Theor Appl Genet (2002) 105:1027–1037
DOI 10.1007/s00122-002-1047-5

J.-Y. Hoarau · L. Grivet · B. Offmann
L.-M. Raboin · J.-P. Diorflar · J. Payet
M. Hellmann · A. D'Hont · J.-C. Glaszmann

Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTLs for yield components

Received: 4 June 2001 / Accepted: 17 January 2002 / Published online: 19 September 2002

Abstract The genetics of current sugarcane cultivars (*Saccharum* spp.) is outstandingly complex, due to a high ploidy level and an interspecific origin which leads to the presence of numerous chromosomes belonging to two ancestral genomes. In order to analyse the inheritance of quantitative traits, we have undertaken an extensive Quantitative Trait Allele (QTA) mapping study based on a population of 295 progenies derived from the selfing of cultivar R570, using about 1,000 AFLP markers scattered on about half of the genome. The population was evaluated in a replicated trial for four basic yield components, plant height, stalk number, stalk diameter and brix, in two successive crop-cycles. Forty putative QTAs were found for the four traits at $P = 5 \times 10^{-3}$, of which five appeared in both years. Their individual size ranged between 3 and 7% of the whole variation. The stability across years was improved when limiting threshold stringency. All these results depict the presence in the genome of numerous QTAs, with little effects, fluctuating slightly across cycles, on the verge to being perceptible given the experimental resolution. Epistatic interactions were also explored and 41 independent di-genic interactions were found at $P = (5 \times 10^{-3})^2$. Altogether the putative genetic factors revealed here explain from 30 to 55% of the total phenotypic variance depending on the trait. The tentative assignment of some QTAs to the ancestral genomes showed a small majority of contributions as expected from the ancestral phenotypes. This is the first extensive QTL mapping study performed in cultivated sugarcane.

Keywords AFLP · Agronomic traits · Quantitative trait allele (QTA) · Polyploidy · Sugarcane

Communicated by H.F. Linskens

J.-Y. Hoarau (✉) · J.-P. Diorflar · J. Payet · M. Hellmann
CERF, BP315, 97494 Sainte-Clotilde, Ile de la Reunion, France
e-mail: cerf@guetali.fr
Fax: +33 02 62 29 05 07

L. Grivet · B. Offmann · A. D'Hont · J.-C. Glaszmann
CIRAD, TA 40/03, Avenue Agropolis,
34398 Montpellier Cedex 5, France

L.-M. Raboin
CIRAD, Station de Ligne paradis, 7 chemin de l'IRAT,
97410 St-Pierre, Ile de la Reunion, France

Introduction

The domesticated sugar-producing species *Saccharum officinarum* and the wild *Saccharum spontaneum* are the main taxa involved in the ancestry of current cultivated sugarcane. *S. officinarum* ($x = 10$) is an octoploid with $2n = 80$ (Bremer 1961; Li and Price 1967; D'Hont et al. 1998) and *S. spontaneum* ($x = 8$) displays a range of ploidy level probably varying between 6 and 12 ($2n = 48$–128) (Panje and Babu 1960; Sreenivasan et al. 1987; D'Hont et al. 1998). No close diploid relative is known for sugarcane. *S. officinarum* and *S. spontaneum* differ for many morphological traits; in particular, *S. officinarum* has few thick stalks with high sucrose content whereas *S. spontaneum* has many thin stalks yielding little or no sucrose. *S. spontaneum* has also been a donor of important attributes related to disease resistance and adaptability to various stress conditions.

Cultivars are clones propagated by stem cuttings. The first 'modern' cultivars were produced from a few interspecific crosses performed between *S. officinarum* and *S. spontaneum* at the beginning of the century, followed by a series of backcrosses to *S. officinarum*. Since then, breeding has been based on the intercrossing of the best existing cultivars followed by clonal selection. Cultivars are aneuploids with a number of chromosomes varying between 100 to 130, among which approximately 10% to 25% are contributed by *S. spontaneum* (Sreenivasan et al. 1987; D'Hont et al. 1996; Piperidis and D'Hont 2001). The chromosome pairing scheme has still to be well understood. Various levels of preferential pairing have been recorded and a small proportion of disomic-like pairing behavior has been observed.

Genetic analysis of agronomic traits using molecular markers has been limited so far in sugarcane. One difficulty when addressing Quantitative Trait Loci (QTLs) detection in this high polyploid, is the construction of a saturated map covering well all homo(eo)logous chromosomes. A second difficulty is the possibility that Quantitative Trait Allele (QTA) effects may be small, anticipated from the many segregating alleles present at key loci for the investigated traits. This latter difficulty could even be greater in current cultivated materials that are the result of high selection pressure for several generations. Therefore, the study of a large population with a high number of markers is necessary, more than in any diploid or amphidiploid species. A third difficulty is the important investment required for field trials to collect agronomic data with accurate estimates. Unlike many grasses, which appear to have distinct tillering and elongation stages, sugarcane has stalks that elongate while new tillers are still being formed. Thus a single plant will have tillers of many different heights and maturity stages (Lingle 1998). Moreover competition for light in sugarcane is very important in small plots (Tovey et al. 1973). Therefore agronomic trials have to be based on numerous plants per clone (across plots and replicates) to provide data of sufficient accuracy.

First studies on QTL detection in sugarcane were published by Sills et al. (1995) and Guimaraes et al. (1997). They were however based on small mapping populations (44 and 100 individuals, respectively) and map coverage was sparse, particularly in Sills et al. (1995). An important contribution toward genetic dissection of sugar stalk content in ancestral species, *S. officinarum* and *S. spontaneum*, was recently published by Ming et al. (2001) who initiated a systematic candidate-gene approach. However each genotype was evaluated based on a limited number of plants (one plant replicated three times).

Although a specific theory for QTL detection in polyploids recently started to be developed (Doerge and Craig 2000), no tools are yet available to systematically investigate QTLs when discrepancies for marker coverage are high between homologous chromosomes, and when the number of homologous chromosomes is variable from one homology class to another and pairing is irregular, as is the case in a typical sugarcane cultivar genome. The only possible approach remains to investigate the presence for QTAs, individually, on each chromosome. In that perspective, single dose markers (Wu et al. 1992) are the most useful. We tried to explore Mendelian factors underlying yield components in elite sugarcane material. For doing so, we constructed a map with AFLP markers based on a population of 295 progenies derived from the selfing of cultivar R570 (Hoarau et al. 2001). This cultivar is highly valuable for many traits and is the leading commercial variety in Réunion, Guadeloupe, Mauritius and in Vietnam. The map was constructed with 939 simplex markers, of which 887 were assembled into 120 cosegregation groups (CGs) whose cumulated genetic length was 5,850 cM. Despite the extensive genotyping efforts, large chromosome segments, specifically those inherited from *S. officinarum*, were left uncovered by markers. We estimated that the map coverage may be around one-third of the total genome of R570. However, the coverage efficiency in a search for QTLs was estimated to be about half of the genome, when taking into account a window of 10 cM on both sides of unsaturated CGs (about a hundred) and unlinked markers.

The specific purpose of the present study was to investigate the feasibility of detecting marker-trait associations for important agronomic traits in elite materials, by estimating the number, the distribution and the magnitude of QTA effects which could emerge from an extensive experiment, comparing the QTAs identified across two successive crop-cycles and examining the results with respect to the presumed species origin of the QTAs.

Materials and methods

Agronomic trial and field data

A segregating population of 295 clones derived from the selfing of cultivar R570 was planted on July 1994 in a randomized complete block design with five replicates at Centre d'Essai de Recherche et de Formation (CERF), Le Gol Station, Réunion (lowland tropical environment, 20 m elevation, 21°S). Each clone was planted in a basic plot of a 1.5-m row with four cuttings. Rows were spaced 1.5-m apart. Standard commercial cultivation practices were used.

Sugar content as estimated by a brix value (percent of soluble solids), stalk diameter, stalk length and stalk number were recorded on each individual plot, a few days prior to harvest in plant canes (June 1995) and in the first ratoon crop (July 1996). Brix (BR) was measured with a hand refractometer on the juice of a sampling punch taken at half-height of the stalk. Five stalks randomly chosen were considered to estimate BR. The number of millable stalks (NS) was counted for the whole row-plot. Five stalks per plot were chosen at random to evaluate stalk diameter (SD) and stalk length (SL). SD was recorded at mid-length of the stalk. SL was measured from ground level to the last visible dewlap.

AFLP genotyping

The AFLP genetic mapping experiment is reported in a companion paper (Hoarau et al. 2001). A set of 1,180 polymorphic markers were produced using 37 AFLP primer combinations. Among them 939 were found to be simplex (3:1). Only these markers were exploited for the map construction. The remaining 241 duplex, triplex and higher multiplex markers were not considered for map construction. However, both simplex and multiplex markers were used for QTL detection. The nomenclature for marker names, cosegregation groups (CGs) and linkage groups (LGs) are defined in Hoarau et al. (2001).

As also described in Hoarau et al. (2001), the specific origin of markers (either *S. officinarum* or *S. spontaneum*) and of chromosome segments, was assessed by comparing the banding patterns of R570 with those of a sample of *S. officinarum* clones, for most primer combinations. A marker was assigned a *S. officinarum* origin when a band was present in at least one *S. officinarum* clone, and a *S. spontaneum* origin when it was absent from all those clones. This identification procedure is possibly wrong in a number of cases, since the *S. officinarum* sample may not include all ancestors of R570, and moreover a band present in *S. officinarum* may not be specific for this species, but common to *Saccharum*

sp.. To take this into account, we developed a procedure to attribute an origin to a chromosome by integrating the information over the adjacent markers. This may introduce an apparent local discrepancy between the origin of a chromosome segment and the origin of a given marker included in it; in this case, it is the segment origin that is taken into account for general statistical computations.

Analyses of quantitative traits

Analyses of variance were performed for each attribute and each crop-cycle on the basis of the following model:

$$P_{ij} = \mu + G_i + B_j + e_{ij},$$

where P_{ij} is the phenotypic observation of progeny individual genotype G_i (i = 1,…,295) in block B_j (j = 1,…,5) μ is the overall mean and e_{ij} is the residual error. Block effect was considered as fixed and genotype effect as random. For each trait and each year, an estimate of the genetic variance σ_g^2 was inferred from the expected mean squares of the corresponding ANOVA computed by SAS (SAS 1990). Broad-sense heritabilities at the experimental design level (h^2) were determined from the ratio between genetic (σ_g^2) and phenotypic (σ_g^2) variance, with $\sigma_p^2 = \sigma_g^2 + \sigma_e^2/5$, where σ_e^2 is the error variance. Normality of traits (mean value over the five replications) were examined with the test of Shapiro and Wilks available in the UNIVARIATE procedure of SAS. Genetic correlations (r_g) between traits within each crop were determined by:

$$r_g = \sigma_{XY}/\sigma_X\sigma_Y,$$

where σ_{XY}, σ_X^2 and σ_Y^2 indicate the genetic covariance between X and Y and their respective genetic variance. Genetic covariance was calculated according to Gallais (1990) from an analysis of covariance computed with the MANOVA procedure of SAS, where the total sum of products is partitioned into components according to the sources of variation, namely genotypes, blocks and error.

In order to test the effect of the crop-cycle C_k (k = 1995, 1996), and the interaction between crop-cycle and genotype (G_i C_k), the following model was used:

$$P_{ijk} = \mu + G_i + B_{j/k} + Y_k + G_iC_k + e_{ijk}.$$

The crop-cycle effect corresponds to the physiological difference of the plant growth separating the plant crop (grown from original stalk cuttings) and the first ratoon crop (derived from a rhizomatous plateau). However, in our experiment, crop-cycle is confounded with the year effect (climatic environment).

QTL Analysis

QTL detection was performed for each marker by using a one-way ANOVA to test the contrast for a given trait between individuals that have and those that do not have the marker. Both simplex markers, yielding "S-QTAs", and multiplex markers, yielding "M-QTAs", were used. For significant associations, the proportion of the total phenotypic variance explained (R^2) was calculated using the sums of squares provided by the ANOVA. The QTA effect was calculated as the difference between the mean of progenies that have the marker and the mean of progenies that lack it. The choice of the threshold for type-I error was made while taking into account the number of tests performed but also the small effect anticipated for QTAs. If we consider the 939 simplex markers and the 241 multiplex markers for which a test should be performed, and if we suppose that a global type-I error of 0.05 is desired, the Bonferoni procedure (Rice 1989) would give a threshold of around 5×10^{-5} for each individual test. This is very conservative given the numerous markers that are closely linked. Moreover the global type-I error should be increased to keep the type-II error at a reasonable level considering the expected small effect of QTAs and the exploratory purpose of the present study. We thus performed a first round of analyses by using an individual threshold of 0.005. We then performed a second round of analyses for those markers that had been identified and a modified threshold corresponding to $P = 0.05$.

The number of expected false QTAs can be estimated according to the P value selected. We can roughly estimate the number of independent segregating factors along the portion of the genome covered by markers by dividing the total map length, about 9,000 cM after extending each map bit by 10 cM at both ends, by a distance of 30 cM, separating two independent segregating factors. This leads to a number of 300 virtual independent markers. For an experiment on eight traits and a P value of 0.005, this leads to a number of 12 ($300 \times 8 \times 0.005$) artifactual QTAs. Note that this number varies only between 7 and 18 if one takes between 20-cM and 50-cM as the average distance between independent markers.

In order to investigate the portion of genetic variance that could be explained by possible epistasis, we tested digenic linear × linear interactions between simplex markers, in the search for pairs of interacting factors, termed "int-QTA pairs". However, the highly unbalanced theoretical segregation for any two independent markers M1 and M2 [166 (M1M2):55 (M1–):55 (–M2):18 (–)] does not permit an evenly accurate estimation of the four class means. Therefore, in order to avoid spurious interactions, we retained for analysis only the 675 simplex markers having less than 50 missing molecular data items, which resulted in examining 227,475 possible pairs. The atypical format of the map we use (many CGs, unsaturated map very unevenly covered by markers) did not allow us to simply define a threshold for a type-I error. To be homogeneous, the threshold of the type-I error for interaction was chosen as the square of the basic previous threshold, $P = (0.005)^2 = 0.000025$.

Multiple regression was used to determine the contribution of the whole set of QTAs to the phenotypic variance.

Results

Quantitative trait analysis

The global quantitative genetic features are summarized in Table 1. The clone effect was highly significant ($P < 0.0001$) for all traits analyzed in both the plant crop and the first ratoon. Broad-sense heritabilities were generally high, indicating good control of the environmental error. They are comparable to broad-sense heritabilities reported in similar experiments with less replications, but with a larger entry mean basis due to either a multi-location estimate (Gravois and Milligan 1992; Milligan et al. 1990a) or to a multi-crop estimate (Kang et al. 1983; Milligan et al. 1990a).

Trait distributions are given on Fig. 1. The distribution range shifted in the second year toward higher values for BR, and toward lower values for SL and SN due to more adverse conditions that affected the vigor of plants but was more favorable for ripening. The distribution could be considered normal for SD and SL in both years and for BR in 1996 according to the Shapiro and Wilk test at $P = 0.05$, but was not for the other traits.

The phenotypic correlation for a single trait between the years was 0.70 or more. Genetic correlations between traits within years are also reported. The negative correlation between SD and SN is in agreement with numerous studies (Brown et al. 1969; Mariotti 1971, Kang et al. 1983; Milligan et al. 1990b; Singh et al. 1995; Sunil and

Table 1 Summary statistics from analyses of variance and covariance: correlations between traits are given, as well as broad-sense heritabilities (h^2) and genetic variance (σ_g^2) for each trait/crop

Traits	Traits[a]				h^2		σ^2_g			
	BR	SL	SD	SN	1995	1996	1995	1996	Units	Mean
BR	0.69	0.30	–0.11	0.35	0.81	0.77	1.37	0.85	$(\%)^2$	1.11
SL	0.15	0.79	0.05	0.17	0.83	0.81	309	411	$(cm)^2$	360
SD	–0.27	0.00	0.80	–0.56	0.91	0.84	4.7	3.3	$(mm)^2$	4
SN	0.46	0.47	–0.40	0.78	0.81	0.86	14.7	25.3	$(m^{-1})^2$	20

[a] Genetic correlations between traits in 1995 and in 1996 are given above and below the diagonal, respectively. On the diagonal is indicated the phenotypic correlation of a trait between years. Code attributes: BR, brix; SL, stalk length; SD, stalk diameter; and SN, stalk number

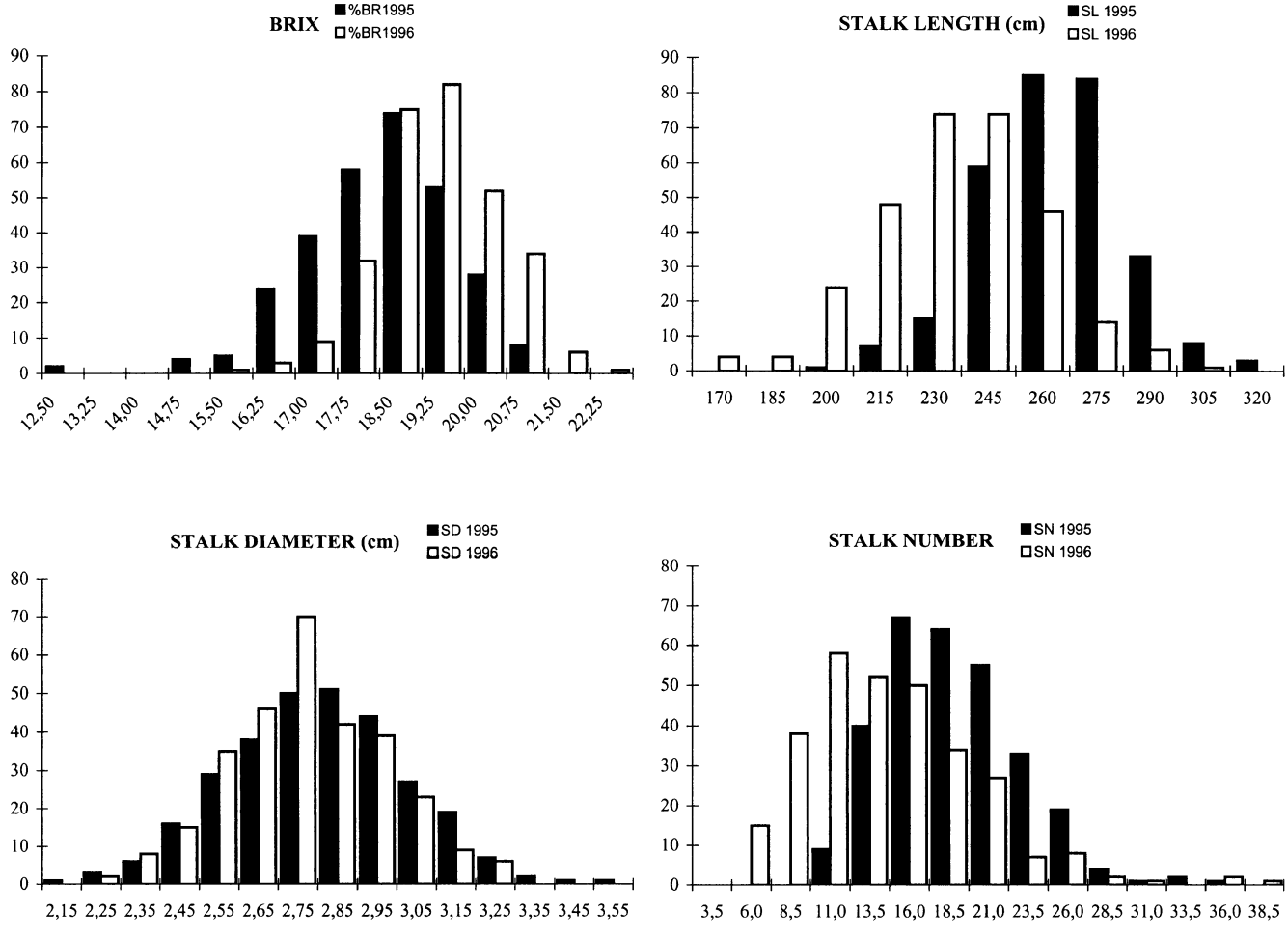

Fig. 1 Distribution of brix, stalk length, stalk diameter and stalk number among 295 progenies (individual phenotypic means over five replications)

Lawrence 1996; Sukchain et al. 1997). It may reflect a partial retention within modern hybrid material of the well-known contrasted association of SD and SN observed within each ancestral species. The positive correlation between BR and SN differed from most previous reports, since this correlation is usually either negative (Brown et al. 1969; Sunil and Lawrence 1996; Suckchain et al. 1997) or almost null (Tai et al. 1980; Kang et al. 1983; Milligan et al. 1990b; Singh et al. 1995). This unusual positive correlation is echoed by weaker, but also positive, correlations of the two traits with stalk length. This suggests that a variation existed among the progeny for a general speed of growth and development, that has an impact on the three traits; less vigorous clones would then be opposed to more vigorous clones for all traits. This would build up positive correlations between the traits, which can be different from those correlations that describe the potential ultimate morphology of the clones when they have reached their full development. This has to be taken into account for QTL data interpretation.

The effect of the crop-cycle was highly significant ($P = 0.0001$) for the four traits. Interaction between genotype and year was also highly significant ($P = 0.0001$) for all traits.

The magnitude of the genetic variances σ_g^2 presented in Table 1 is between those observed in an advanced stage of a selection program (Milligan et al. 1990a; Gravois and Milligan 1992) and those observed in a large *Saccharum* sp. germplasm collection (Sunil and Lawrence 1996). The genetic variances observed here remain in the range observed for biparental crosses, although in its lower part, when compared with various earlier studies (Mariotti 1971; Hogarth et al. 1981; Kang et al. 1983; Singh et al. 1995).

Simplex marker-trait associations

The marker-trait associations significantly detected at $P = 0.005$ are listed in Table 2. Considering all traits measured for each crop cycle, a total of 126 associations were found significant at $P = 0.005$, of which 48 (38%) were also significant at $P = 0.001$ (Tables 2 and 3). The 126 associations corresponded to 92 different markers, 88 of which were distributed over 27 cosegregation groups (CGs) while the remaining four markers were unlinked (Table 2). A single significant marker-trait association was detected for 14 of the 27 CGs, whereas several were significant (from 2 up to 13) for the other 13 CGs.

Marker-trait associations involving closely linked markers probably reflect the effects of the same S-QTA, although the presence of several S-QTAs can not be excluded when the markers are scattered over a large genetic length, as in CGs 4, 15 and 29. For the sake of simplification, we chose to consider a single S-QTA per CG for a given trait; 45 S-QTAs could be recorded at $P = 0.005$ (Tables 2 and 3). The number of S-QTAs per trait and year varies between 3 and 9. As earlier estimated, the threshold chosen probably allowed around 12 artifactual S-QTAs among these 45.

The proportion of the total phenotypic variation explained by a single marker (R^2) was generally small and ranged from 3 to 7%. Assuming a maximum of one S-QTA per CG, the distribution of the magnitude effect of the 45 S-QTAs displayed the typical asymmetrical shape skewed towards a QTA of small effect: two S-QTAs exhibited an effect of 7% and two others an effect of 6%, while 11, 10, and 20 S-QTAs had respectively an R^2 of 5%, 4% and 3%. All markers showing a R^2 of 6% or 7% were significant at $P = 0.001$. This was also the case for most markers exhibiting an effect of 5% (16/20) or of 4% (25/35), but none of the markers showing an effect of 3% were significant at this threshold.

Among the 45 putative S-QTAs detected, 21 could be assigned to a Linkage Group (LG) (Hoarau et al. 2001). Among these 21, two S-QTAs for SD were on LG I, two S-QTAs for SL were on LG VIII and three S-QTAs for SN and two S-QTAs for SL belonged to LG X. These S-QTAs located on a common LG might correspond to different alleles of the same locus, although the use of AFLP does not allow the map alignment that would be necessary for testing this hypothesis.

Eighteen markers mapping on seven distinct CGs were found involved in associations with two characters in at least 1 of the 2 years. The direction of the contribution of each CG for the two characters was always in accordance with the sign of the genetic correlation between the two traits (Table 1). Among these cases, three participate in the usual correlations: CGs 29 and 61, that decrease SD and increase SN, and CG 63 that increases SD and decreases SN. The other four may most likely relate to a general vigor or earliness: CG4, that increases SN and BR; CGs 15 and 28, that increase SL and SN; CG 74, that decreases SD and SL.

Analyses of multiplex markers

Possible associations between the 241 multiplex markers and the four traits investigated were tested by ANOVA, as for simplex markers. Considering all the traits, 12 multiplex markers were detected at $P = 0.005$, among which 11 were found associated with one trait and one with two traits. The segregation ratio fitted a duplex hypothesis (ratio ranging from 15:1 to 35:1, depending on the hypothesis for chromosome pairing) for nine markers and both duplex and triplex hypotheses (ratio higher than 63:1) for the three others. We tentatively gave a map position to the 12 multiplex markers by testing linkage with the simplex markers of the map (threshold were LOD = 5 and $\theta = 0.35$ to limit the risk of spurious linkage). Two, three and seven multiplex markers appeared to be linked to 3, 2 and 1 region(s) respectively.

Among the 19 M-QTAs thus revealed, 17 were found to be independent. Three of them had already been detected as bearing a S-QTA for the same trait, thanks to simplex marker information, thus indicating that in these particular cases multiplex and simplex markers may reveal the same QTA. The 12 multiplex markers associated with traits at $P = 0.005$ add at least 14 (17–3) new putative genetic factors (Table 3).

Digenic interactions

A total of 41 independent putative interactions (Table 3), that are 41 int-QTA pairs, were significant at $P = (0.005)^2$, among which 33 were detected in one crop-cycle and four others in both crop-cycles. From 2 to 9 independent interactions were revealed for a given trait/crop-cycle, which is in the same range as the number of S-QTAs detected per trait/crop-cycle (Table 3). Fourteen of these 41 interactions were still significant at $P = 10^{-5}$, but none of them were significant at $P = 10^{-6}$. A significant S-QTA for the same trait had already been detected nearby 6 of the 82 markers involved in the 41 interactions. The remaining 76 markers were not linked to any S- or

Table 2 Significant associations between simplex markers and agronomic traits at $P = 0.005$. When an association was significant for a given year, the association for the other year was reported up to $P = 0.05$. The corresponding percentage of variance (R^2) is also given, as well as the phenotypic effect attributed to an individual marker

CG[a]	Marker	Origin[b]	Distance cM[c]	Stalk diameter (cm)						Stalk length (cm)						Stalk number (m^{-1})						Brix (%)					
				1995			1996			1995			1996			1995			1996			1995			1996		
				Pd	R^2	Effect	Pd	R^2	Effect	Pd	R^2	Effect	Pd	R^2	Effect	Pd	R^2	Effect	Pd	R^2	Effect	Pd	R^2	Effect	Pd	R^2	Effect
1[I4]	agcctc8	Os	35.1	***	3%	–0.098	*	2%	–0,067																		
2[I9]	acccat23	Oo	6.7	***	3%	+0,085																					
	aggcat4	Os	7	***	3%	+0,085	*	1%	+0,047																		
	actctg3	Oo	7.8	***	3%	+0,087																					
	aaccac27	O–	15.6	***	3%	+0,084																					
	accctt6	Ss	57.8							*	2%	–6,61	***	3%	–8.96												
4[III1]	acactaK	Oo	103.9													***	3%	+1,53	*	2%	+1,47						
	accctg11	S–	127.2													****	4%	+1,86	***	3%	+1,95						
	agccat6	So	133.7													*****	4%	+1,95	***	3%	+2,16						
	aggcat8	Ss	136.3													******	5%	+2,14	****	4%	+2,33						
	aggctg9	Ss	137.6													******	6%	+2,28	*****	5%	+2,50						
	aagctc4	S–	140.4													*****	5%	+1,99	***	3%	+2,02						
	accctg16	So	149.7													******	6%	+2,20	*****	4%	+2,46	*	2%	+0,36	****	4%	+0,45
	accctg25	Ss	152.8													*****	5%	+2,10	*****	4%	+2,44	*	1%	+0,34	*****	5%	+0,51
	acgcta17	Ss	155.2													******	7%	+2,38	******	5%	+2,71	*	2%	+0,38	****	5%	+0,51
7[III2]	aagcac4	Oo	24.4							*	2%	+7,43	***	3%	+10,92												
8[IV1]	aggctg14	S–	4.7				****	4%	+0,099																		
	aagcag28	Ss	7.7	*	2%	+0,072	*****	5%	+0,110																		
	agcctc24	So	9.9				****	4%	+0,102																		
	aggcat20	S–	13.7	*	2%	+0,079	******	6%	+0,112																		
	aagcaa18	S–	17.5	*	2%	+0,069	*****	4%	+0,094																		
	aagctt24	S–	22.4	*	1%	+0,064	******	5%	+0,106																		
	agcctt13	Ss	25.2				*****	4%	+0,097																		
	acccaa21	Ss	32.2				****	3%	+0,089																		
14[VIII1]	actcac27	So	82.1										****	6%	+10,83												
	aggctaF	Ss	117.7										***	5%	+10,09												
15[VIII2]	acactg6	–	12.3										***	3%	+8,52												
	agccag2	Ss	13.7							*	2%	+5,43	***	3%	+8,77												
	aggctc21	Ss	16.1							*	2%	+5,64	***	3%	+9,34				***	3%	+2,12						
	acgcta4	Ss	16.4							*	1%	+5,16	***	3%	+9,38				***	3%	+2,06						
	aagctc19	So	17.8							*	2%	+5,95	****	4%	+10,00	*	2%	+1,22	*****	4%	+2,51						
	acccaa22	Ss	24.4							*	2%	+5,96	***	3%	+9,36												
	aagcac22	S–	31							**	2%	+6,87	***	3%	+9,36												
	accctg17	So	36.7													*	2%	+1,16	***	3%	+2,08						
	agccta4	S–	40.3							**	2%	+6,48	*****	5%	+10,78	*	1%	+1,11	****	4%	+2,39						
	aagcag13	Ss	41.9							*	1%	+4,89	***	3%	+8,30				***	3%	+2,16						
	agcctt5	S–	43.6																***	4%	+2,24						
	actctt5	So	47.5							*	2%	+5,90	***	3%	+8,93				***	3%	+2,11						
	acccac25	Ss	76.4							*	2%	+6,83	***	3%	+10,12												
	aagcta7	S–	84.6							*	2%	+7,88	***	4%	+12,18												
	actcat5	So	95							*	2%	+7,67	***	3%	+10,98												
22[IX4]	agccaa11	So														**	3%	–1,59	***	4%	–2.08						

Table 2 (continued)

CG[a]	Marker	Origin[b]	Distance cM[c]	Stalk diameter (cm)						Stalk length (cm)						Stalk number (m^{-1})						Brix (%)					
				1995			1996			1995			1996			1995			1996			1995			1996		
				P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect
28[X2]	Aaccac15	S–	132.5							***	3%	+7,96	***	3%	+9,75				***	3%	+2,46						
	aggcat12	S–	141													*	2%	+1,53	****	4%	+2,61						
29[X3]	actctt25	Ss	4.3																			***	3%	–0.55	*	2%	–0,32
	aggcat16	S–	53.9	***	3%	–0.086	*	2%	–0,062							*	2%	+1,44	***	3%	+2,39						
	aaccac13	S–	55.7	***	3%	–0.091	**	2%	–0,063							*	2%	+1,44	***	4%	+2,43						
	aagctt36	So	56.9	***	3%	–0,090	**	3%	–0,066							*	2%	+1,41	***	3%	+2,32						
	accctg24	Ss	61.6	****	4%	–0,105	****	4%	–0,086							*	2%	+1,40	***	3%	+2,18						
–	accctg14	Ss	64.9	***	4%	–0.102	***	4%	–0.085																		
	acccat16	Oo	67.6	***	3%	–0.092	*****	5%	–0,095																		
	acccaa14	Oo	71.1	***	3%	–0.089	*	2%	–0,061																		
	acactc6	Oo	78.3	*****	4%	–0,104	***	3%	–0.072																		
	aagcaa20	O–	98.3	****	4%	–0,105	***	3%	–0.082																		
30[X12]	accctt14	Oo	24.2							*	2%	+7,27	***	3%	+9,91												
	aagcat3	Oo	28.3							*	2%	+5,84	*****	5%	+11,13												
	acgctt4	Oo	34										***	3%	+8,75												
31	aggcat10	Oo														***	3%	–1.76									
35	accctc14	Oo																				***	3%	+0,54	***	3%	+0,45
36	actcac31	Oo											***	3%	+9,26												
37	accctt16	–																				***	4%	–0.57	***	3%	–0.42
48	acgctt24	–														*	2%	+1,22	****	5%	+2,46						
51	aggcta13	Oo								*	2%	+5,67	***	3%	+9,27												
61	acgcag9	Oo		****	5%	–0,101										***	4%	+1,71									
63	actcac25	Oo	8.7				***	3%	+0,085																		
	aagctt39	O–	42.9				******	7%	+0,113										***	5%	–2.38						
	aggctg1	Oo	43.8	*	1%	+0,068	*****	4%	+0,100							*	2%	–1,39	******	5%	–2,96						
	aagctc1A	Oo	60.1																***	4%	–2.47						
	agccag11	Oo	61.9																***	3%	–2.35						
	acgcag17	Oo	71													*	2%	–1,29	*****	4%	–2,67						
67	actcat20	Ss	4.9																						***	3%	–0.41
	agccta5	Ss	7.1																						***	3%	–0.39
70	actctt4	Oo																				*	2%	–0,41	***	3%	–0.42
74	accctgL	Ss	31.8	***	3%	–0.093	*	2%	–0,064	***	3%	–8.74															
	accctg18	Ss	35.8	***	3%	–0.092																					
	acccaa10	S–	39.7	***	3%	–0.092																					
	aagctc11	Ss	43.8	****	4%	–0,104	*	2%	–0,058																		
	agcctc7	Ss	45	****	4%	–0,106	*	2%	–0,058																		
	acgcta20	Ss	46.9	****	4%	–0,109	*	2%	–0,065																		
	aagcaa16	–	49.6	***	3%	–0.096	*	2%	–0,054																		
77	accctg13	Oo																				*	2%	+0,38	*****	4%	+0,48

Table 2 (continued)

CG[a]	Marker	Origin[b]	Distance cM[c]	Stalk diameter (cm)						Stalk length (cm)						Stalk number (m^{-1})						Brix (%)					
				1995			1996			1995			1996			1995			1996			1995			1996		
				P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect	P[d]	R^2	Effect
82	acactg39	Oo	0	****	4%	+0,125	*	2%	+0,076																		
	actctgG	Oo	23	***	3%	+0,098																					
86	aagcac20	–														***	3%	–1.72	*	2%	–1,72						
111	aggctc14	So	22.6				***	3%	–0,080																		
	aagctt11	Ss	25.9				***	4%	–0,082																		
	acgctt16		30.3				***	5%	–0,089																		
116	acactt13									*	1%	–5,32	***	3%	–8.86							*	2%	+0,45	***	3%	+0,45
Unlinked acacta4		–o																				*	2%	–0,41	***	3%	–0.46
Unlinked aagcag15		–o																									
Unlinked accctgT										****	5%	+10,06	*	2%	+8,78												
Unlinked agccaa23		–o								***	5%	+10,76															

[a] In brackets the corresponding RFLP CG is given, the Roman numeral being indicative of the homology group number
[b] In lower-case:simple marker origin. In capitals:multi-marker origin of the chromosomic segment including the marker (Hoarau et al. 2001) 'o' is for *S. officinarum* and 's' is for *S. spontaneum*
[c] Cumulated distance from the upper distal marker of the CG
[d] Level of significance for marker-trait association : 0.05,*; 0.01,**; 0.005,***; 0.001,****; 0.0005,*****; 0.0001,******

M-QTA.The R^2 of interactions ranged between 6% and 8%.

Given the few CGs of the map already assigned to a LG (Hoarau et al. 2001), the identity of the LG of both interacting markers could be established in four cases out of 41. In all these four instances the two LGs involved were different (LGs VIII and X in three cases, and LGs II and IX in one case), thus showing an interaction of epistatic nature among alleles belonging to different loci.

Consistency of QTA detection across years

The detection of significant genetic effects appears to be relatively inconsistent between the two crop-cycles, whatever the type of QTA detected, since most marker-trait associations that were significant for one crop-cycle failed to be detected for the other at the same threshold. Indeed, as summarised in Table 3, stability across crop-cycles was observed for only five S-QTAs represented by 16 simplex-trait associations (Table 2), one M-QTA, at $P = 0.005$, and four int-QTA pairs, at $P = (0.005)^2$. In comparison 35 S-QTAs, 11 M-QTA and 33 int-QTA pairs were revealed in one of the two crops only. The stability of the S- or M-QTAs was further investigated by decreasing the stringency and allowing associations with lower significance ($P = 0.05$) to appear. This investigation was successful for 26 S-QTAs (located on 23 CGs and three unlinked markers) and failed for the nine others (located on eight CGs and one unlinked marker) (Table 2). For the multiplex markers, a stability across years was recovered for 6 out of 12 M-QTAs (data not shown). It is remarkable that the associations were always in the same direction between the 2 years, even when the marker-trait associations were not significant at $P = 0.05$ (data not shown).

The same is true for the digenic interactions. The direction of the effect was the same in the two crop-cycles, not only for each of the four pairs of int-QTAs that were perceptible in both years, but also for all the 33 other pairs detected once (data not shown).

QTA contribution to genetic variance

Table 3 gives the R^2 calculated by multiple regression for different models including either the simplex markers alone, the simplex and the multiplex markers, the di-marker interactions, or the three types together. Depending on which trait/year is considered, the S-QTAs detected at $P = 0.005$ altogether explained between 9% ($df = 261$) and 28% ($df = 179$) of the genetic variation. The smallest and the highest values reflected the segregation of three and seven putative S-QTAs, respectively. The addition of the multiplex markers in the regression model provided a slight R^2 increase, not higher than 10%. On average the amount of variation explained per QTA is 3%.

Table 3 Number of S-QTAs, M-QTAs and int-QTAs detected for yield components and the estimated proportion of genotypic variance explained. Code attributes are: SD, Stalk Diameter; SL: Stalk Length; SN : Stalk Number; BR: Brix

Item	SD (cm)		SL (cm)		SN		BR(%)		Total
	1995	1996	1995	1996	1995	1996	1995	1996	
Associations detected									
Simplex markers[a]	24(8)	19(12)	4(1)	24(4)	12(8)	29(11)	3(0)	11(4)	126(48)
Simplex markers common to both years[a]	5(1)		1(0)		8(5)		2(0)		16(6)
Deduced S-QTAs[ab]	6(4)	4(3)	4(1)	9(3)	4(1)	7(5)	3(0)	8(2)	45(19)
Deduced S-QTAs common to both years[ab]	1(1)		1(0)		1(1)		2(0)		5(2)
M-QTAs[a]	3(0)	5(1)	0(0)	3(0)	1(0)	0(0)	1(0)	1(0)	14(1)
M-QTAs common to both years[a]	1(0)		0(0)		0(0)		0(0)		1(0)
Int-QTA pairs[c]	2(0)	2(1)	5(2)	2(1)	8(3)	8(2)	9(2)	6(3)	41(14)
Int-QTA pairs common to both years[c]	1(0)		0(0)		2(0)		1(0)		4(0)
R^2 of a multiple regression[d]:									
(1) S-QTAs [e]	17%	19%	9%	25%	14%	28%	9%	24%	
	(194)	(210)	(174)	(173)	(214)	(179)	(261)	(242)	
(2) S-QTAs and M-QTAs [e]	20%	28%	9%	29%	21%	28%	9%	26%	
	(189)	(178)	(174)	(165)	(206)	(179)	(258)	(238)	
(3) int-QTAs[f]	14%	15%	28%	13%	41%	36%	45%	40%	
	(246)	(244)	(246)	(274)	(221)	(237)	(193)	(223)	
(4) S-QTAs, M-QTAs and int-QTAs	30%	35%	40%	36%	44%	51%	46%	55%	
	(170)	(161)	(144)	(155)	(155)	(147)	(173)	(187)	

[a] Detected at $P = 0.005$ (at $P = 0.001$)
[b] One S-QTA counted per CGs or unlinked marker
[c] Detected at $P = 0.000025$ (at $P = 0.00001$)
[d] Total degre of freedom of the regression model indicated in parenthesis, representing the number of no missing molecular data across all factors
[e] All those detected at $P = 0.005$ are considered
[f] All those detected at $P = 0.000025$ are considered

The genetic variance explained by di-marker interactions ranged between 13% ($df = 274$) and 45% ($df = 193$), and was roughly proportional to the number of di-marker interactions considered in the model for each trait. It is remarkable that in five of eight cases the global effect of epistatic interactions exceeded the global effect of S- and M-QTAs (SL in 1 year and SN and BR in both years).

When cumulating all effects into a single model, the proportion of the total genetic variance explained ranged between 30% and 55%; it was close to one-third for SL and SD, and one-half for SN and BR. These estimates were calculated on the basis of more than half of the population, as indicated by the total *df* (144–187 range). The phenotypic variation was slightly better accounted for by segregating markers in 1996 than in 1995 for all characters, except SL.

QTA effect and chromosome origin

Both putative marker origin and segment origin are provided in Table 2 for the simplex markers. This information allowed us to compare the direction of the effect of a QTA (whether its presence increases or decreases the trait value) with the contribution expected from the ancestral species. If we consider the three traits that are markedly contrasted between the two ancestral species, namely BR, SN and SD, one counts 16 CGs associated with these traits whose segmental specificity is entirely of a single origin (either *S. officinarum* or *S. spontaneum* from one end to the other). Only five of these 16 cases do not respect the expected ancestral contribution, CGs 1, 8, 22 61 and 70. Besides, it is interesting to note that the three CGs (4, 15 and 28) involved in unusual correlations between BR, SN and SL that may relate to general vigor, display a *S. spontaneum* origin.

The multiplex markers, most of which have a *S. officinarum* specificity (data not shown), yielded a contradictory picture, many of them contributing to a small SD. In the origin of the simplex markers involved in di-marker interactions there appears no tendency for the relation between the origin of the markers and their involvement in epistatic (positive or negative) interactions. This denotes no particular trace of coadaptation among the genes present in the same species or, conversely, of higher heterosis between genes derived from distinct species.

Discussion

The aim of the present study was to explore the Mendelian base underlying the yield components in the high-polyploid aneuploid, poly-specific genetic background of current sugarcane cultivars. This was performed thanks to an extensive experiment associating a large progeny size (close to 300 individuals) and a high number of markers (more than 1,000 including the multiplex ones).

The trials have been successful. With heritabilities ranging between 77% and 91%, one can from sugarcane hardly expect better-results for a QTL analysis. The mapping was only partial, since about half of the genome can be considered left unmarked.

As our purpose was exploratory, we chose to limit the type-II error, i.e. the risk to fail to detect genuine QTAs, at the expense of the type-I error, i.e. the risk to keep false QTAs. With the threshold that we retained, a minimum of 45 additive S-QTAs were detected for the eight variables (four traits observed over two years) on the basis of simplex associations, and 14 M-QTAs on the basis of multiplex associations.

In addition, 41 independent digenic interactions were detected; these factors, however, have to be taken cautiously due to the highly unbalanced sizes of the genotype classes compared with S-QTA detection. This provides a concrete illustration of epistasis which is difficult to assess in sugarcane by conventional quantitative genetic approaches.

Despite the choice not to be very selective, the variance explained by all simplex and multiplex QTAs was at the most 29%. It increased notably when the di-marker interactions were taken into account, reaching sometimes more than 50% (55% for BR in 1996); the exact significance of these interactions, however, remains to be clarified.

The main explanation for the low part of the variation accounted for by the QTAs is their small individual effect. This effect ranged in size between 3% and 7% of the phenotypic variance. The average R^2 per QTA coming from the multiple regression models we used is 3%, which is exactly two-fold smaller than the one averaged by Kearsey and Farquhar (1998) on many diploid studies in an extensive review. This is probably related to the high ploidy coupled with a high heterozygozity, which implies the coexistence of many alleles at each locus. It is likely that the alleles other than the QTA at the same locus have an effect on the same trait. When in a diploid plant one looks for one factor per locus, which is the difference between the two alleles, in sugarcane one has to deal with 8 to 10 or 12 segregating factors per locus. In such a polyploid background with a predominantly polysomic inheritance, one intuitively expects to be able to tag the most (or the least) favorable allele, and only it, among the whole allelic series. This confounding effect of the high ploidy level is probably reinforced in our modern genetic background. Indeed, other studies based on F1 interspecific populations between *S. officinarum* and *Saccharum robustum* (Sills et al. 1995), or between *S. officinarum* and *S. spontaneum* (Ming et al. 2001), detected QTAs with stronger effects, in some cases twice as large as the effects that we observed. Favorable alleles have probably been concentrated into modern cultivars by recurrent breeding, thus diminishing the internal contrast that determines the trait segregation and the magnitude of the QTA effects.

The consistency of the QTAs across crop-cycles appeared rather low. This apparent fluctuation of the QTA effect is congruent with the high genotype × crop-cycle interactions detected for most traits in the present study and in various other quantitative genetic studies. However, it is remarkable that the direction of the effect of the QTAs was conserved over the 2 years in all cases, whether the effect was significant or not. Poorly overlapping sets of QTAs across the 2 years may thus reflect the statistical threshold that truncates the QTA repertoire by retaining those with the highest effect. Altogether, a minimum number of five S-QTAs have been identified for the four traits across both years. The specific origin of the S-QTAs provides a mitigated picture, with only a majority of contributions as expected from the ancestral phenotypes. This depicts the genome of a cultivar as a swarm of numerous QTAs each with a little effect. Note, however, that the map used here covers approximately half of the genome and that the above considerations may have to be revised when a better coverage has been achieved.

The strategy for improving the map coverage has already been discussed (Hoarau et al. 2001). The recent availability of a set of locus-specific markers such as microsatellites (Cordeiro et al. 2000, and the D'Hont mSSCIR1 to mSSCIR78 EMBL data base) will be of great utility. Having access to the whole genome with established homology relationships will enable us to estimate the full genotypic combination of a given plant at marker loci linked with QTLs, and thus to attempt the identification of allelic QTAs and the evaluation of gene dosage effects.

The genetic analysis for yield components can also be improved by refining the parameters that should be monitored. Indeed our study has pointed out specific difficulties related to the continuing evolution of most traits, including those in the later part of the crop-cycle. The use of crop-physiology models may suggest biologically sounder traits and parameters that may help achieve general understanding of the genetics of the crop. The same is true for brix, which should obviously be complemented by more-specific sucrose-related traits and biochemical analyses. The present study will help identify population subsamples that could be advantageously assayed with a more complete approach.

References

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

Brown AHD, Daniels J, Latter BDH (1969) Quantitative genetics of sugarcane. II. Correlation analysis of continuous characters in relation to hybrid sugarcane breeding. Theor Appl Genet 39:1–10

Cordeiro GM, Taylor GO, Henry RJ (2000) Characterisation of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploid species. Plant Sci 2:161–168

D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann J-C (1996) Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Mol Gen Gent 250:405–413

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

Doerge RW, Craig BA (2000) Model selection for quantitative trait locus analysis in polyploids. Proc Natl Acad Sci USA 97:7951–7956

Gallais A (1990) Théorie de la sélection en amélioration des plantes. Masson editions, Paris

Gravois KA, Milligan SB (1992) Genetic relationships between fiber and sugarcane yield components. Crop Sci 32:62–67

Guimaraes CT, Sills GR, Sobral BWS (1997) Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proc Natl Acad Sci USA 94:14,261–14,266

Hoarau J-Y, Offmann B, D'Hont A, Risterucci A-M, Roques D, Glaszmann J-C Grivet L (2001) Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp). I. Genome mapping with AFLP markers. Theor Appl Genet 103:84–97

Hogarth DM, Wu KK, Heinz DJ (1981) Estimating genetic variance in sugarcane using a factorial cross design. Crop Sci 21:21–25

Kang MS, Miller JD, Tai PYP (1983) Genetic and phenotypic path analyses and heritability in sugarcane. Crop Sci 23:643–647

Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants; where are we now? Heredity 80:137–142

Li HW, Price S (1967) Chromosome numbers of noble sugarcane clones. Proc Int Soc Sugar Cane Technol 12:884–886

Lingle SE (1998) Sugarcane. In: Smith DJ, Hamel C (eds) Crop yield, physiology and processes. Springer, Berlin Heidelberg, pp 287–310

Mariotti JA (1971) Associations among yield and quality components in sugarcane hybrid progenies. Proc Int Soc Sugar Cane Technol 14:297–302

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

Milligan SB, Gravois KA, Bischoff KP, Martin FA (1990b) Crop effects on genetic relationships among sugarcane traits. Crop Sci 30:927–931

Ming R, Liu S-C, Moore PH, Irvine JE, Paterson AH (2001) QTL analysis in a complex autopolyploid: genetic control of sugar content in sugarcane. Genome Res 11:2075–2084

Panje RR, Babu CN (1960) Studies in *Saccharum spontaneum*. Distribution and geographical associations of chromosome number. Cytologia 25:152–172

Piperidis G, D'Hont A (2001) Chromosome composition analysis of various *Saccharum* interspecific hybrids by genomic in situ hybridization (GISH). Proc Int Soc Sugarcane Technol 24: 565–566

Rice WR (1989) Analyzing tales of statistical tests. Evolution 43:223–225

SAS Institute Inc (1990) SAS procedures guide, version 6, 3rd edn. SAS Intitute Inc, Cary, North Carolina, USA

Sills G, Bridges W, Al-Janabi S, Sobral B (1995) Genetic analysis of agronomic traits in a cross between sugarcane (*Saccharum officinarum* L.) and its presumed progenitor (*S. robustum* Brandes & Jesw. Ex. Grassl). Mol Breed 1:355–363

Singh RK, Singh S, Singh GP (1995) Genetic variability and yield relationships in adult sugarcane. Sugar Cane 1:18–21

Sreenivasan TV Ahloowalia BS, Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, Amsterdam, pp 211–253

Sukchain, Devinder Sandhu, Saini GS (1997) Inter-relationships among cane yield and commercial cane sugar and their component traits in the autumn plant crop of sugarcane. Euphytica 95:109–114

Sunil HK, Lawrence MJ (1996) Quantitative genetics of sugarcane. I. A large-scale evaluation of Saccharum germplasm. Sugar Cane 6:3–10

Tai PYP, Miller JD, Gill BS, Chew V (1980) Correlations among characters of sugarcane in two intermediate selection stages. Proc Int Soc Sugar Cane Technol 17:1119–1127

Tovey DA, Glasziou KT, Farquhar RH, Bull TA (1973) Variability in radiation received by small plots of sugarcane due to differences in canopy heights. Crop Sci 13:240–242

Wu KK, Burnquist W, Sorrels ME, Tew TL, Moore PH, Tanksley SD (1992) The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor Appl Genet 83:294–300