occasional concerns expressed by breeders about possible lack of control of stray pollen and pollen contamination of crosses in the past in sugarcane breeding programs, and possibility of occasional but cumulative mistakes in propagation of clonal populations across years.

The population showed no clear discontinuities in overall structure, either in the pedigree information (partly shown in Supplementary Table 1s via the parentages), or in principal component analysis on coancestry coefficients matrix generated from pedigree (data not shown), or in percentages of genome originated from a group using all the marker data by STRUCTURE (data not shown). However, there are obviously structural features in the population based on the pedigree information, with some clones being closely related (e.g. from the same parents). The presence of clear structural features but without strong grouping discontinuities is typical of parent collections used in sugarcane breeding programs. In this situation, it would be expected that marker–phenotype associations could easily arise due to uneven contributions of particular parents or ancestors, without necessarily being due to physical linkage of markers and QTL affecting the phenotypes of interest.

The groups differed very significantly ($P \leq 0.01$) for the average of the ratings of the member clones for Fiji leaf gall, pachymetra root rot, and smut, but not for leaf scald (Table 3), indicating that population structure is associated with a significant proportion of variation in resistance for the former three diseases.

When population structure at the eight-group level was taken into account, the number of AFLP markers showing significant association when tested for thresholds used for individual comparisons almost halved except for leaf scald (Table 1). This was less obvious for the SSR markers where more markers were significant for Fiji leaf gall and leaf scald. The different pattern for leaf scald was expected because of its nonsignificant differences of disease ratings among groups. All markers showing a significant association with resistance level before population structure was accounted for were also significantly associated when only the three-group level was considered (data not shown). This indicates that this high level of grouping was inadequate in eliminating spurious associations due to population structure.

All markers significant at $P \leq 0.001$ based on threshold levels used for individual comparisons before population structure was taken into account for each of the diseases are listed in Table 4. In all cases these markers accounted for greater than 7% of the variance in disease resistance ratings. For each marker, evidence for marker × group interaction and marker within-groups variance was examined via an analysis of variance using the models explained above.

Marker × group interactions were significant ($P \leq 0.05$) for no markers for Fiji leaf gall, one marker for leaf scald, one marker for *Pachymetra*, and four markers for smut. Most of these interactions were manifest as differences in size of marker effects between groups without a reversal of the effects, but there were also cases of effects being reversed between the groups (data not shown). These results indicate that in most cases, the marker effects were not dependent on differences in genetic background within the population used, but in some cases the marker effects varied in different backgrounds.

The numbers of markers that initially were significant at $P \leq 0.001$ and also retained at $P \leq 0.05$ within groups were 4 out of 9 (56%) for Fiji leaf gall, 5 out 5 for leaf scald (100%), 5 out of 12 (42%) for *Pachymetra*, and 12 out of 20 (60%) for smut. This indicates that most marker–trait associations found in the whole population could not be attributed to population structure, although a significant proportion of marker–trait associations could be attributed to this cause.

Multiple regression

Table 5 presents the results of multiple regression using a small number of markers as predictors for the four diseases. The percentage of variation explained by multiple markers varied from 26.2 for leaf scald to 58.6 for smut on all markers. It should be noted that only slightly less of the variation could be explained by a small number of markers. For example, 53.7% of the variation of smut ratings could by explained by four markers, namely, ACA-CTA91, ACC-CAC2, ACC-CTC50, and ACG-CTG6.

Discussion

Markers apparently associated with disease resistance could be readily found in this study. The results suggested that most of these associations could not be attributed to population structure effects. Given these results it would appear that markers linked to important QTL in sugarcane populations similar to those we have used may be easily found without need for a very high density of genome coverage. This result is consistent with the breeding history of sugarcane and the presence of extensive linkage disequilibrium remaining in populations being generated within modern sugarcane breeding programs (Janoo et al. 1999). In fact, this preliminary study showed a higher occurrence of linkage equilibrium occurring in our population. With the Fisher exact test, we found 736 out of 569,778 pairs of markers were significantly associated ($P = 8.78 \times 10^{-8}$), compared with 59 out of 540,688 reported by Janoo et al. (1999).

The number of markers used in this study would not provide for an extensive genome coverage in sugarcane given the large genome size, but with extensive linkage disequilibrium a moderate number of markers (less than 1,500) may still be able to detect many marker–trait associations. Nevertheless, an even larger number of markers would be expected to provide more markers in more consistent coupling phase linkage to a greater number of QTL, and therefore explain a greater proportion of variation than detected in this study.

There was some evidence for marker effects changing with different genetic backgrounds as indicated by marker × group interaction effects. This may be due to changes in frequency of particular marker–QTL linkages in different backgrounds, or recombination events in some ancestors, or to epistasis effects.

Due to the absence of strong grouping discontinuities, the exact number of groups is difficult to determine. We arbitrarily split our population into eight groups based on the Q value from STRUCTURE, which appeared to give us a reasonable classification after examination of their pedigree. Further classification may be necessary for our entire breeding population. However, with this sample population, any further classification may weaken the capacity to detect the effect of population structure because of the small number of individuals within each group. It should be noted that analysis of population structure and resolution of optimal group number and composition in breeding populations such as that employed in this study is complex, and justifies further attention.

There are potential advantages and disadvantages with association studies done in populations from across multiple parents compared with studies focused on specific bi-parental populations. Use of populations derived from many parents in breeding programs avoids problems relating to specificity of marker–QTL associations in different bi-parental populations, and can more often make use of phenotypic data already obtained in breeding programs for other purposes without need for special research orientated trials. However, on the other hand, it should be easier to identify marker–trait associations in specific cross populations because some QTL effects affected by complex allele interactions in a diverse population may present as easily detected additive effects within a specific cross-population. Also, if marker–trait associations are found in a single cross-population, it can be directly inferred that these are due to physical linkage between the marker and QTL rather than possibly

population structure effects. Population specific QTL mapping will be practically useful for breeding strategies that seek to exploit the genetic variance within the population being studied. It may be valuable, for example, in sugarcane breeding programs exploiting specific crosses of high commercial value but which still retain significant genetic load (i.e. deleterious alleles), or for breeding programs seeking to introgress exotic genotypes into elite genetic backgrounds.

Only one marker was associated with resistance to more than one disease (ACC-CTA23—Fiji leaf gall and pachymetra root rot). The four diseases examined in this study varied widely in their causal agents and their modes of infection. It is therefore not unexpected that there may be no association between the genes for resistance to the diseases. However, recent studies on the disease resistance in a number of crops have shown that resistance genes often occur in gene islands (Panstruga et al. 1998; Wei et al. 2002; Brooks et al. 2002), so it would be expected that some QTLs for resistance for one disease may also be associated with resistance to other diseases. There was limited evidence of this in our study on the four diverse sugarcane diseases.

Overall, the promising results found in this study suggest a potential useful role for marker-assisted selection in sugarcane breeding programs based on similar marker–trait studies. However, despite strong evidence of the presence of marker–trait associations that were not due to population structure, there is still a possibility that residual structure contained within the groups identified in this study remained and could be causing marker–trait associations. In verifying the associations reported here, and in conducting definitive research in future, a family-based experimental design would be useful, in which the within-family variation explained by markers can be examined. In such designs, marker–trait associations within families could be attributed to physical linkage without concern about possible population structure effects, with the latter being retained among families.

Acknowledgments This study was conducted with financial support from the Sugar Research and Development Corporation (Australia), CSIRO Plant Industry, BSES Ltd, and the CRC for Sugarcane Industry Innovation through Biotechnology. We are grateful to Mark Jackson for competently conducting the laboratory work to generate the marker data.

References

Aitken KS, Jackson PA, McIntyre CL (2005) A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. Theor Appl Genet 110:789–801

Arceneaux G (1967) Cultivated sugarcanes of the world and their botanical derivation. Proc Int Soc Sugar Cane Technol 6:116–128

Bremer G (1961) Problems in breeding and cytology of sugarcane. IV. The origin in the increase of chromosome number in species hybrids *Saccharum*. Euphytica 10:325–342

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Brooks SA, Huang L, Gill BS, Fellers JP (2002) Analysis of 106 kb of contiguous DNA sequence from the D genome of wheat reveals high gene density and a complex arrangement of genes related to disease resistance. Genome 45(5):963–972

Camus-Kulandaivu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barruaud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172:2449–2463

Cardon LR, Bell JI (2001) Association study designs for complex diseases. Nat Rev Genet 2:91–99

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361:598–604

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963–971

Comstock JC (2000) Smut. In: Rott P, Bailey RA, Comstock JC, Croft BJ, Saumtally AS (eds) A guide to sugarcane diseases. CIRAD and ISSCT, Montpellier, pp 181–185

Cordeiro GM, Taylor GO, Henry RJ (2000) Characterisation of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polymorphic species. Plant Sci 155:161–168

Croft BJ, Braithwaite KS (2006) Management of an incursion of sugarcane smut in Australia. Australas Plant Pathol (in press)

Croft BJ, Magarey RC (1989) A review of research into pachymetra root rot, an important new fungal disease of sugarcane. Proc Int Sugar Cane Technol 20:686–694

Daniels J, Roach BT (1987) Taxonomy and evolution. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 7–142

D'Hont A, Glaszmann JC (2001) Sugarcane genome analysis with molecular markers: a first decade of research. In: Hogarth DM (ed) International Society of Sugar Cane Technologists. Proceedings of the XXIV Congress, Brisbane, Australia, 17–21 September 2001, pp 556–559

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, Ison D, Alix K, Roux C, Glaszmann JC (1998) Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. Genome 41:221–225

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schüler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker–trait associations: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol Breed 13:93–102

Grivet L, Arruda P (2001) Sugarcane genomics: depicting the complex genome of an important tropical crop. Curr Opin Plant Biol 5:122–127

Hoisington D (1992) Laboratory protocols. CIMMYT applied molecular genetics laboratory. CIMMYT, Mexico

Jannink J-L, Walsh B (2002) Association mapping in plant populations. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CABI, Wallingford, UK, pp 59–68

Janoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. Theor Appl Genet 99:1053–1060

Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988) GM[3–5,13,14] and type 2 diabeties mellitus: an association in American Indians with genetic admixture. Amer J Hum Genet 43:520–526

Lawes Agricultural Trust (Rothamsted Experimental Station) (2005) GenStat® for Windows™, 8th edn. VSN International Ltd., Oxford

Magarey RC, Croft BJ Bull JI, Greet AD, James A (2005) The resistance of the Australian sugarcane germplasm to diseases in Queensland. Proc Aust Soc Sugar Cane Technol 27:199–210

Panstruga R, Buschges R, Piffanelli P, Schulze-Lefert P (1998) A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res 26:1056–1062

Parisseaux B, Bernado R (2004) In silico mapping of quantitative trait loci in maize. Theor Appl Genet 109:508–514

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Remmington DL, Thornsberry JM, Matsuola Y, Wilson LM, Whitt SR et al (2001) Sturcture of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Ridley AW, Dhileepan K, Johnson KN, Allsopp PG, Nutt KA, Walter GH, Croft BJ (2006) Is the distribution of Fiji leaf gall in Australian sugarcane explained by variation in the vector Perkinsiella saccharicida? Australas Plant Pathol (in press)

Rott P, Davis MJ (2000) Leaf scald. In: Rott P, Bailey RA, Comstock JC, Croft BJ, Saumtally AS (eds) A guide to sugarcane diseases. CIRAD and ISSCT, Montpellier, pp 38–44

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielson D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping accounting for multiple levels of relatedness. Nat Genet 38:203–208

Wei F, Wing RA, Wise RP (2002) Genome dynamics and evolution of the *Mla* (Powdery Mildew) resistance locus in barly. Plant Cell 14:1903–1917