

## RESEARCH ARTICLE

# A genome-wide association study identified loci for yield component traits in sugarcane (*Saccharum* spp.)

Fernanda Zatti Barreto<sup>1</sup> , João Ricardo Bachega Feijó Rosa<sup>2,3</sup> , Thiago Willian Almeida Balsalobre<sup>1</sup> , Maria Marta Pastina<sup>2,4</sup>, Renato Rodrigues Silva<sup>5</sup>, Hermann Paulo Hoffmann<sup>1</sup>, Anete Pereira de Souza<sup>6,7</sup>, Antonio Augusto Franco Garcia<sup>2</sup>, Monalisa Sampaio Carneiro<sup>1</sup> \*

**1** Departamento de Biotecnologia e Produção Vegetal e Animal, Centro de Ciências Agrárias, Universidade Federal de São Carlos, Araras, São Paulo, Brasil, **2** Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brasil, **3** Centro de Pesquisa e Desenvolvimento de Cultivares de Soja, Setor de Pesquisa e Desenvolvimento, FTS Sementes S.A., Ponta Grossa, Paraná, Brasil, **4** Centro de Pesquisa e Desenvolvimento, Embrapa Milho e Sorgo, Sete Lagoas, Minas Gerais, Brasil, **5** Instituto de Matemática e Estatística, Campus Samambaia, Universidade Federal de Goiás, Goiânia, Goiás, Brasil, **6** Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil, **7** Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Campinas, São Paulo, Brasil

 These authors contributed equally to this work.

\* [monalisa@ufscar.br](mailto:monalisa@ufscar.br)



## OPEN ACCESS

**Citation:** Barreto FZ, Rosa JRBF, Balsalobre TWA, Pastina MM, Silva RR, Hoffmann HP, et al. (2019) A genome-wide association study identified loci for yield component traits in sugarcane (*Saccharum* spp.). PLoS ONE 14(7): e0219843. <https://doi.org/10.1371/journal.pone.0219843>

**Editor:** David D. Fang, USDA-ARS Southern Regional Research Center, UNITED STATES

**Received:** January 23, 2019

**Accepted:** July 2, 2019

**Published:** July 18, 2019

**Copyright:** © 2019 Barreto et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was supported by the following institutions and/or grant programs: FINEP (Financiadora de Estudos e Projetos), FAPESP (Fundação de Amparo à Pesquisa de São Paulo - 08/52197-4 to APdS) and INCT-Bioetanol (Instituto Nacional de Ciência e Tecnologia do Bioetanol - FAPESP 08/57908-6 to MSC, and Conselho Nacional de Desenvolvimento Científico

## Abstract

Sugarcane (*Saccharum* spp.) has a complex genome with variable ploidy and frequent aneuploidy, which hampers the understanding of phenotype and genotype relations. Despite this complexity, genome-wide association studies (GWAS) may be used to identify favorable alleles for target traits in core collections and then assist breeders in better managing crosses and selecting superior genotypes in breeding populations. Therefore, in the present study, we used a diversity panel of sugarcane, called the Brazilian Panel of Sugarcane Genotypes (BPSG), with the following objectives: (i) estimate, through a mixed model, the adjusted means and genetic parameters of the five yield traits evaluated over two harvest years; (ii) detect population structure, linkage disequilibrium (LD) and genetic diversity using simple sequence repeat (SSR) markers; (iii) perform GWAS analysis to identify marker-trait associations (MTAs); and iv) annotate the sequences giving rise to SSR markers that had fragments associated with target traits to search for putative candidate genes. The phenotypic data analysis showed that the broad-sense heritability values were above 0.48 and 0.49 for the first and second harvests, respectively. The set of 100 SSR markers produced 1,483 fragments, of which 99.5% were polymorphic. These SSR fragments were useful to estimate the most likely number of subpopulations, found to be four, and the LD in BPSG, which was stronger in the first 15 cM and present to a large extension (65 cM). Genetic diversity analysis showed that, in general, the clustering of accessions within the subpopulations was in accordance with the pedigree information. GWAS performed through a multilocus mixed model revealed 23 MTAs, six, three, seven, four and three for soluble solid content, stalk height, stalk number, stalk weight and cane yield traits, respectively.

Tecnológico (CNPq) - 574002/2008-1 to MSC). FZB received a master's fellowship from CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Finance Code 001). TWAB received doctoral fellowships from FAPESP (10/50091-4). JRBFR received master fellowships from FAPESP (10/06702-9). APS and AAFG received research fellowships from CNPq. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

These MTAs may be validated in other populations to support sugarcane breeding programs with introgression of favorable alleles and marker-assisted selection.

## Introduction

Sugarcane (*Saccharum* spp.) is an important industrial crop and a vital component for food and energy security, providing sucrose, bioethanol and bioelectricity [1,2]. Sugarcane is cultivated in mainly tropical and subtropical areas and has a very high photosynthetic efficiency and a complex genome due to its variable ploidy levels, frequent aneuploidy, and large genome size of approximately 10 gigabases (Gb) [3–8]. Modern sugarcane cultivars have chromosome numbers ranging from 100 to 130, are vegetatively propagated, and result from the selection of populations derived from outcrossing heterozygous parents [8–10]. Brazil is the world's largest sugarcane producer, and its productivity increased 66% in tons of sugarcane per hectare from 1975 to 2010, partially due to the growing area expansion and improvements in agricultural practices [2,10,11].

Sugarcane breeding programs concentrate efforts to release cultivars adapted to different environments that have high yields in terms of biomass production and sucrose content as well as resistance to diseases. However, the breeding process is expensive and requires approximately 15 years of experimentation and selection to obtain one or a few cultivars. Briefly, every year, crosses between accessions generate hundreds of thousands of F1 progenies, and the individuals reaching the final stages of selection are commonly evaluated over several harvests in multi-environment trials (METs) to identify those with the potential to become new cultivars [10–13]. Even with the adoption of better agricultural practices and selection strategies in the early stages of breeding programs, which attempt to measure and isolate the environmental effects of genetic factors [13–16], the genetic gains to quantitative traits have declined in recent years for sugarcane and other crops [17–19].

Clearly, there is a need to complement the classical breeding of sugarcane with other tools, such as molecular approaches, which have been applied for other crops [20–22]. Quantitative trait locus (QTL) mapping and genome-wide association studies (GWAS) are strategies to understand the genetic architecture of complex traits and include a first step of marker-assisted selection (MAS) [4,6,19,22]. To employ these strategies in outcrossing heterozygous species, such as sugarcane, we need to consider that, for each segregating locus, different numbers of segregating alleles may exist, and the single-dose markers currently available for mapping studies show only some of the genetic information [8,23]. This limitation is more evident in the traditional QTL mapping approach, which may identify genomic regions with low resolution, usually due to the smaller amount of available markers and also limited to the genetic composition of the biparental population under study. Nevertheless, attempts to associate phenotype and genotype and the development of new data analysis strategies have been significantly advanced [23–27].

On the other hand, GWAS has been widely used to identify marker-trait associations (MTAs) in genetically diverse populations of plants [20,21,28–32]. GWAS is based on linkage disequilibrium (LD) due to physical linkage, which is reportedly extensive in sugarcane [33–36]. This LD value is assigned to a recent breeding history, characterized by a strong foundation bottleneck followed by a small number of intercrossing cycles, which significantly reduces the frequency of recombination events. The high extent of LD in sugarcane indicates that a high density of markers may not be critical for performing GWAS [36–38] and that single-

dose markers might be appropriate for this purpose; indeed, mapping models for loci with high allelic dosages are under development [26]. Although high-throughput marker systems are available, mainly for single nucleotide polymorphism (SNP) genotyping, the lack of appropriate methods for analyzing complex species such as sugarcane hinders the applicability of new molecular breeding tools [23,26,27]. In this context, single-dose markers, such as simple sequence repeats (SSRs) and target region amplification polymorphisms (TRAPs) could be used to characterize genome variation, investigate population structure and genetic diversity and thus enable GWAS [37,39–41]. In addition, despite the potential for using LD-based association studies to identify MTAs, a few studies on yield-related traits in sugarcane have been published [18,35–42].

For the latter, several algorithms and software have been developed to improve statistical power, increase computational efficiency, and reduce spurious associations in the GWAS approach [43]. Among GWAS algorithms, FarmCPU [44], which uses a multilocus linear mixed model (MLMM), is considered an efficient alternative to control for spurious associations [45–47]. Indeed, combinations of various methods for multilocus GWAS have also been used to identify causal associations and control the false positive rate [43,47,48].

In the current assignment, our objectives were to (i) estimate, through a mixed model, the adjusted means and genetic parameters of the five yield traits evaluated over two harvest years in a diversity panel composed of ancestral and modern sugarcane accessions; (ii) detect population structure, LD and genetic diversity using SSR markers; (iii) perform GWAS analysis to identify MTAs; and (iv) annotate the sequences giving rise SSR markers that had fragments associated with target traits to search for putative candidate genes.

## Materials and methods

### Plant material and phenotypic traits

In this study, 134 accessions (S1 Table) of the Brazilian Panel of Sugarcane Genotypes (BPSG) were used. BPSG is a mini core collection from the germplasm bank of RIDESA (Inter-University Network for the Development of Sugarcane Industry), and the accessions were chosen according to the following criteria: i) relevant Brazilian cultivars, ii) main parents for Brazilian breeding programs; iii) cultivars from countries that grow sugarcane; iv) cultivars used as parents in mapping programs [25,49]; and v) representatives of the *Saccharum* species complex. The BPSG accessions represent an important genetic background in Brazilians breeding programs.

The 134 accessions of BPSG were planted in a field experiment performed in 2013 at the Agricultural Science Center of the Federal University of São Carlos (UFSCar) in Araras City, São Paulo State, Brazil. Araras is located at 22°21'25"S, 47°23'3"W at an altitude of 611 m; the experimental area soil is Typic Eutroferric Red Latosol. The experimental design consisted of a randomized complete block, which was fully replicated four times. The plots consisted of two rows 3 m long and spaced 1.5 m apart. Each plot was composed of 12 presprouted seedlings at the planting of the experiment in 2013. The experimental plants were harvested when they were approximately 18 months of age during the plant cane and first ratoon. The BPSG was evaluated for five yield components: soluble solid content (BRIX, in °Brix), stalk height (SH, in m), stalk number (SN), stalk weight (SW, in kg), and cane yield (TCH, in t ha<sup>-1</sup>). Phenotypic yield trait data were collected according to Balsalobre et al. [12]. Briefly, a 10-stalk sample per plot was taken for analysis of the BRIX and SH. The weight of the 10 stalks was added to the total weight of the plot (SW) to estimate the TCH, which was calculated as the product between the SW of a linear meter and the amount of linear meters in one ha (6667 linear

meters compose one ha with a spacing of 1.5 m). The SN was estimated by directly counting the stalks in each plot.

### Statistical analysis of phenotypic data

A multiharvest mixed model produced the joint adjusted means. The analysis was conducted for each trait using GenStat 19th edition [50] based on restricted maximum likelihood (REML) and the following linear mixed model:

$$y_{imkuv} = \mu + h_m + b_{km} + g_{imk} + r_{umk} + c_{vmk} + e_{imkuv}$$

where  $y_{imkuv}$  is the phenotype of the  $i^{\text{th}}$  accession, evaluated in the  $m^{\text{th}}$  harvest, located in the  $u^{\text{th}}$  row and the  $v^{\text{th}}$  column inside the  $k^{\text{th}}$  replication;  $\mu$  is the overall mean;  $h_m$  is the fixed effect of the  $m^{\text{th}}$  harvest ( $M = 1, \dots, M; M = 2$ );  $b_{km}$  is the fixed effect of the  $k^{\text{th}}$  replication ( $k = 1, \dots, K; K = 4$ ) at the  $m^{\text{th}}$  harvest;  $g_{imk}$  is the random effect of the  $i^{\text{th}}$  accession ( $i = 1, \dots, I; I = 134$ ) at the  $m^{\text{th}}$  harvest evaluated in the  $k^{\text{th}}$  replication;  $r_{umk}$  and  $c_{vmk}$  are the random effects of the  $u^{\text{th}}$  row and  $v^{\text{th}}$  column, both evaluated at the  $m^{\text{th}}$  harvest and  $k^{\text{th}}$  replication; and  $e_{imkuv}$  is the random residual error. In addition, for the SN, SW, and TCH traits, the number of clumps per plot was included in the mixed model as a fixed covariate. Aiming to model the accession effects, the genetic variance–covariance (VCOV) matrix  $\mathbf{G} = \mathbf{G}_M \otimes \mathbf{I}_g$ , i.e.,  $\mathbf{g} \sim \mathbf{N}(\mathbf{0}, \mathbf{G})$  was considered, where  $M$  is the number of harvests, and  $\otimes$  represents the Kronecker product of both the genetic  $\mathbf{G}_M$  and identity  $\mathbf{I}_g$  matrices with the respective dimensions of  $2 \times 2$  and  $1 \times 134$ . For the  $\mathbf{G}_M$  matrix, four structures (identity, ID; diagonal, DIAG; first order autoregressive homogenous, AR1; and first order autoregressive heterogeneous, AR1(het)) were examined and compared via Akaike [51] (AIC; [51]) and Bayesian (BIC; [52]) information criteria [53]. For the residue, a structure of variance heterogeneity was assumed for the different harvests. For each trait, the fixed effects were tested using the Wald statistics test and were retained in the model if statistically significant ( $P < 0.05$ ). After the  $\mathbf{G}_M$  matrix structure selection, the adjusted means for accessions and genetic parameters for each evaluated trait were obtained. The phenotypic ( $\hat{\sigma}_p^2$ ) and genotypic ( $\hat{\sigma}_g^2$ ) variances were used for calculating heritability in the broad sense on an individual-plant basis ( $\hat{H}^2 = \hat{\sigma}_g^2 / \hat{\sigma}_p^2$ ). The  $\hat{\sigma}_p^2$  value was determined from  $\hat{\sigma}_p^2 = \hat{\sigma}_g^2 + \hat{\sigma}_e^2 + \hat{\sigma}_r^2 + \hat{\sigma}_c^2$ , where  $\hat{\sigma}_e^2$  was the residual variance,  $\hat{\sigma}_r^2$  was the variance for row effects and  $\hat{\sigma}_c^2$  was the variance for column effects [54].

### DNA extraction, molecular markers and genotyping

Approximately 3.0 g of tissue from the leaf primordia of each accession was collected, and the genomic DNA was extracted according to methods described by Aljanabi et al. [55]. The SSR markers were amplified based on the procedures described by Oliveira et al. [56], and the amplified fragments were visualized as described by Creste et al. [57]. A total of 100 SSR primers were used, of which 86 were from expressed sequences (EST-SSR) [58,59] and 14 were of genomic origin [60]. These markers were selected because they met one or more of the following criteria: i) high polymorphic information content (PIC); ii) high discrimination power (DP); and iii) present in previously published sugarcane genetic maps.

Due to the polyploid and complex nature of sugarcane, the amplified SSR fragments, which cannot depict ploidy levels and allele dosages, were evaluated as dominant markers [61,62], i.e., the presence of fragments suggested that an allele for a given locus was present in at least one of the chromosomes that comprised a homologous group, while the absence of fragments suggested that this same allele was not present in any chromosome. Thus, the fragments were classified as binary, i.e., (1) indicated a fragment was present, and (0) indicated a fragment was

absent. When amplification failed, NA (nonamplified) was used to indicate missing data. The polyacrylamide gels were manually evaluated with the support of a light box, and a binary matrix formed by the combination of the detected fragments with the analyzed accessions was constructed.

### Population structure and genetic diversity

Population structure was analyzed by a discriminant analysis of principal components (DAPC) [63] using SSR data in the adegenet package [64], which is available in R software [65], as described by Jombart and Collins [66] and Deperi et al. [67]. Briefly, the `find.clusters` function was used to detect the number of clusters in the BPSG. This function uses K-means clustering, which decomposes the total variance of a variable into between-group and within-group components. The best number of subpopulations has the lowest associated BIC. A cross validation function (`xval.Dapc`) and optimal  $\alpha$ -score function (`optim.a.score`) were used to confirm the correct number of principal components (PCs) to be retained. The optimal number of PCs to retain is associated with the lowest root mean square error and with the highest optimized  $\alpha$ -score. The subpopulations indicated by DAPC were plotted in a scatterplot considering the first and second linear discriminants. Additionally, a genetic dissimilarity matrix was calculated via a simple matching (SM) method using Darwin software [68] based on the SSR information. Then, the resulting matrix was plotted as a phylogram using the neighbor-joining (NJ) algorithm [69]. In addition, bootstrap analysis was performed as described by Efron [70] and Efron and Tibshirani [71] to verify whether the number of fragments evaluated was sufficient to distinguish the accessions. The coefficients of variation are graphically shown as boxplots for each sampling with different numbers of fragments.

### Kinship matrix

The kinship coefficient was calculated between pairs of accessions using the `kinship2` package [72] in R, considering the accessions of all generations and assigning the value 0 when the parents were unknown. Based on the estimated kinship coefficients, a kinship matrix ( $K$ ) was generated.

### Linkage disequilibrium analysis

Marker data were used to assess the level of LD in the BPSG as described by Raboin et al. [35]. Briefly, Fisher's exact probability was used to test for associations between SSR fragments that were common to both the association mapping population and the SP80-180 and SP80-4966 integrated genetic map [56]. For each pair of markers, a contingency table (presence versus absence) was established, and the Fisher probability was computed using the `exact2x2` package in R software [73]. To control for error due to multiple testing, we used the false discovery rate (FDR) procedure [74] with an initial threshold of 5%. A Bonferroni-corrected threshold was also verified. The Fisher ( $-\log P$ ) logarithmic probabilities of the associations between only linked fragments were plotted with the respective genetic distances [75] in centimorgans (cM).

### GWAS analysis

GWAS analysis was conducted using both the Genomic Association and Prediction Integrated Tool (GAPIT, [76]) and FarmCPU [44] methods in R software. To carry out GWAS analyses using the SSR data obtained in the BPSG, the fragments were reclassified, with (2) indicating the presence of a fragment and (0) indicating the absence of a fragment. The retained PC obtained in DAPC analysis was used as a covariate in the FarmCPU procedure, while the

kinship matrix and retained PC were used in the GAPIT analysis. To control for type I errors due to multiple testing, the adjusted  $p$ -value less than 1% following an FDR controlling procedure [77] and Bonferroni-corrected threshold with 1% were used to declare significant MTAs by GAPIT and FarmCPU, respectively. To determine which of the tested methods best fit the data, we plotted the quantile-quantile (QQ) plot, i.e., the QQ negative log<sub>10</sub>-transformed observed  $p$ -values obtained for each MTA, against their expected distribution under the null hypothesis of no genetic association. For significant MTAs detected by FarmCPU, the phenotypic variance explained for each SSR fragment was estimated one at a time using a linear model with the `lm` function in R software.

## Sequence annotation

Functional annotation of the loci associated with traits was performed using the available sequences that gave rise to the SSR marker. These sequences were annotated using i) the non-redundant NCBI database with  $e$ -values  $\leq 1 \times 10^{-3}$  through BLASTX and ii) the Phytozome website [78], which was used to align the data against the *Viridiplantae* protein databases.

## Results

### Phenotypic data

The VCOV models selected for the  $G_M$  matrix were based on AIC and BIC criteria. AR1(het) had the lowest AIC and BIC values, which indicated that it was the best model for all evaluated traits (BRIX, SH, SN, SW and TCH) (S2 Table). This result supports heterogeneous genetic variances between harvests and correlations between successive harvests and provides a systematic explanation of the existing temporal dependence. The ranges, adjusted means and estimates of the components of variance, coefficients of variation, and broad-sense heritability on an individual-plant basis for the five traits evaluated for the BPSG over the two harvest years (plant cane and first ratoon) are summarized in Table 1. The TCH trait had the highest variation, i.e., the accession RB925268 (295.60 t ha<sup>-1</sup>) was 7.6 times greater than the accession POJ2878 (38.90 t ha<sup>-1</sup>). The SN trait also showed high variation, i.e., the accession IN84-58 (290.64 stalks) was 7.03 times greater than the accession POJ2878 (41.34 stalks). On the other hand, the BRIX trait had a relatively low variation, i.e., the accession TUC71-7 (22.55° Brix) was 1.48 times greater than the accession IN84-58 (15.14° Brix).

Estimates for  $\hat{H}^2$  ranged from 0.48 (TCH) to 0.67 (SN) and from 0.49 (TCH) to 0.65 (SN) in the first and second harvests, respectively. For genetic ( $\hat{\sigma}_G^2$ ) and phenotypic ( $\hat{\sigma}_P^2$ ) variances, higher and lower values were observed for the TCH and SH traits, respectively. The lowest coefficients of genetic ( $CV_G$ ) and phenotypic ( $CV_P$ ) variations were for the BRIX trait, while the higher values for  $CV_G$  and  $CV_P$  were for SN, SW and TCH.

Pairwise genotypic correlations among the five evaluated traits, considering both harvests (plant cane and first ratoon), are shown in Fig 1. In total, eight significant genotypic correlations ( $P < 0.05$ ) were observed between the evaluated traits in the BPSG. According to the degree of correlation between traits, correlations were grouped into low ( $\leq 0.35$ ), moderate (0.36–0.70) and strong ( $\geq 0.71$ ) categories [12]. Thus, four interactions were classified as low (BRIX–SH, BRIX–SW, BRIX–TCH and SH–SN), four interactions were classified as moderate (BRIX–SN, SN–SW, SN–TCH and SH–TCH), and two interactions were classified as strong (SH–SW and SW–TCH). The correlation of BRIX–SN was negative.

**Table 1. Ranges, adjusted means, estimates of components of genetic variance ( $\hat{\sigma}_G^2$ ) and phenotypic variance ( $\hat{\sigma}_P^2$ ), coefficients of genetic variation ( $CV_G$ ) and phenotypic variation ( $CV_P$ ), and broad-sense heritability on an individual-plant basis ( $\hat{H}^2$ ) for BRIX, SH, SN, SW and TCH for the BPSG over two harvest years (plant cane (1) and first ratoon (2)).**

Trait	Range	Adjusted means	$\hat{\sigma}_{G(1)}^2$	$\hat{\sigma}_{G(2)}^2$	$\hat{\sigma}_{P(1)}^2$	$\hat{\sigma}_{P(2)}^2$	$CV_{G(1)}$	$CV_{G(2)}$	$CV_{P(1)}$	$CV_{P(2)}$	$\hat{H}_{(1)}^2$	$\hat{H}_{(2)}^2$
BRIX (°Brix)	15.14–22.55	18.65	1.45	0.96	1.09	0.72	0.06	0.05	0.05	0.04	0.57	0.57
SH (m)	1.51–3.08	2.64	0.08	0.08	0.06	0.06	0.11	0.11	0.09	0.09	0.57	0.57
SN	41.34–290.64	81.77	482.60	1197.00	232.10	645.40	0.27	0.42	0.18	0.31	0.67	0.65
SW (kg)	47.80–251.80	152.30	2182.00	1602.00	1576.00	1267.00	0.31	0.26	0.26	0.23	0.58	0.56
TCH (t ha <sup>-1</sup> )	38.90–295.60	181.40	2740.00	2127.00	2891.00	2192.00	0.29	0.25	0.30	0.26	0.48	0.49

<https://doi.org/10.1371/journal.pone.0219843.t001>

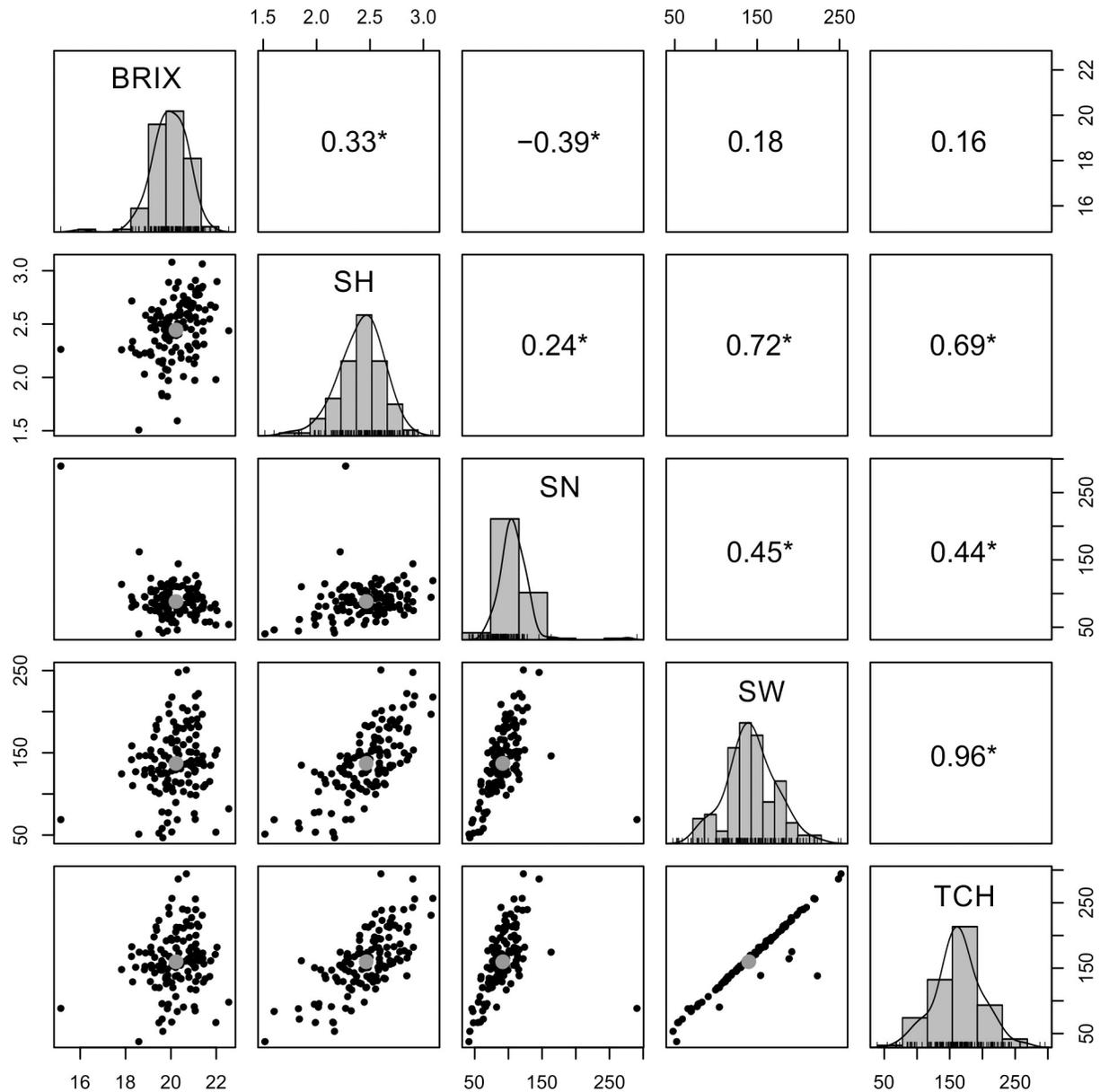
### Polymorphisms of SSR markers

The use of 100 SSR markers generated 1483 fragments, 1476 of which were polymorphic (99.5%), in the 134 accessions of the BPSG. Considering all polymorphic fragments, 484 (32.8%) were produced by SSR dinucleotides, 689 (46.7%) were produced by SSR trinucleotides, and 303 (20.5%) were produced by SSR tetranucleotides. The number of fragments ranged from four (ESTC52 and ESTC55) to 36 (ESTA31), with an average of 14.83 fragments per SSR. Species-specific fragments were observed for the ancestral accessions Badila (*S. officinarum*) at ESTB45 and SMC319; Ganda Cheni (*S. barberi*) at ESTB45, ESTB118, ESTA51, and ESTC17; and especially IN84-58 (*S. spontaneum*) at CIR23, ESTA26, ESTA61, CIR55, ESTB69, ESTA33, ESTB94, ESTA63, CIR18, ESTB63, CIR36, ESTB45, ESTA16, ESTC55, ESTA48, SMC222 and CIR25.

### Population structure and genetic diversity

Four subpopulations were detected according to the lowest BIC value derived by the find.clusters function (S1 Fig). DAPC analysis was performed using the detected number of subpopulations (Fig 2). Seven first PCs (25.5% of variance conserved) from principal component analysis (PCA) (S2 and S3 Figs) and three discriminant eigenvalues were retained. All accessions were classified in each subpopulation with a membership coefficient equal to 1, suggesting that there were no admixtures and that the BPSG was structured (S4 Fig). A total of 42 fragments with the largest contribution to subpopulation identification were detected, with 24 fragments assigned to linear discriminant 1 and 18 fragments assigned to linear discriminant 2 (S3 Table and S5 Fig).

The phylogram using the SM genetic distance among accessions also suggested the presence of four subpopulations. A total of 99.25% of the group assignments made by the DAPC analysis were also made by the phylogram (Fig 3). Only accession SP70-1284 was assigned to different groups by the NJ phylogram and DAPC methods. The genetic dissimilarity ranged from 0.06 (between accessions IAC68-12 and IAC64-257, in subpopulation 3) to 0.45 (between accessions SP70-1005 and RB855589, in subpopulations 2 and 1, respectively), with an average value of 0.31 (S6 Fig). Overall, the clusters inside subpopulations were in accordance with the pedigree information. This result was verified by full-sib accessions within the subpopulations, as was the case for the accessions RB845197, RB845210 and RB845257 in subpopulation 3, which originated from the crossing between cultivars RB72454 and SP70-1143, and for the cultivars SP80-1816, SP80-1842 and SP80-3280 in subpopulation 2, which originated from the crossing between the cultivars SP71-1088 and H57-5028. In addition, the ancestral accessions Maneria (*Saccharum sinense*) and Ganda Cheni (*S. barberi*) were placed in subpopulation 2, the ancestral accessions Badila (*S. officinarum*) and IN84-58 (*S. spontaneum*) were positioned



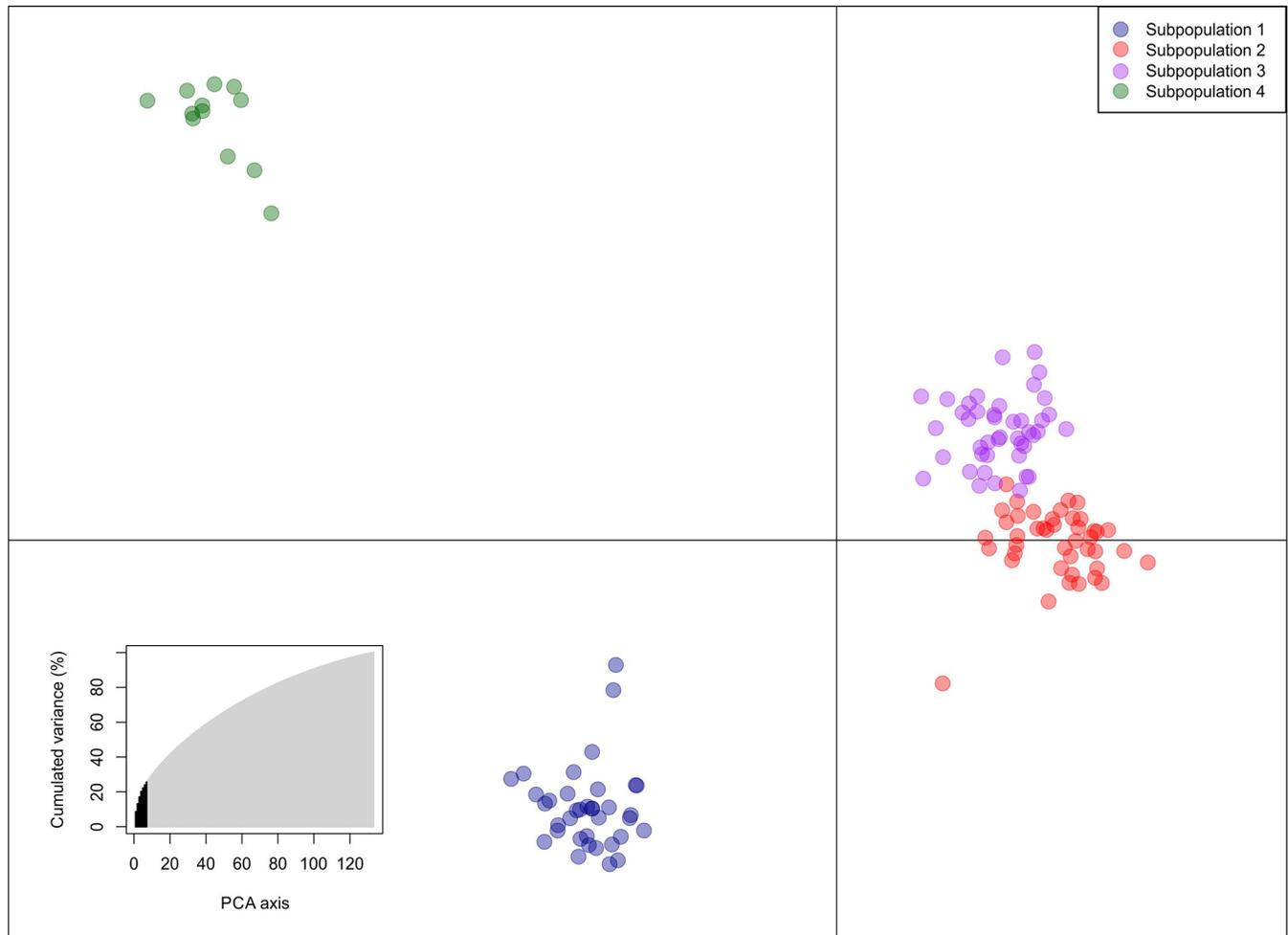
**Fig 1. Genotypic correlation between yield traits evaluated in the BPSG.** For each trait, the histograms of the adjusted means (diagonal), scatterplots (below diagonal), and values of the genotypic correlation (above diagonal) between pairs of traits are shown. \*Significant at the 5% global level ( $P < 0.05$ ).

<https://doi.org/10.1371/journal.pone.0219843.g001>

in subpopulation 1, and the ancestral accession White Transparent (*S. officinarum*) was positioned in subpopulation 4.

### Linkage disequilibrium analysis

Fisher's ( $-LogP$ ) logarithm probabilities were plotted against the distances, in centiMorgan (cM), between linked marker fragments in the same cosegregation group of the SP80-180 × SP80-4966 integrated genetic map (S7 Fig). This strategy corresponded to 60 of the 5151 associations between the 102 common SSR fragments being present in both the BPSG and the integrated genetic map. Although few significant associations reported LD (5 of the 60



**Fig 2. DAPC for the BPSG.** The axes represent the first two linear discriminants (LD). The dots represent accessions grouped in subpopulations, each with a different color. The cumulative variance values, in percentages, of the PCs are shown in the lower left corner of the figure; the eigenvalues of the seven first PCs retained by PCA are in black.

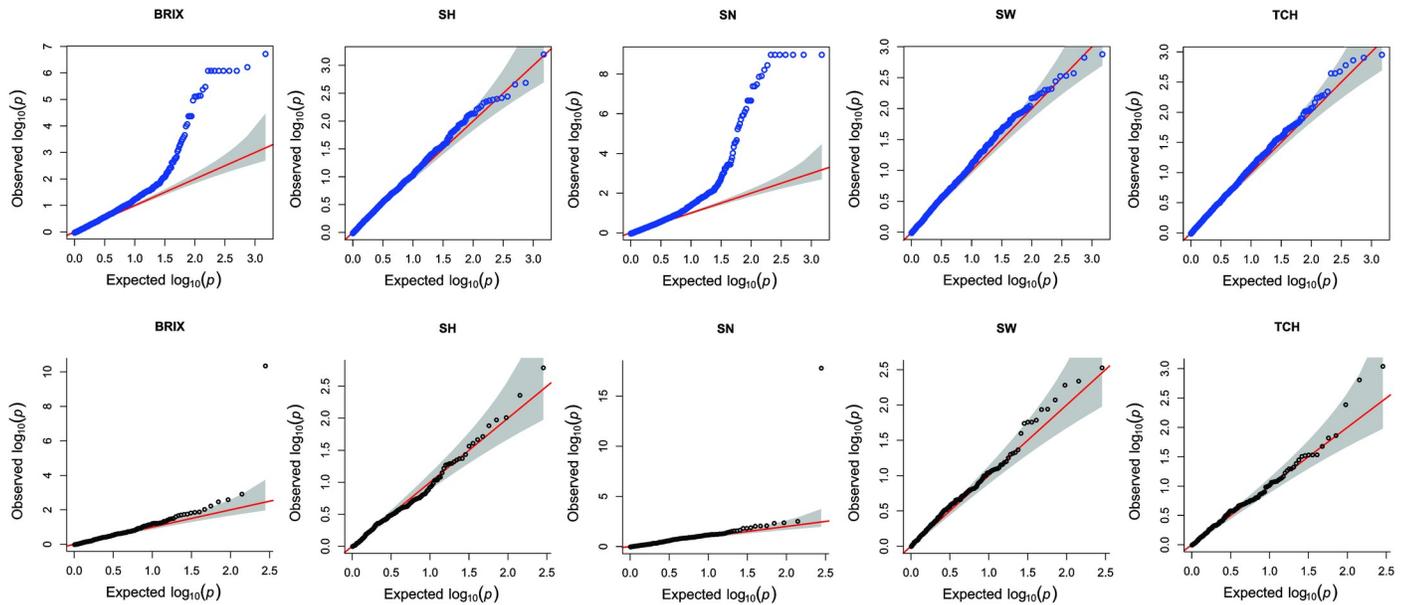
<https://doi.org/10.1371/journal.pone.0219843.g002>

associations), the population showed good evidence of LD decay in relation to genetic distance. The strongest LD appeared in the first 15 cM, mainly in the first 5 cM, and clear decay occurred over distances. In addition, LD was noted between fragments at 65 cM in the same cosegregation group, indicating preferential associations in larger extensions.

### GWAS analysis

The QQ plots obtained with FarmCPU and GAPIT software for phenotypic traits are presented individually in Fig 4. The results show that FarmCPU compared to GAPIT better fit the data by reducing false positives, mainly for the BRIX and SN traits. Therefore, we considered the MTAs identified by FarmCPU to be more reliable than those identified by GAPIT and thus present the results of the former. For the BRIX, SH, SN, SW and TCH traits, 6, 3, 7, 4 and 3 MTAs were detected, respectively, with a Bonferroni-corrected threshold of 1% (Table 2). The SSR fragment ESTB61\_15 was negatively and positively associated with BRIX and SN, respectively. ESTB61\_15 is a species-specific fragment for *S. spontaneum* (IN84-58). The three SSR fragments associated with TCH were also associated with SW, and two of these fragments





**Fig 4.** QQ plots using GAPIT (graphs with blue dots) and FarmCPU (graphs with black dots) software. The dotted lines show the 95% confidence intervals for the QQ plots under the null hypothesis of no association between the SSR fragment and the trait.

<https://doi.org/10.1371/journal.pone.0219843.g004>

**Table 2.** BRIX, SH, SN, SW and TCH MTAs, *p*-values, effect estimates and amounts of phenotypic variance explained (adjusted R-squared) when using the MLM implemented in FarmCPU.

Trait	Code	SSR fragment	<i>p</i> -value	Effect	R <sup>2</sup>
BRIX	m93	ESTA61_07	0.009004	-0.22	0.03
BRIX	m101	ESTA61_15	4.29E-11	-2.79	0.20
BRIX	m131	CIR55_06	0.001163	0.24	0.02
BRIX	m139	CIR55_14	0.002446	-0.32	0.14
BRIX	m515	CIR51_04	0.005713	0.26	0.01
BRIX	m797	ESTB133_10	0.003254	0.27	0.01
SH	m745	SMC319_16	0.004327	-0.42	0.07
SH	m752	SMC248_08	0.001605	0.09	0.14
SH	m839	ESTC19_12	0.009647	-0.08	0.07
SN	m101	ESTA61_15	1.50E-06	104.29	0.43
SN	m138	CIR55_13	0.003936	21.06	0.02
SN	m522	CIR51_11	0.008232	5.37	0.04
SN	m650	ESTB111_05	0.008690	-11.21	0.02
SN	m664	ESTB111_19	0.002828	7.22	0.01
SN	m738	SMC319_09	0.007744	-5.59	0.03
SN	m921	ESTB130_16	0.004402	7.57	0.03
SW	m522	CIR51_11	0.004541	12.01	0.02
SW	m738	SMC319_09	0.002941	-12.98	0.03
SW	m937	ESTB130_32	0.008385	10.92	0.05
SW	m1070	SMC222_01	0.005182	-13.88	0.02
TCH	m522	CIR51_11	0.000901	15.94	0.03
TCH	m738	SMC319_09	0.001533	-15.78	0.03
TCH	m1070	SMC222_01	0.004051	-16.33	0.02

<https://doi.org/10.1371/journal.pone.0219843.t002>

**Table 3. Functional descriptions of the sequences that Gave Rise to SSR markers associated with the BRIX, SH, SN, SW and TCH traits as determined using BLASTX and phytozome (NA: No available sequence).**

SSR marker	Traits	Description	e-value
ESTA61	BRIX, SN	Cortical cell-delineating protein [ <i>Sorghum bicolor</i> ]	6.3E-166
ESTB111	SN	Exonuclease DPD1, chloroplastic/mitochondrial [ <i>Sorghum bicolor</i> ]	3.1E-113
ESTB130	SN, SW	Auxin response factor 5 [ <i>Sorghum bicolor</i> ]	7.0E-87
ESTB133	BRIX	Vacuolar fusion protein MON1 [ <i>Sorghum bicolor</i> ]	4.9E-57
ESTC19	SH	DVL family protein [ <i>Zea mays</i> ]	6.3E-77
CIR51	BRIX, SN, SW, TCH	Near to Cytochrome P450 transcript region [ <i>Sorghum bicolor</i> ]	9.1E-57
CIR55	BRIX, SN	Uncharacterized protein [ <i>Sorghum bicolor</i> ]	9.8E-60
SMC222	SW, TCH	NA	-
SMC248	SH	NA	-
SMC319	SH, SN, SW, TCH	NA	-

<https://doi.org/10.1371/journal.pone.0219843.t003>

found for seven out of the ten SSR markers significantly associated with homologies for *Sorghum bicolor* (for the BRIX, SN and SW traits) and *Zea mays* (for the SH trait). A functional description of the sequences showed possible candidate genes for all traits except for TCH. Despite this result, the CIR51 marker, which was found near (approximately 5.3 kb) the *cytochrome P450* transcript region in *S. bicolor*, had fragments significantly associated with TCH in addition to BRIX, SN and SW. Overall, the homologies found for significant SSR markers associated with BRIX (ESTA61, ESTB133) suggest a role in the accumulation and trafficking of lipids and sucrose, while the homologies for significant SSR markers associated with SH (ESTC19), SN (ESTB111, ESTB130) and SW (ESTB130) were related to plant growth and development.

## Discussion

The complexity of the sugarcane genome and the quantitative nature of sugar- and yield-related traits are challenging for geneticists and breeders searching for higher genetic gains for this crop. Moreover, assessing genetic variables free of environmental effects and estimating their real genotypic value are extremely important for breeding purposes. Here, the genetic information obtained with SSR markers was able to efficiently distinguish ancestral and improved accessions of the BSPG due the high polymorphism and presence of unique alleles in some accessions, such as IN84-58 (*S. spontaneum*), Badila (*S. officinarum*) and Ganda Cheni (*S. barberi*). The identification of new alleles controlling sugar and yield metabolism in alternative *Saccharum* species and the introduction of these alleles into core germplasms would be one way to overcome obstacles in sugarcane breeding, increasing the productivity of commercial cultivars [79]. Following this strategy, association mapping is a powerful tool to identify genes and favorable alleles that could be used for the introgression process. In the present study, using the GWAS approach, we were able to detect MTAs for all five evaluated traits (BRIX, SH, SN, SW and TCH), mainly due to the presence of LD in the BSPG and by the analysis strategies employed.

The model selection approach used in this study for phenotypic data analysis can capture the heterogeneity of variance and more complex covariance structures (AR1(het)) at the genetic level, thereby improving the predictive accuracy directly related to heritability and genetic gain [12,53,80,81]. In the AR1(het) model selected for all traits (S2 Table), the

correlations between harvest decay with time and each harvest have their own genetic variance [53]. Indeed, sugarcane production decreases with harvests; therefore, the differential expression of genes across harvests can be suggested. On the other hand, the use of more locations and harvest years would probably permit the adjustment of other variance and covariance structures [12].

The phenotypic range for each trait reflected the high genetic variability of the BSPG, and the broad-sense heritability values showed that much of the observed phenotypic variation can be attributed to differences at the genotypic level (Table 1). Therefore, the significant genotypic correlations among traits could indicate biological processes that are of considerable evolutionary interest and result from genetic or physiological features [82,83]. The SH, SN and SW traits are involved in plant development and are therefore important parameters in breeding programs that increase genetic gains in terms of cane yield. The MTAs discovered for any of these three traits might potentiate plant development, mainly because the SW, SH and SN traits were significantly associated with the five evaluated traits, and SW was part of the two most strongly detected correlations (SH–SW and SW–TCH) (Fig 1). Similar genotypic correlation results among these traits have been reported in previous studies [39,84].

In addition to genotypic correlations, genetic variability is essential to breeders for the generation of improved cultivars. In the present study, population structure and genetic diversity were assessed in the BSPG through DAPC analysis and by a genetic dissimilarity matrix calculated with the SM distance and visualized as an NJ phylogram, both based on SSR markers. DAPC analysis divided the BSPG into four subpopulations (Fig 2), and this result was confirmed by the NJ phylogram of the SM distance of the whole population (Fig 3). To obtain the population structure, some studies have reported similar or better results for DAPC analysis than for the Bayesian model-based method [63,85–87] implemented in STRUCTURE software [88–90]. In addition, for complex genomes, several assumptions are not fulfilled with the use of STRUCTURE; therefore, the applicability of this algorithm may be limited in sugarcane [34,37,38,91]. The NJ phylogram showed that the subpopulations contained some clusters formed by family relatedness. These results suggested that the BSPG could be affected by population structure and relatedness, which is in agreement with the history of sugarcane breeding [10,18,92].

LD is affected by genetic and nongenetic factors, such as recombination, genetic drift, population stratification, genetic relatedness, mutation, selection and linkage [93,94]. Therefore, the population structure and family relatedness of the BSPG could be responsible for the detected LD, which was stronger in the first 15 cM and present in a large extension, i.e., at 65 cM, similar to the results of Raboin et al. [35] and Wei et al. [38]. Recently, Yang et al. [36] showed a large extent of LD, with lengths of 962.4 Kbp, 2739.2 Kbp and 3573.6 Kbp for *S. spontaneum*, *S. officinarum* and modern hybrids, respectively. The existence of a large LD extent and, consequently, the presence of large gene clusters indicate that a high density of markers is not required to detect MTAs by the GWAS approach in sugarcane. Thus, single-dose markers could be useful for this purpose as an initial step. On the other hand, the LD caused by population structure and familial relatedness can promote false positive detection in GWAS analysis [95–97], and to avoid these spurious associations, the models consider covariates (population structure matrix and/or kinship matrix) to adjust the association tests on markers. In addition, confounding between these covariates and testing markers also produces false negatives [44,96].

The QQ plots obtained with GAPIT software showed that the association tests were inflated and resulted in false positives when compared with the QQ plots generated by FarmCPU software, mainly for the BRX and SN traits (Fig 4). The compressed mixed linear model (CMLM), implemented in GAPIT, is a single-locus model that tests one marker at a time and

maintains the kinship matrix constant for all markers [76]. On the other hand, FarmCPU, a multilocus model, implements a fixed model that contains the testing markers and covariates (multiple associated markers and PCs) and a random model that contains the kinship matrix. This kinship matrix is adjusted based on the testing markers and covariates of the fixed model [44]. Therefore, the differences in the analysis procedures could explain the occurrence of false positives by GAPIT, which fails to match the true genetic model of complex traits that are controlled by numerous loci simultaneously [48], such as those evaluated in the present study. In GAPIT, other associated loci nearby or elsewhere in the genome will sometimes disrupt with the tested marker and result in spurious associations, especially when the effects of the other loci are large [98]. In addition, in GAPIT, covariate information could overlap (kinship matrix and PCs), as previous studies have shown that the PCs from PCA also include part of the family relatedness [99,100]; therefore, the seven PCs retained by the DAPC analysis, which explained 25.5% of the variance, provided some information about relatedness and population structure for GWAS analysis. Finally, the more reliable MTAs detected with the FarmCPU approach could be attributed to the use of only retained PCs of DAPC as a covariate and the MLM, which was able to remove the confounding between the tested markers and covariates [44].

The GWAS analysis with FarmCPU software revealed 23 MTAs associated with five traits when the Bonferroni-corrected threshold was set to 1% (Table 2). All but four MTAs showed a low percentage of explained phenotypic variation, with values ranging from 1% to 7%. These low values may be due to the high ploidy level of sugarcane and the quantitative inheritance of the evaluated traits [39]. In addition, the SSR fragments are treated as dominant in polyploid species, such as sugarcane, and thus do not capture the allelic dosage information of homologous chromosomes [101]. Despite that Fickett et al. [42] obtained 6299 SNPs and 235 InDels through a high-throughput genotyping system, only 27 markers were significantly associated with six traits (stalk number, stalk height, stalk diameter, °Brix, pol and fiber) and explained no more than 14.3% of the phenotypic variation. Therefore, genetic studies on polyploidy species, like sugarcane, are obviously delayed when compared to those on crops with minor genetic complexity. New methods of analysis are still in development to increase the understanding of complex genomes and enable mapping and association studies with further levels of allelic information [24,26,102]. Despite this, four MTAs with the highest percentages of explained phenotypic variation (43% for SN with ESTA61\_15, 20% for BRIX with ESTA61\_15, 14% for BRIX with CIR55\_14 and 14% for SH with SMC248\_08) indicate that the presence of at least one copy of the allele could also be important for driving strategies in breeding programs. The SSR fragment ESTA61\_15, a species-specific fragment present in *S. spontaneum* accession IN84-58, was positively and negatively associated with the SN and BRIX traits, respectively. ESTA61\_15 may be a unique allele that causes important phenotypic variation. Previous studies detected MTAs for the SW [37,39], SN [18,39,40,42], SH [18,39,40,42] and BRIX [18,40–42] traits, and the percentages of phenotypic variation found in the present study were similar for SW, SH and BRIX and higher for SN. Therefore, these MTAs may be validated as an initial approach to support breeding programs with introgression or selection processes [37,41,42].

To understand the plant metabolism functions of the SSR marker regions associated with traits and search for candidate genes, we annotated the available sequences from which the associated markers originated. Thus, the sequence that produced the ESTA61 marker showed similarity with *cortical cell-delineating protein*, which is a member of the alpha-amylase inhibitors, lipid transfer and seed storage (AAI-LTSS) protein family according to SMART annotation in Phytozome [78]. This result suggests differential lipid transport and sucrose accumulation performances between *S. spontaneum* and other BPSG accessions [103,104].

The ESTB133 marker, also associated with BRIX, showed similarity with the *vacuolar fusion protein MON1*, which is a member of the MON1/SAND protein family. In *Arabidopsis*, the MON1 and CCZ1 proteins form a complex that is critical for vacuolar trafficking, vacuole biogenesis, and plant growth. The *mon1* mutants show pleiotropic growth defects, fragmented vacuoles, and altered vacuolar trafficking [105]. Therefore, the accumulation and vacuolar trafficking of the sucrose in sugarcane could be affected by alteration of this marker region.

For the ESTB111 marker, which was associated with SN, similarity with *exonuclease DPD1, chloroplastic/mitochondrial* could indicate that the nucleotides, i.e., purines and pyrimidines, released during the leaf senescence process provide nitrogen, sugar and phosphate to maintain or increase the plant tillering ability [106]. Likewise, the ESTB130 marker, which was associated with SN and SW, showed similarity to *auxin response factor 5* (ARF5), which acts as a transcriptional activator of auxin-responsive promoter elements. This homology suggests that a modification in the ARF5 protein could affect plant growth and development and consequently affect the weight and stalk production of sugarcane [107–109]. For the SH trait, the significantly associated marker ESTC19 showed similarity to DVL family proteins. In *Arabidopsis*, the overexpression of DVL1 was associated with plants with a shortened stature, smaller and rounder rosette leaves, clustered inflorescences, shortened pedicles, and siliques with pronged tips resembling horns [110]. Thus, this result suggests that the ESTC19 marker also plays a role in sugarcane plant development.

The GWAS analysis with FarmCPU software, which used population structure information derived from DAPC analysis as a covariate, was able to detect MTAs with efficient control of spurious associations in sugarcane. In addition, the verification of possible candidate genes for MTAs showed the importance of providing insights into gene networks that are related to the expression of target traits. This approach has great potential for assisting breeding programs in increasing the genetic gain rate of target traits. However, the development of statistical approaches to enable mapping association with markers in multiple doses is important to enhance the probability of finding higher numbers of significant associations and, consequently, increase the use of molecular markers in breeding programs of outcrossing heterozygous species, such as sugarcane.

## Supporting information

**S1 Table. Names, parents and origins of 134 accessions of the BPSG.**  
(PDF)

**S2 Table. Selected models for the  $G_M$  matrix and number of estimated parameters ( $n_{par}$ ) considering each trait separately.** The Akaike (AIC) and Bayesian (BIC) information criteria were used to compare the structures of the variance–covariance matrix. The models for the  $G_M$  matrix were selected according to the lowest value of the BIC criterion for BRIX as °Brix, stalk height (SH) in m, stalk number (SN) by direct counting, stalk weight (SW) in kg and cane yield (TCH) in t ha<sup>-1</sup> for BPSG over two harvest years (plant cane and first ratoon). Bold numbers represent the smallest AIC and BIC values.  
(PDF)

**S3 Table. SSR fragments with the largest values of contribution to subpopulation identification detected through the loadingplot function.** A threshold of 0.005 was used to declare the major contributions. LD: linear discriminant.  
(PDF)

**S1 Fig. Number of subpopulations (clusters) vs. BIC values.** The x-axis represents the different number of subpopulations that could be presented in the Brazilian Panel of Sugarcane

Genotypes (BPSG). The y-axis represents the BIC value associated with each number of subpopulations.

(PDF)

**S2 Fig. Cross validation of DAPC.** The x-axis represents the number of PCs retained in each DAPC. The y-axis represents the proportion of successful outcome prediction. Each dot represents the individual replicate of the analysis.

(PDF)

**S3 Fig. Numbers of retained principal components (PCs) vs.  $\alpha$ -score values.** The  $\alpha$ -score on the y-axis depicts the difference between the proportion of successful reassignment of the analysis (observed discrimination) and the values obtained using random groups (random discrimination). The x-axis represents the number of retained PCs for each random group. The spline interpolation approximates the optimal number of PCs to be retained.

(PDF)

**S4 Fig. Distribution of the accessions of BPSG (y-axis) into 4 subpopulations (x-axis) obtained through the DAPC.** The red regions inside the columns indicate the set of accessions grouped in the corresponding subpopulation according to the membership probabilities.

(PDF)

**S5 Fig. Variable contributions (SSR Fragments) for linear discriminants 1 (LD1) and 2 (LD2).** A threshold of 0.005 was used to declare major contributions.

(PDF)

**S6 Fig. Heatmap with genetic dissimilarities among accessions of the BPSG.** The subpopulations obtained through discriminant analysis of principal components (DAPC) are also shown.

(PDF)

**S7 Fig. Plot of linkage disequilibrium ( $-\log P$ ) and genetic distance (cM) in the SP80-180 and SP80-4966 genetic map.** The thresholds corresponding to the Bonferroni and false discovery rate (FDR) corrections are indicated on a logarithmic scale, showing significant and nonsignificant associations above and below their values, respectively. Genetic distances were obtained through the Kosambi mapping function.

(PDF)

## Acknowledgments

We gratefully acknowledge Sandro Augusto Ferrarez for providing support in the field experiment.

## Author Contributions

**Conceptualization:** Fernanda Zatti Barreto, João Ricardo Bachega Feijó Rosa, Thiago Willian Almeida Balsalobre, Hermann Paulo Hoffmann, Anete Pereira de Souza, Antonio Augusto Franco Garcia, Monalisa Sampaio Carneiro.

**Data curation:** Fernanda Zatti Barreto, João Ricardo Bachega Feijó Rosa, Thiago Willian Almeida Balsalobre, Monalisa Sampaio Carneiro.

**Formal analysis:** Fernanda Zatti Barreto, João Ricardo Bachega Feijó Rosa, Thiago Willian Almeida Balsalobre, Maria Marta Pastina, Renato Rodrigues Silva, Antonio Augusto Franco Garcia.

**Funding acquisition:** Monalisa Sampaio Carneiro.

**Investigation:** Fernanda Zatti Barreto, João Ricardo Bachega Feijó Rosa, Thiago Willian Almeida Balsalobre, Anete Pereira de Souza, Antonio Augusto Franco Garcia.

**Methodology:** João Ricardo Bachega Feijó Rosa, Thiago Willian Almeida Balsalobre, Maria Marta Pastina, Renato Rodrigues Silva, Anete Pereira de Souza, Antonio Augusto Franco Garcia, Monalisa Sampaio Carneiro.

**Project administration:** Monalisa Sampaio Carneiro.

**Supervision:** Monalisa Sampaio Carneiro.

**Writing – original draft:** Fernanda Zatti Barreto, João Ricardo Bachega Feijó Rosa, Thiago Willian Almeida Balsalobre, Maria Marta Pastina, Renato Rodrigues Silva, Hermann Paulo Hoffmann, Anete Pereira de Souza, Antonio Augusto Franco Garcia, Monalisa Sampaio Carneiro.

**Writing – review & editing:** Fernanda Zatti Barreto, João Ricardo Bachega Feijó Rosa, Thiago Willian Almeida Balsalobre, Anete Pereira de Souza, Antonio Augusto Franco Garcia, Monalisa Sampaio Carneiro.

## References

1. Saini JK, Saini R, Tewari L. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech*. 2015; 5: 337–353. <https://doi.org/10.1007/s13205-014-0246-5> PMID: 28324547
2. Bordonal RDO, Carvalho JLN, Lal R, de Figueiredo EB, de Oliveira BG, la Scala N. Sustainability of sugarcane production in Brazil. A review. *Agron Sustain Dev*. 2018; 38: 13.
3. D'Hont A, Ison D, Alix K, Roux C, Glaszmann JC. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome*. 1998; 41: 221–225.
4. Hotta CT, Lembke CG, Domingues DS, Ochoa EA, Cruz GMQ, Melotto-Passarim DM, et al. The biotechnology roadmap for sugarcane improvement. *Trop Plant Biol*. 2010; 3: 75–87.
5. de Setta N, Monteiro-Vitorello CB, Metcalfe CJ, Cruz GM, del Bem LE, Vicentini R, et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics*. 2014; 15: 540. <https://doi.org/10.1186/1471-2164-15-540> PMID: 24984568
6. Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, et al. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat Commun*. 2018; 9: 2638. <https://doi.org/10.1038/s41467-018-05051-5> PMID: 29980662
7. Thirugnanasambandam PP, Hoang NV, Henry RJ. The challenge of analyzing the sugarcane genome. *Front Plant Sci*. 2018; 9: 616. <https://doi.org/10.3389/fpls.2018.00616> PMID: 29868072
8. Vieira MLC, Almeida CB, Oliveira CA, Tacuatiá LO, Munhoz CF, Cauz-Santos LA, et al. Revisiting meiosis in sugarcane: chromosomal irregularities and the prevalence of bivalent configurations. *Front Genet*. 2018; 9: 213. <https://doi.org/10.3389/fgene.2018.00213> PMID: 29963076
9. Cheavegatti-Gianotto A, de Abreu HMC, Arruda P, Filho JCB, Burnquist WL, Creste S, et al. Sugarcane (*Saccharum X officinarum*): a reference study for the regulation of genetically modified cultivars in Brazil. *Trop Plant Biol*. 2011; 4: 62–89. <https://doi.org/10.1007/s12042-011-9068-3> PMID: 21614128
10. de Moraes LK, Aguiar MS, Silva PA, Câmara TMM, Cursi DE, Júnior ARF, et al. Breeding of sugarcane. In: Cruz VMV, Dierig DA, editors. *Industrial crops: breeding for bioenergy and bioproducts*. New York, NY: Springer; 2015. pp. 29–42.
11. Dal-Bianco M, Carneiro MS, Hotta CT, Chapola RG, Hoffmann HP, Garcia AA, et al. Sugarcane improvement: how far can we go? *Curr Opin Biotechnol*. 2012; 23: 265–270. <https://doi.org/10.1016/j.copbio.2011.09.002> PMID: 21983270
12. Balsalobre TW, Mancini MC, Pereira GDS, Anoni CO, Barreto FZ, Hoffmann HP, et al. Mixed modeling of yield components and brown rust resistance in sugarcane families. *Agron J*. 2016; 108: 1824–1837.
13. Guilly S, Dumont T, Thong-Chane A, Barau L, Hoarau J-Y. Analysis of multi-environment trials (MET) in the sugarcane breeding program of Réunion Island. *Euphytica*. 2017; 213: 213.

14. Jackson P, McRae T, Hogarth M. Selection of sugarcane families across variable environments I. Sources of variation and an optimal selection index. *Field Crops Res.* 1995; 43: 109–118.
15. Stringer JK, Cox MC, Atkin FC, Wei X, Hogarth DM. Family selection improves the efficiency and effectiveness of selecting original seedlings and parents. *Sugar Tech.* 2011; 13: 36–41.
16. Yang K, Jackson P, Wei X, Chen X, Liu J, Zhao J, et al. Selecting for multiple traits in complex production systems: a case study of sugarcane in China. *Agric Syst.* 2017; 152: 110–120.
17. Pardey PG, Beddow JM, Hurley TM, Beatty TKM, Eidman VR. A bounds analysis of world food futures: global agriculture through to 2050. *Aust J Agric Resour Econ.* 2014; 58: 571–589.
18. Gouy M, Rousselle Y, Thong-Chane A, Anglade A, Royaert S, Nibouche S, et al. Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. *Euphytica.* 2015; 202: 269–284.
19. Xu Y, Li P, Zou C, Lu Y, Xie C, Zhang X, et al. Enhancing genetic gain in the era of molecular breeding. *J Exp Bot.* 2017; 68: 2641–2666. <https://doi.org/10.1093/jxb/erx135> PMID: 28830098
20. Sukumaran S, Reynolds M, Lopes M, Crossa J. Genome-wide association study for adaptation to agronomic plant density: a component of high yield potential in spring wheat. *Crop Sci.* 2015; 55: 1–11.
21. Cao K, Zhou Z, Wang Q, Guo J, Zhao P, Zhu G, et al. Genome-wide association study of 12 agronomic traits in peach. *Nat Commun.* 2016; 7: 13246. <https://doi.org/10.1038/ncomms13246> PMID: 27824331
22. Li H, Rasheed A, Hickey LT, He Z. Fast-forwarding genetic gain. *Trends Plant Sci.* 2018; 23: 184–186. <https://doi.org/10.1016/j.tplants.2018.01.007> PMID: 29426713
23. Garcia A, Mollinari M, Marconi T, Serang OR, Silva R, Vieira M-L, et al. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci Rep.* 2013; 3: 3399. <https://doi.org/10.1038/srep03399> PMID: 24292365
24. Margarido GR, Pastina MM, Souza AP, Garcia AA. Multi-trait multi-environment quantitative trait loci mapping for a sugarcane commercial cross provides insights on the inheritance of important traits. *Mol Breed.* 2015; 35: 175. <https://doi.org/10.1007/s11032-015-0366-6> PMID: 26273212
25. Balsalobre TWA, Pereira GS, Margarido GRA, Gazaffi R, Barreto FZ, Anoni CO, et al. GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics.* 2017; 18: 72. <https://doi.org/10.1186/s12864-016-3383-x> PMID: 28077090
26. Gerard D, Ferrao LFV, Garcia AAF, Stephens M. Genotyping polyploids from messy sequencing data. *Genetics.* 2018; 210: 789–807. <https://doi.org/10.1534/genetics.118.301468> PMID: 30185430
27. Pereira GS, Garcia AAF, Margarido GRA. A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids. *BMC Bioinformatics.* 2018; 19: 398. <https://doi.org/10.1186/s12859-018-2433-6> PMID: 30382832
28. Sardos J, Rouard M, Hueber Y, Cenci A, Hyma KE, van den Houwe I, et al. A genome-wide association study on the seedless phenotype in banana (*Musa* spp.) reveals the potential of a selected panel to detect candidate genes in a vegetatively propagated crop. *PLoS One.* 2016; 11: e0154448. <https://doi.org/10.1371/journal.pone.0154448> PMID: 27144345
29. Xiao Y, Liu H, Wu L, Warburton M, Yan J. Genome-wide association studies in maize: praise and stargaze. *Mol Plant.* 2017; 10: 359–374. <https://doi.org/10.1016/j.molp.2016.12.008> PMID: 28039028
30. Mwadzingeni L, Shimelis H, Rees DJ, Tsilo TJ. Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. *PLoS One.* 2017; 12: e0171692. <https://doi.org/10.1371/journal.pone.0171692> PMID: 28234945
31. Sun C, Zhang F, Yan X, Zhang X, Dong Z, Cui D, et al. Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai valley of China. *Plant Biotechnol J.* 2017; 15: 953–969. <https://doi.org/10.1111/pbi.12690> PMID: 28055148
32. Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 2017; 18: 161. <https://doi.org/10.1186/s13059-017-1289-9> PMID: 28838319
33. Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann J. Linkage disequilibrium among modern sugarcane cultivars. *Theor Appl Genet.* 1999; 99: 1053–1060.
34. Wei X, Jackson PA, McIntyre CL, Aitken KS, Croft B. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theor Appl Genet.* 2006; 114: 155–164. <https://doi.org/10.1007/s00122-006-0418-8> PMID: 17047910
35. Raboin LM, Pauquet J, Butterfield M, D'Hont A, Glaszmann JC. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theor Appl Genet.* 2008; 116: 701–714. <https://doi.org/10.1007/s00122-007-0703-1> PMID: 18196216

36. Yang X, Song J, Todd J, Peng Z, Paudel D, Luo Z, et al. Target enrichment sequencing of 307 germplasm accessions identified ancestry of ancient and modern hybrids and signatures of adaptation and selection in sugarcane (*Saccharum* spp.), a 'sweet' crop with 'bitter' genomes. *Plant Biotechnol J*. 2019; 17: 488–498. <https://doi.org/10.1111/pbi.12992> PMID: 30051590
37. Racedo J, Gutierrez L, Perera MF, Ostengo S, Pardo EM, Cuenya MI, et al. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biol*. 2016; 16: 142. <https://doi.org/10.1186/s12870-016-0829-x> PMID: 27342657
38. Wei X, Jackson PA, Hermann S, Kilian A, Heller-Uszynska K, Deomano E. Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. *Genome*. 2010; 53: 973–981. <https://doi.org/10.1139/G10-050> PMID: 21076513
39. Banerjee N, Siraree A, Yadav S, Kumar S, Singh J, Kumar S, et al. Marker-trait association study for sucrose and yield contributing traits in sugarcane (*Saccharum* spp. hybrid). *Euphytica*. 2015; 205: 185–201.
40. Siraree A, Banerjee N, Kumar S, Khan MS, Singh PK, Kumar S, et al. Identification of marker-trait associations for morphological descriptors and yield component traits in sugarcane. *Physiol Mol Biol Plants*. 2017; 23: 185–196. <https://doi.org/10.1007/s12298-016-0403-x> PMID: 28250594
41. Ukoskit K, Posudsavang G, Pongsiripat N, Chatwachirawong P, Klomsa-Ard P, Poomipant P, et al. Detection and validation of EST-SSR markers associated with sugar-related traits in sugarcane using linkage and association mapping. *Genomics*. 2019; 111: 1–9. <https://doi.org/10.1016/j.ygeno.2018.03.019> PMID: 29608956
42. Fickett N, Gutierrez A, Verma M, Pontif M, Hale A, Kimbeng C, et al. Genome-wide association mapping identifies markers associated with cane yield components and sucrose traits in the Louisiana sugarcane core collection. *Genomics*. 2018. <https://doi.org/10.1016/j.ygeno.2018.12.002> PMID: 30529701
43. Li J, Tang W, Zhang Y-W, Chen K-N, Wang C, Liu Y, et al. Genome-wide association studies for five forage quality-related traits in *Sorghum* (*Sorghum bicolor* L.). *Front Plant Sci*. 2018; 9: 1146. <https://doi.org/10.3389/fpls.2018.01146> PMID: 30186292
44. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet*. 2016; 12: e1005767. <https://doi.org/10.1371/journal.pgen.1005767> PMID: 26828793
45. Liu W, Maccaferri M, Chen X, Laghetti G, Pignone D, Pumphrey M, et al. Genome-wide association mapping reveals a rich genetic architecture of stripe rust resistance loci in emmer wheat (*Triticum turgidum* ssp. *dicoccum*). *Theor Appl Genet*. 2017; 130: 2249–2270. <https://doi.org/10.1007/s00122-017-2957-6> PMID: 28770301
46. Martinez SA, Godoy J, Huang M, Zhang Z, Carter AH, Campbell KAG, et al. Genome-wide association mapping for tolerance to preharvest sprouting and low falling numbers in wheat. *Front Plant Sci*. 2018; 9: 141. <https://doi.org/10.3389/fpls.2018.00141> PMID: 29491876
47. Zhang Y, Liu P, Zhang X, Zheng Q, Chen M, Ge F, et al. Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. *Front Plant Sci*. 2018; 9: 611. <https://doi.org/10.3389/fpls.2018.00611> PMID: 29868068
48. Xu Y, Yang T, Zhou Y, Yin S, Li P, Liu J, et al. Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. *Front Plant Sci*. 2018; 9: 1311. <https://doi.org/10.3389/fpls.2018.01311> PMID: 30233634
49. Cardoso-Silva CB, Costa EA, Mancini MC, Balsalobre TWA, Canesin LEC, Pinto LR, et al. *De novo* assembly and transcriptome analysis of contrasting sugarcane varieties. *PLoS One*. 2014; 9: e88462. <https://doi.org/10.1371/journal.pone.0088462> PMID: 24523899
50. VSN International. *Genstat for windows*. Hemel Hempstead, UK: VSN International; 2017.
51. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974; 19: 716–723.
52. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978; 6: 461–464.
53. Pastina MM, Malosetti M, Gazaffi R, Mollinari M, Margarido GR, Oliveira KM, et al. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theor Appl Genet*. 2012; 124: 835–849. <https://doi.org/10.1007/s00122-011-1748-8> PMID: 22159754
54. Holland JB, Nyquist WE, Cervantes-Martínez CT. Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev*. 2003; 22: 9–112.
55. Aljanabi SM, Forget L, Dookun A. An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. *Plant Mol Biol Report*. 1999; 17: 281.

56. Oliveira KM, Pinto LR, Marconi TG, Margarido GRA, Pastina MM, Teixeira LHM, et al. Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. *Mol Breed*. 2007; 20: 189–208.
57. Creste S, Neto AT, Figueira A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Report*. 2001; 19: 299–306.
58. Pinto LR, Oliveira KM, Ulian EC, Garcia AA, de Souza AP. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. *Genome*. 2004; 47: 795–804. <https://doi.org/10.1139/g04-055> PMID: 15499394
59. Oliveira KM, Pinto LR, Marconi TG, Mollinari M, Ulian EC, Chabregas SM, et al. Characterization of new polymorphic functional markers for sugarcane. *Genome*. 2009; 52: 191–209. <https://doi.org/10.1139/g08-105> PMID: 19234567
60. Cordeiro GM, Taylor GO, Henry RJ. Characterisation of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploid species. *Plant Sci*. 2000; 155: 161–168. PMID: 10814819
61. Wu KK, Burnquist W, Sorrells ME, Tew TL, Moore PH, Tanksley SD. The detection and estimation of linkage in polyploids using single-dose restriction fragments. *Theor Appl Genet*. 1992; 83: 294–300. <https://doi.org/10.1007/BF00224274> PMID: 24202510
62. Cordeiro GM, Pan Y-B, Henry RJ. Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. *Plant Sci*. 2003; 165: 181–189.
63. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010; 11: 94. <https://doi.org/10.1186/1471-2156-11-94> PMID: 20950446
64. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008; 24: 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129> PMID: 18397895
65. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013
66. Jombart T, Collins C. A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0. London, UK: Imperial College London, MRC Centre for Outbreak Analysis and Modelling; 2015.
67. Deperi SI, Tagliotti ME, Bedogni MC, Manrique-Carpintero NC, Coombs J, Zhang R, et al. Discriminant analysis of principal components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *PLoS One*. 2018; 13: e0194398. <https://doi.org/10.1371/journal.pone.0194398> PMID: 29547652
68. Perrier X, Flori A, Bonnot F. Data analysis methods. In: Hamon P, Seguin M, Perrier X, Enfield JCG, editors. Genetic diversity of cultivated tropical plants. Montpellier, France: Science Publishers; 2003. pp. 43–76.
69. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4: 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454> PMID: 3447015
70. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat*. 1979; 7: 1–26.
71. Efron B, Tibshirani RJ. An introduction to the bootstrap. Boca Raton, FL: Chapman & Hall; 1993.
72. Sinnwell JP, Therneau TM, Schaid DJ. The kinship2 R package for pedigree data. *Hum Hered*. 2014; 78: 91–93. <https://doi.org/10.1159/000363105> PMID: 25074474
73. Fay MP, Hunsberger SA, Nason M, Erin G. Package 'exact2x2'. 27 July 2018. Available from: <https://cran.r-project.org/web/packages/exact2x2/exact2x2.pdf> Cited 10 December 2018.
74. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003; 100: 9440–9445. <https://doi.org/10.1073/pnas.1530509100> PMID: 12883005
75. Kosambi DD. The estimation of map distances from recombination values. *Ann Eugen*. 1943; 12: 172–175.
76. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012; 28: 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444> PMID: 22796960
77. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995; 57: 289–300.
78. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 2012; 40: D1178–D1186. <https://doi.org/10.1093/nar/gkr944> PMID: 22110026

79. de Silva JAG, Costa PMDA, Marconi TG, Barreto EJDS, Solís-Gracia N, Park J-W, et al. Agronomic and molecular characterization of wild germplasm *Saccharum spontaneum* for sugarcane and energy-cane breeding purposes. *Sci Agric*. 2018; 75: 329–338.
80. Kelly AM, Smith AB, Eccleston JA, Cullis BR. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Sci*. 2007; 47: 1063–1070.
81. Malosetti M, Ribaut J-M, van Eeuwijk FA. The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Front Physiol*. 2013; 4: 44. <https://doi.org/10.3389/fphys.2013.00044> PMID: 23487515
82. Jamoza J, Owuoche J, Kiplagat O, Opile W. Broad-sense heritability estimation and correlation among sugarcane (*Saccharum* spp. hybrids) yield and some agronomic traits in western Kenya. *Int J Agric Policy Res*. 2014; 2: 16–25.
83. Soomro AF, Junejo S, Ahmed A, Aslam M. Evaluation of different promising sugarcane varieties for some quantitative and qualitative attributes under Thatta (Pakistan) conditions. *Int J Agric Biol*. 2006; 8: 195–197.
84. Shadmehr A, Ramshini H, Zeinalabedini M, Almani MP, Ghaffari M, Darbandi AI, et al. Phenotypic variability assessment of sugarcane germplasm (*Saccharum officinarum* L.) and extraction of an applied mini-core collection. *Agriculture*. 2017; 7: 55.
85. Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, et al. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc Natl Acad Sci U S A*. 2013; 110: 8057–8062. <https://doi.org/10.1073/pnas.1217133110> PMID: 23630259
86. Pometti CL, Bessega CF, Saidman BO, Vilardi JC. Analysis of genetic population structure in *Acacia caven* (*Leguminosae*, *Mimosoideae*), comparing one exploratory and two Bayesian-model-based methods. *Genet Mol Biol*. 2014; 37: 64–72. PMID: 24688293
87. Ertiro BT, Semagn K, Das B, Olsen M, Labuschagne M, Worku M, et al. Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. *BMC Genomics*. 2017; 18: 777. <https://doi.org/10.1186/s12864-017-4173-9> PMID: 29025420
88. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155: 945–959. PMID: 10835412
89. Corander J, Waldmann P, Sillanpää MJ. Bayesian analysis of genetic differentiation between populations. *Genetics*. 2003; 163: 367–374. PMID: 12586722
90. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes*. 2007; 7: 574–578. <https://doi.org/10.1111/j.1471-8286.2007.01758.x> PMID: 18784791
91. Huang E, Aitken K, George A. Association studies. In: Henry R, Kole C, editors. *Genetics, genomics and breeding of sugarcane*. Enfield, NH: Science Publishers; 2010. pp. 43–68.
92. Matsuoka S, Garcia A, Calheiros G. Hibridação em cana-de-açúcar. In: Borém A, editor. *Hibridação artificial de plantas*. Viçosa: Editora UFV; 1999. pp. 221–254.
93. Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol*. 2003; 54: 357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907> PMID: 14502995
94. Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor Appl Genet*. 2005; 111: 723–730. <https://doi.org/10.1007/s00122-005-2057-x> PMID: 15997389
95. Newman DL, Abney M, McPeck MS, Ober C, Cox NJ. The importance of genealogy in determining genetic associations with complex traits. *Am J Hum Genet*. 2001; 69: 1146–1148. <https://doi.org/10.1086/323659> PMID: 11590549
96. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. 2006; 38: 203–208. <https://doi.org/10.1038/ng1702> PMID: 16380716
97. Müller BU, Stich B, Piepho HP. A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants. *Heredity*. 2011; 106: 825–831. <https://doi.org/10.1038/hdy.2010.125> PMID: 20959861
98. Wei W, Mesquita ACO, Figueiró ADA, Wu X, Manjunatha S, Wickland DP, et al. Genome-wide association mapping of resistance to a Brazilian isolate of *Sclerotinia sclerotiorum* in soybean genotypes mostly from Brazil. *BMC Genomics*. 2017; 18: 849. <https://doi.org/10.1186/s12864-017-4160-1> PMID: 29115920

99. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006; 2: e190. <https://doi.org/10.1371/journal.pgen.0020190> PMID: 17194218
100. McVean G. A genealogical interpretation of principal components analysis. *PLoS Genet.* 2009; 5: e1000686. <https://doi.org/10.1371/journal.pgen.1000686> PMID: 19834557
101. Garcia AA, Kido EA, Meza AN, Souza HM, Pinto LR, Pastina MM, et al. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theor Appl Genet.* 2006; 112: 298–314. <https://doi.org/10.1007/s00122-005-0129-6> PMID: 16307229
102. Bourke PM, Voorrips RE, Visser RGF, Maliepaard C. Tools for genetic studies in experimental populations of polyploids. *Front Plant Sci.* 2018; 9: 513. <https://doi.org/10.3389/fpls.2018.00513> PMID: 29720992
103. Casu RE, Dimmock CM, Chapman SC, Grof CP, McIntyre CL, Bonnett GD, et al. Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. *Plant Mol Biol.* 2004; 54: 503–517. <https://doi.org/10.1023/B:PLAN.0000038255.96128.41> PMID: 15316286
104. Casu RE, Rae AL, Nielsen JM, Perroux JM, Bonnett GD, Manners JM. Tissue-specific transcriptome analysis within the maturing sugarcane stalk reveals spatial regulation in the expression of cellulose synthase and sucrose transporter gene families. *Plant Mol Biol.* 2015; 89: 607–628. <https://doi.org/10.1007/s11103-015-0388-9> PMID: 26456093
105. Cui Y, Zhao Q, Gao C, Ding Y, Zeng Y, Ueda T, et al. Activation of the Rab7 GTPase by the MON1-CCZ1 complex is essential for PVC-to-vacuole trafficking and plant growth in *Arabidopsis*. *Plant Cell.* 2014; 26: 2080–2097. <https://doi.org/10.1105/tpc.114.123141> PMID: 24824487
106. Sakamoto W, Takami T. Chloroplast DNA dynamics: copy number, quality control and degradation. *Plant Cell Physiol.* 2018; 59: 1120–1127. <https://doi.org/10.1093/pcp/pcy084> PMID: 29860378
107. Ha CV, Le DT, Nishiyama R, Watanabe Y, Sulieman S, Tran UT, et al. The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress. *DNA Res.* 2013; 20: 511–524. <https://doi.org/10.1093/dnares/dst027> PMID: 23810914
108. Korasick DA, Westfall CS, Lee SG, Nanao MH, Dumas R, Hagen G, et al. Molecular basis for auxin response factor protein interaction and the control of auxin response repression. *Proc Natl Acad Sci U S A.* 2014; 111: 5427–5432. <https://doi.org/10.1073/pnas.1400074111> PMID: 24706860
109. Li S-B, Xie Z-Z, Hu C-G, Zhang J-Z. A review of auxin response factors (ARFs) in plants. *Front Plant Sci.* 2016; 7: 47. <https://doi.org/10.3389/fpls.2016.00047> PMID: 26870066
110. Hsu PY, Benfey PN. Small but mighty: functional peptides encoded by small ORFs in plants. *Proteomics.* 2018; 18: e1700038. <https://doi.org/10.1002/pmic.201700038> PMID: 28759167