Field Crops Research 288 (2022) 108678

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

Nutritional composition of commercial sugarcane (*Saccharum* spp.) genotypes evaluated over regrowth cycles in different environments in Brazil

Monalisa S. Carneiro [a,*], Rodrigo Gazaffi [b], Nara F. Pistarini [c], Rodolpho T.G. Lorencetti [c], Eduardo A. Bressan [d], Edson F. da Silva [e], Giancarlo C.X. Oliveira [d], Maria Teresa M.R. Borges [c]

[a] *Plant Biotechnology Laboratory, Department of Biotechnology, Vegetal and Animal Production, Center of Agricultural Sciences, Federal University of São Carlos, Highway Anhanguera Km 174, Araras CEP 13600-970, SP, Brazil*
[b] *Department of Biotechnology, Vegetal and Animal Production, Center of Agricultural Sciences, Highway Anhanguera Km 174, Araras CEP 13600-970, SP, Brazil*
[c] *Technological Analysis and Simulation Laboratory, Department of Agroindustrial Technology and Rural Socioeconomics, Center of Agricultural Sciences, Federal University of São Carlos, Highway Anhanguera Km 174, Araras CEP 13600-970, SP, Brazil*
[d] *Evolution Laboratory, Department of Genetics, "Luiz de Queiroz" Agricultural College, University of São Paulo, 11 Pádua Dias Ave., Piracicaba CEP 13418-900, SP, Brazil*
[e] *Plant Breeding Laboratory, Biology Department, Federal Rural University of Pernambuco, Dom Manoel de Medeiros St., Recife CEP 52171-930, PE, Brazil*

ARTICLE INFO

Keywords:
Phenotypic plasticity
Genetic correlation
Heritability
Variance-covariance structures
Geographical variation

ABSTRACT

Understanding genetic variation, genotype × environment interactions and associations between nutritional component traits is crucial for the development of improved sugarcane cultivars. Additionally, this information can be useful to identify a wide range of variation in these traits by taking into account different varietal profiles, production environments and years. Here, we report for the first time the evaluation of G × E interactions analyzed for characteristics associated with the "nutritional composition" of sugarcane, as well as the "technological composition". This study involved the analysis of 20 nutritional composition characteristics evaluated in 20 commercial sugarcane varieties in six production environments analyzed across two agricultural years. The present study generated an unpublished database of 14,300 experimental data points on sugarcane nutritional composition that can be used in studies on the substantial equivalence for future genetically modified cultivars. The genetic heritability and genotypic correlation of sugarcane compositional characteristics were also estimated. For the traits reducing sugar of juice, lipids, ash and protein, the heritability value was zero because their genetic variances were also estimated as zero. Thus, all the phenotypic variation associated with these traits is due to non-heritable factors. For all traits, the difference between the minimum and maximum value of the adjusted means exceeded at least 50%. The observed results indicated that there was a strong influence of GxE interactions on the phenotypic variations of the characteristics associated with the nutritional and technological composition of sugarcane.

1. Introduction

Sugarcane (*Saccharum* spp.) is one of the most economically important grasses in the world, showing a high bioenergy potential due to its C4 nature and high biomass yield. The main product is sugar, but it is also an important bioenergy feedstock exemplified by bioethanol and bioelectricity; more recently, interest has also relied on bioplastics (Furlan et al., 2013; Hoang et al., 2015; Saini et al., 2015; Hiloidhari et al., 2018; Silveira et al., 2018). Genetically, sugarcane has a complex genome with a size of approximately 10 Gb due to a high and variable ploidy level, with frequent aneuploidy (D'Hont, 2005; Piperidis et al., 2020; Thirugnanasambandam et al., 2018).

Currently, sugarcane breeding aims to obtain new varieties with tolerance to abiotic stress (e.g., drought), resistance to pests and

Abbreviations: CV, coefficient of variation; G × E, genotype × environment interaction; MET, multi-environment trials; VCOV, variance-covariance matrix; CTNBio, National Technical Biosafety Commission; Gb, gygabase; GM, genetically modified.

* Correspondence to: Center of Agricultural Sciences, Federal University of São Carlos, Highway Anhanguera Km 174, Araras CEP 13600-970, SP, Brazil.
E-mail address: monalisa@ufscar.br (M.S. Carneiro).

https://doi.org/10.1016/j.fcr.2022.108678
Received 20 December 2021; Received in revised form 26 August 2022; Accepted 3 September 2022
Available online 22 September 2022

diseases, mechanical harvesting features (e.g., sprout, tillering), and mainly high sucrose content (with high quality) per hectare; the last can be achieved either from increasing biomass productivity or sugar content (de Morais et al., 2015). The sugarcane pipeline breeding can vary between different programs. However, a similar structure can be defined, i.e., the procedure starts with several crosses among superior genotypes (frequently commercial cultivars) to obtain a population with genetic variability, followed by several selection stages until the target genotype is identified. The selection stages explore sugarcane clonal propagation ability; for this, two major consequences are observed: the hole genetic variation can be exploited by either additive and non-additive effects over stages; and, historically, a few recombination events are observed from the ancestors to the modern varieties (Yadav et al., 2020). According to each breeding program, obtaining a new cultivar can vary from 11 to 12 years (Cheavegatti-Gianotto et al., 2011; Cursi et al., 2021).

Sugarcane history has indicated that the breeding program follows a recurrent selection process (Jackson, 2005), in which the newly selected genotypes are included as new parents for another cycle. This approach is a successful example of breeding, worldwide data indicate that sugarcane yield increasing41% during the last 50 years, and sugar production increasing by 1–1.5% per year (Gouy et al., 2013, 2015). On the other hand, yield gains rate are decreasing over time, suggesting a possible plateau in sugar accumulation (Dal-Bianco et al., 2012). New sugarcane breeding strategies are essential to overcome this limitation and reduce the demanding time to obtain a new cultivar.

In this aim, Biotechnology is a highlighted strategy, integrating new tools for breeding (Dal-Bianco et al., 2012), such as 'omics' data, molecular markers, and Genetic Engineering. The development of sugarcane genetically modified (GM) cultivars has received special attention as a possible alternative to increase genetic gain. In 2017, the National Technical Biosafety Commission (CTNBio), responsible for the evaluation and commercial approval of genetically modified cultivars in Brazil, approved the cultivation of the first GM sugarcane developed to control the sugarcane borer in Brazilian fields (Gianotto et al., 2019). Approval for commercial planting of GM cultivars depends on a rigorous interdisciplinary risk assessment process in which the potential environmental impact and the food and feed safety are evaluated (Kennedy et al., 2018).

A critical aspect of biosafety analysis of GM cultivars is the risk assessment for human and animal health. Normative Resolution No. 05 of CTNBio states that GM organisms used as food must be evaluated for substantial equivalence to the non-modified cultivar. The Organization for Economic Co-operation and Development (OECD) developed the concept of substantial equivalence to ensure that new biotechnology-derived foods are safe as their conventional counterparts. Specifically, for sugarcane, the OECD recommends that the new cultivar be analyzed for its main component contents (moisture, crude protein, lipids, ashes, fibers, and sucrose) (OECD, 2011). There is national scientific literature on the topic (Azevedo et al., 2003; de Souza França et al., 2004; Santos et al., 2006), but most studies have focused on the use of sugarcane as silage. As required by CTNBio and recommended by the OECD, there is currently no database containing the nutritional information of national sugarcane cultivars.

In this context, the lack of information about the nutritional characteristics of conventional sugarcane cultivars creates a challenging scenario to assess substantial equivalence with GM sugarcane. For this, we aim to obtain a reference database of substantial equivalence of future genetically modified cultivars, as recommended by the OECD report, for the sugarcane Nutritional Composition traits and Technological Composition. These data are collected based on Brazilian cultivated germplasm evaluated under multi-environmental trials.

2. Materials and methods

2.1. Plant material and experimental conditions in the fields

A collection of twenty relevant commercial sugarcane cultivars was used in this study. The criteria for choosing the evaluated varieties were proportion of the planted area that they occupy 63.2%, maturation with respect to sucrose accumulation and adaptability to different production environments. Varieties from the main Brazilian breeding programs were also chosen, such as RIDESA public program (Inter-university Network for the Development of the Sugarcane Industry), which develop varieties with the acronym RB; COPERSUCAR's private program (produced SP varieties) that ended its activities in 2003, and since 2004, it has been managed by the Sugarcane Technology Center (CTC), which has been developing CTC varieties; CanaVialis, whose varieties are named CV; and IAC program (Agronomic Institute of Campinas) that produced varieties with the acronyms IACSP and IAC (Cursi et al., 2021). According to the maturation period of the cultivars, they can be classified as early (RB855156, RB855453, RB965917, RB966928, CV7231, CTC9, CTC17 and CTC21) or late cultivars (encompasses the medium/late cultivars: RB92579, RB835054, RB867515, RB965902, IACSP955000, IACSP955094, CV7870, SP81–3250, SP83–2847, CTC4, CTC15 and CTC20), thus this study encompassed materials from the main Brazilian breeding programs for sugarcane. Still, twelve of them were cited as the most cultivated in Brazil in 2019, in which the Brazilian sugarcane census was based on 6.8 mi hectares (Braga Jr et al., 2021). These cultivars were already present in the national census since 2011, except for CV7870, which appeared in the census in 2014. Regarding the regional censuses, 18 of them were present in at least one regional area, an indication that the cultivar sample also includes several environmental production areas.

The locations chosen in this research represent the main growing areas of sugarcane in Brazil. The States of São Paulo, Paraná, Pernambuco, and Goiás together represent 71.5% of the sugarcane growing area in Brazil. The trials were conducted in six environments: Conchal (lat 22°24'S, long 47°06'W and alt 591 m asl, in the state of São Paulo), Jaboticabal (lat 21°16'S, long 48°23'W and alt 615 m asl, in the state of São Paulo), Taciba (lat 22° 23'S, long 51°17'W and alt 416 m asl, in the state of São Paulo), Rolândia (lat 23°18'S, long 51°22'W and alt 730 m asl, in the state of Parana), Montividiu (lat 17°26'S, long 51°10'W and alt 821 m asl, in the state of Goias), and Carpina (lat 07°35'S, long 34°15'W and alt 184 m asl, in the state of Pernambuco). The climates were Cwa (Jaboticabal and Conchal), Aw Montividiu, As (Carpina), Cfa (Rolândia and Taciba), according to Köppen (Fig. 1) (Supplementary Table 4). At each location, the experimental design consisted of a randomized block design, which was fully replicated three times. Trial plots consisted of two 3 m-long rows spaced 1.4 m apart. Weed, fertilizer and pest and disease management proceeded per commercial farm practice. The experimental plots were installed in 2014, and harvesting, according to maturation period, was carried out in 2015 (plant cane) and 2016 (first ratoon). A 10-stalk sample was taken for analysis of the nutritional composition, and three replicates at each location were evaluated.

2.2. Nutritional composition analysis

Sugarcane cultivars were evaluated for 20 nutritional composition traits described in Supplementary Table 1. In 2015, sugarcane stalk samples from the Conchal, Jaboticabal, Taciba, Rolândia and Montividiu environments were collected and sent directly to the laboratory where they were processed as indicated in Supplementary Fig. 1. All the samples from 2016 and also the sample from Carpina of 2015 were collected, disintegrated and homogenized in the field, and approximately 1.5 kg were packed and frozen for transportation to the

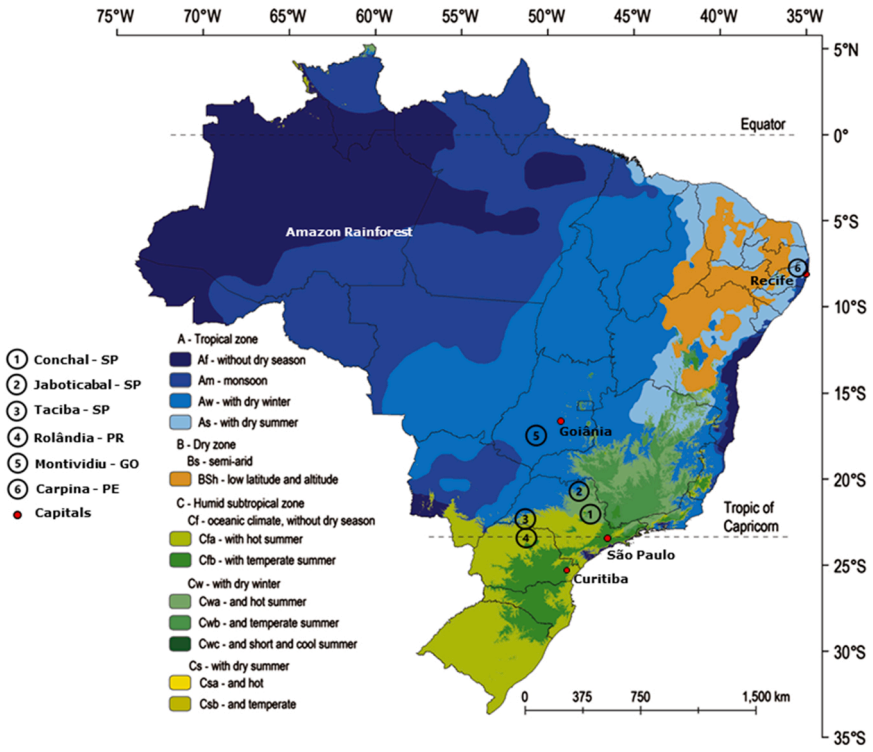

Fig. 1. Geographical positioning of the six environments in which the experiments were carried out across different climatic regions of Brazil (Adapted from Alvares et al., 2014).

laboratory.

All samples were analyzed according to Supplementary Figs. 1 and 2. In the laboratory, the samples were thawed, and from an initial mass of 1.5 kg, subsamples of 100 g were used to quantify the moisture level based on the mass difference after drying them in an air circulation oven at 100–105 °C to constant weight (AOAC 935.29). A fraction of approximately 200 g was dried to a constant weight and ground for use in the analysis of the nutritional composition.

A fraction of 500 g was applied using a hydraulic press with a pressure of 250 kg/cm^2 per minute for one minute to extract and separate the wet bagasse from the juice. Wet bagasse was used to determine the fiber content according to the Tanimoto method, ABNT NBR 16225 (2013)) (Supplementary Table 1). The juice was submitted to an analysis of soluble solids in a Reichert R2I300 refractometer (ABNT NBR 16223:2013) and of sucrose by polarimetry using a Micronal-Saccharomat polarimeter (ABNT NBR 16224:2013). Before performing the polarimetric analysis, the juice went through a clarification process involving a mixture composed of aluminum chloride, calcium hydroxide and Celite (ICUMSA Method GS5/7–28) and subsequent filtration. The fiber content was determined from the pressing residue, where the fiber was calculated by the sugarcane payment system method by sucrose content (F_{PCTS}%). Additionally, the pressing residue was weighed after drying to a constant weight and determined by the Tanimoto method (F_{tan}%). From these determinations, it was possible to obtain the estimated values of juice reducing sugar %, cane reducing sugar and total cane reducing sugar %, for the two methods applied (payment system and Tanimoto method), including AR%, ARC_{PCTS}%, ARC_{Tan}%, $ARTC_{PCTS}$%, and $ARTC_{Tan}$%. The value of total recoverable sugar in kg/ton cane (ATR%) was also calculated. For the nutritional composition, the moisture content (U%) and ash (ash%) values were determined by drying and burning at 100–105 °C and 600–650 °C, respectively (AOAC 935.92 and 942.05); protein (Pt%) was determined by the Kjeldahl method, where the total amount of nitrogen was quantified (AOAC 2001.11); total lipids (Lp%) were obtained by ether extraction (AOAC 2003.06); FDN and FDA were determined by ANCOM Method 12 and Method 13; and finally, the hemicellulose content (HC%) was obtained by an algebraic method based on the difference between FDN and FDA. The data generated in this study are available in the supplementary material (Supplementary Table 2).

2.3. Statistical analyses

The first step was to consider a statistical model to obtain the estimates of the variance components for each trait. The genotypic, location and harvest sources of variation were modeled, as well as their interactions and the maturation period. The statistical design of completely randomized blocks was used with three replications. The statistical model is.

(Eq. 1):

$$y_{ijkwt} = \mu + G_i + L_j + H_k + B_{w(jk)} + M_t + GL_{ij} + GH_{ik} + LH_{jk} + GLH_{ijk} + GLB_{ijw(k)} + e_{ijkwt}, \quad (1)$$

where y_{ijkwt} is the phenotypic value for the i-th genotype (i = 1,…,20), (i = 1,…,20), classified as the t-th maturation group at the j-th location (j=1,…,6), the k-th harvest (k = 1 and 2) and the w-th repetition (w=1,2,3); μ is the intercept of the model; G_i is the random effect for genotypes with $N(0, \sigma^2_G)$; L_j and H_k are the fixed effects for location and harvests, respectively; $B_{w(jk)}$ is the fixed effects for the blocks nested in the j-th location and the k-th harvest; M_t is the covariate for maturation period, and the genotypes were classified as early or late maturing lines; LH_{jk} is the fixed location by harvest interaction; GL_{ij} and GH_{ik} are the random two-way interactions for genotype by location [$N(0, \sigma^2_{GL})$] and genotype by harvest [$N(0, \sigma^2_{GH})$], respectively; GLH_{ijk} is the three-way random interaction term between genotype, location and harvest, with $N(0, \sigma^2_{GLH})$; $GLB_{ijw(k)}$ is the error term originated from the repeated measures observations (over harvests); and e_{ijkw} is the error term, with $N(0,\sigma^2_E)$. The estimates were obtained by a mixed models approach using GenStat 16 software (Payne et al., 2009). Based on variance component

estimates, phenotypic variation ($\widehat{\sigma}_p^2 = \widehat{\sigma}_G^2 + \widehat{\sigma}_{GL}^2 + \widehat{\sigma}_{GH}^2 + \widehat{\sigma}_{GLH}^2 + \widehat{\sigma}_{GLB}^2 + \widehat{\sigma}_E^2$) and heritability ($h^2 = \frac{\widehat{\sigma}_G^2}{\widehat{\sigma}_P^2}$) were obtained. The coefficients of variation (CV) were calculated considering the error term ($CV_R = \frac{\widehat{\sigma}_E}{\widehat{\mu}}$) and the genetic variation ($CV_G = \frac{\widehat{\sigma}_G}{\widehat{\mu}}$) to measure the experimental precision and the genetic variability available, respectively. The fixed effects were tested using the Wald's test.

In the second round of analyzes for each trait, a flexible approach was assumed to better model the interactions in a more realistic way, so we let different values of variance and genetic covariance be assumed and included heteroscedasticity for the error term, i.e., a different variance may be estimated for the error term of each combination of location and harvest. For this situation, the following model was considered (Eq. 2):

$$y_{ijkwt} = \mu + L_j + H_k + LH_{jk} + B_{jkw} + M_t + GLH_{ijk} + e_{ijkwt}, \quad (2)$$

where all elements in this statistical model were already described in the previous one, except GLH_{ijk}, which is the random effect for three-way interactions, now assuming $N(0, \Sigma_{LH})$, in which $\Sigma_{LH} = \Sigma_L \bigotimes \Sigma_H$; here, Σ_{LH} represents the variance-covariance (vcov) matrix for combinations of location and harvest, which is obtained by the Kronecker product of vcov for location (Σ_L, with 6 ×6 dimensions) and for harvest (Σ_H, with 2 ×2 dimensions). Several structures for Σ_L and Σ_H were assumed. Three basic matrices for Σ were tested: homogeneous genetic variance (ID), heterogeneous genetic variance (DIAG) and unstructured model (UN). The combination of these three structures for location and harvest provided nine possible situations. To select the best combination for Σ_L and Σ_H, the Akaike information criterion (AIC) was used. With the selected vcov structure, variance components (genetic and environmental variances) were estimated for each of the twelve combinations of location (six) and harvest (two). Parameters such as heritability ($h^2 = \frac{\widehat{\sigma}_G^2}{\widehat{\sigma}_G^2 + \widehat{\sigma}_E^2}$), residual coefficient of variance and genetic coefficient of variance were also calculated. This procedure was applied for the traits that showed a genetic variance higher than zero. Also, the fixed effects were tested based on Wald's test.

Adjusted means were also predicted, allowing us to verify the minimum, average and maximum values for each trait. However, genotypic correlations between traits were inferred and it was found that these correlations did not vary between locations. This allowed us to obtain several genetic correlation matrices. For the average matrix, we tested all pairwise correlation coefficients ($H_0 : \rho = 0$) with a 5% significance level. The same procedure was applied to verify whether the correlations differed when the maturation time was different, i.e., the genetic correlations between the early- and late-maturing groups were inferred and compared between groups. All graphics and correlation results were obtained using R software and the library ggplot2.

Finally, the similarity between locations was estimated using cluster analysis. Briefly, this procedure is based on the assumption that similar locations tend to show similar genotype performance (Bernardo, 2010). Thus, a clustering analysis was performed for locations using the adjusted means as entries. Usually, this approach can be applied for a given trait, for example, the one that can be classified as the most important. Considering this study is based on nutritional composition traits, a priori, all traits are equally important, the complete interpretation of GxE interaction insight could be applied if this procedure were applied to each trait. However, summarizing the cluster analysis for each trait would not be efficient for interpretation purposes. For this, a multivariate approach was chosen to deal with three-way data (genotypes evaluated in different locations over the years). The Tucker3 approach (Tucker, 1966) is a possible generalization of the two-way principal components analysis. The data array can be decomposed into four structures, corresponding to the genotype (A matrix), the location (B matrix), the trait information (C matrix), and a central array (G) that connects the previous matrices. Here, one can define the number of linear combinations for a genotype, location, and trait. The considered model had four parameters for genotypes (g1, g2, g3, and g4), three for locations (q1, q2, and q3), and three for traits (r1, r2, and r3). Tucker's main advantage is the understanding of the simultaneous association between the levels of each entry. Additionally, considering that the traits have different scales, the adjusted means were previously standardized to 0 and the variances to 1. In this context, the linear combinations for the C matrix were used to obtain new two-way tables composed of genotypes and locations. A clustering algorithm (UPGMA) was applied to the table for locations. The results were indicated by dendrograms. For the decomposition using the Tucker3 model, a similarly named function implemented in the rrcov3way package that runs in the R software was used.

3. Results

In the first step, a joint analysis considering twenty sugarcane varieties evaluated at six different Brazilian sites (environments) over two harvests, i.e., plant cane (2015) and first ratoon (2016), was performed for each of twenty nutritional components. The results are summarized in Table 1, which contains the general average, the genetic and phenotypic variances, the genetic (CV_G) and residual (CV_R) coefficient of variation, and the broad sense heritability.

Nutritional composition traits evaluated across the different chemical-analytical procedures provided similar results (Table 1), i.e., $PolC_{PCTS}$% and $PolC_{Tan}$% (13.59 and 13.49, respectively), $ARTC_{PCTS}$ % and $ARTC_{Tan}$ % (14.91 and 14.80), ARC_{PCTS} % and ARC_{Tan}% (0.60 and 0.60), and F_{Tan}% and F_{PCTS} % (13.76 and 13.38). FDN% showed a similar average (13.15) compared to F_{Tan}% and F_{PCTS}%. The traits with the lowest averages were Lp% (0.53), Pt% (0.54), and Ash% (0.59).

The same pattern was observed for CV_R and CV_G, for example, $PolC_{PCTS}$% and $PolC_{Tan}$% (showed CV_R of 0.08 and CV_G of 0.02 for both traits) and $ARTC_{PCTS}$ % and $ARTC_{Tan}$ % (showed CV_R of 0.07 and CV_G of 0.02 for both traits). The lowest CV_R values were identified for U% (0.02), Pza% (0.04) and Bx% (0.06). On the other hand, the highest values for CV_R were found for Lp% (0.29), Ash% (0.24), Pt% (0.17). The

Table 1

Averages, estimates of the genetic ($\widehat{\sigma}_G^2$) and phenotypic ($\widehat{\sigma}_P^2$) components of variance, coefficients of genotypic (CV_G) and residual (CV_R) variation, and broad-sense heritability of a genotypic mean (h^2) for 20 nutritional composition traits evaluated in 20 commercial cultivars of sugarcane in six different production environments (Conchal, Jaboticabal, Taciba, Rolândia, Montividiu and Carpina, Brazil) over two harvest years (2015 and 2016).

Traits	Average	$\widehat{\sigma}_G^2$	$\widehat{\sigma}_P^2$	CV_R	CV_G	h^2
Bx%	19.35	0.06	1.75	0.06	0.01	0.04
Pol%	16.44	0.08	2.50	0.08	0.02	0.03
$PolC_{PCTS}$%	13.59	0.07	1.76	0.08	0.02	0.04
$PolC_{Tan}$%	13.49	0.11	1.79	0.08	0.02	0.06
Pza%	84.82	0.25	11.52	0.04	0.01	0.02
F_{Tan}%	13.76	0.71	2.49	0.09	0.06	0.29
F_{PCTS} %	13.38	0.24	1.40	0.07	0.04	0.17
$ARTC_{PCTS}$ %	14.91	0.08	1.74	0.07	0.02	0.04
$ARTC_{Tan}$ %	14.80	0.12	1.78	0.07	0.02	0.07
ATR kg.t^{-1} Cane	147.50	6.35	140.04	0.07	0.02	0.05
U %	70.31	0.53	3.66	0.02	0.01	0.15
FDN%	13.15	0.65	2.51	0.09	0.06	0.27
FDA%	8.59	0.36	1.48	0.12	0.07	0.24
HC%	4.58	0.08	1.05	0.22	0.05	0.06
AR%	0.73	0.00	0.01	0.14	0.00	0.00
ARC_{PCTS} %	0.60	0.00	0.01	0.14	0.00	0.00
ARC_{Tan}%	0.60	0.00	0.01	0.14	0.00	0.00
Lp%	0.53	0.00	0.03	0.29	0.00	0.00
Ash%	0.59	0.00	0.02	0.24	0.00	0.00
Pt%	0.54	0.00	0.01	0.17	0.00	0.00

average CV_R was approximately 0.11 (or 11%). CV_G indicates the standardized genetic variability allowing the comparison between traits. The highest values were observed for fiber-associated traits, such as FDA % (0.07), FDN% (0.06), F_{Tan}% (0.06), HC% (0.05), and F_{PCTS}% (0.04). The lowest values for CV_G were zero and were found for Lp%, Ash%, Pt %, ARC_{Tan}%, AR% and ARC_{PCTS}%. The non-zero lowest values were 0.01 (Pza%), 0.01(U%), and 0.01 (Bx%). The sugar-associated traits showed a CV_G varying from 0.01 (Bx%) to 0.02 ($PolC_{Tan}$%).

The heritability follows the same pattern as CV_G was low for all the traits (Table 1), with an average of 0.08 (or 8%). The highest values were also found for fiber-associated traits (F_{PCTS}%, 0.17; F_{tan}%, 0.29; FDN%, 0.27; and FDA%, 0.24). HC% was the only trait in the fiber group with a low heritability value (0.06). On the other hand, lower heritability was observed for traits associated with sugar (Bx%, Pol%, $PolC_{PCTS}$%, $PolC_{Tan}$%, AR%, ARC_{PCTS}%, ARC_{Tan}%, $ARTC_{PCTS}$%, $ARTC_{Tan}$%, and ATR $kg.ton^{-1}$), varying from 0.07 to 0.03. The heritability for U% was 0.15, which is an intermediate value between those for fiber-associated traits and sugar-associated traits. The lowest non-zero heritability was found for Pza% (0.02). For the traits AR%, ARC%, ARC_{Tan}%, Lp%, Ash% and Pt %, the heritability value was zero as a consequence of their genetic variance being estimated as zero. In other words, all the phenotypic variation associated with these six traits is due to a non-heritable factor. Additionally, the same traits showed the highest CV_R values, and the highest heritabilities were associated with a high CV_G. For this reason, we did not use these traits in the following steps.

As the second step, to best capture GxE interactions, the second statistical model was assumed (Eq. 2), where a mixed models approach was used to study the three-way interaction component (GLH_{ijk}) with different structures. Using the AIC criterion, the best variance-covariance matrices for both location and harvest structure were the unstructured for Bx%, Pol%, $PolC_{PCTS}$%, $PolC_{Tan}$%, Pza%, F_{Tan}%, F_{PCTS} %, $ARTC_{PCTS}$, $ARTC_{Tan}$, ATR, U %, FDN%, and FDA%, except for HC%, in which location was modeled as an unstructured matrix and harvest was modeled as an identity (homogenous variance between harvests without a correlation between them) (AIC = 1868.76) (Supplementary Table 3 - Material A). The second best model was the one adopted for the other traits (AIC = 1869.44, difference of 0.68). In this case, the same variance-covariance structure was kept for all the traits. The residual (microenvironmental) term has varied from each combination of location and harvest levels. Notably, the mixed models estimate provided changeable results for genetic and environmental parameters (Supplementary Table 3 - Material B). The 12 heritability and CV_R were obtained for a general overview (Table 2). The average CV_R obtained for a given trait was similar to the one in Table 1, i.e., the highest difference was observed for HC% (0.19 in Table 2 vs. 0.22 in Table 1), while the other differences were based on the third decimal place.

Heritability values varied when the GE interaction was modeled by variance-covariance matrices (Supplementary Table 3 - Material B). For example, F_{Tan} varied from 0.17 (Plant cane at Montevidiu) to 0.73 (First ratoon at Carpina), with a mean of (0.42). FDN% showed the second highest average heritability (0.41), ranging from 0.22 (First ratoon at Conchal) to 0.56 (Plant cane at Taciba). The lowest average heritability was observed for HC% (0.14), ranging from 0.00 (in both years at Carpina) to 0.35 (plant cane at Montevidiu). The sugar-associated traits had average heritability near 0.33 ($ARTC_{Tan}$, $ARTC_{PCTS}$%, ATR, Bx%, Pol%, $PolC_{PCTS}$%, and $PolC_{Tan}$%), and U% showed a value of 0.34. Here, average heritabilities in Table 2 were higher than in the first Table 1. It is also noteworthy that the results of the Wald test for fixed effects, such as location and the interaction between location and year, were always significant, i.e., both are important sources of variation for the data.

Table 2 also presents some descriptive statistics (minimum, maximum, general average, early-maturing average, and late-maturing average), which can define the range for each trait, e.g., Bx% had the lowest mean value (15.84) in Conchal during the first year for RB92579 and the maximum (23.15) in Carpina during the first year for CV7231 (Supplementary Table 3- Material C), with an average of 19.35. The standardized range ($\frac{100(max-min)}{average}$) allows the direct comparison between traits U% and Pza% were the two traits with the smallest range (16.17% and 17.14%, respectively). The highest variations were observed for FDA% (85.03%) and HC% (83.55%). The sugar-related traits showed values of approximately 40%, and for the fiber-related traits the values were at least 57%. Table 2 also shows the averages of an early-maturing group (eight varieties) versus a late-maturing group (twelve varieties). For example, Bx% has a value of 19.64 for the early-maturing group and 19.06 for the late-maturing group, and the groups are significantly distinct from each other. The early- and late-maturing varieties differ for all traits except U%, FDA%, and HC%. However, it should be stressed that these values are very similar in practical purposes.

The association between the traits was also investigated. First, it was obtained the overall mean of each trait, i.e., the adjusted mean

Table 2
Range, averages, residual variation coefficient (CV_R), and broad-sense heritability considering the genotype average (h^2) for 14 nutritional composition traits evaluated in 20 cultivars of sugarcane in six different environments (E) in Brazil over two harvest years (Y) (2015 and 2016). The Akaike information criterion (AIC) was used to estimate the structures of the variance-covariance matrices.

Trait	(Range) Min-Max	Avg[a]	% of Range/Avg	Early Avg[b]	Late Avg[c]	CV_R	h^2	E (p value)[d]	ExY (p value)[e]
$ARTC_{Tan}$ %	11.51–17.35	14.78	39.51	15.15	14.47	0.07	0.33	< 0.001	< 0.001
$ARTC_{PCTS}$ %	11.69–17.88	14.91	41.52	15.21	14.61	0.07	0.32	< 0.001	< 0.001
ATR $kg.t^{-1}$ Cane	117.8–175.6	147.52	39.18	150.20	144.8	0.07	0.33	< 0.001	< 0.001
Bx %	15.84–23.15	19.35	37.78	19.64	19.06	0.06	0.33	< 0.001	< 0.001
FC_{PCTS} %	10.8–18.48	13.33	57.61	13.16	13.51	0.07	0.35	< 0.001	< 0.001
F_{Tan}%	9.16–19.64	13.69	76.55	13.40	13.98	0.09	0.42	< 0.001	< 0.001
FDA%	6.27–13.54	8.55	85.03	8.377	8.734	0.11	0.39	< 0.001	< 0.001
FDN%	9.02–19.35	13.05	79.16	12.61	13.50	0.09	0.41	< 0.001	< 0.001
HC%	2.67–6.48	4.56	83.55	4.513	4.615	0.19	0.14	< 0.001	< 0.001
Pol%	12.8–19.82	16.44	42.70	16.79	16.09	0.08	0.34	< 0.001	< 0.001
$PolC_{PCTS}$ %	10.47–16.45	13.59	44.00	13.89	13.29	0.08	0.32	< 0.001	< 0.001
$PolC_{Tan}$%	10.28–15.94	13.50	41.93	13.84	13.16	0.08	0.33	< 0.001	< 0.001
Pza%	75.41–89.95	84.82	17.14	85.36	84.27	0.04	0.18	< 0.001	< 0.001
U%	64.31–76.07	70.35	16.72	70.23	70.47	0.02	0.34	< 0.001	< 0.001

[a] Avg: average
[b] Early Avg: average of early maturing cultivar
[c] Late Avg: average of late maturing cultivar
[d] E: environmental fixed effect
[e] ExY: environment by harvest interaction

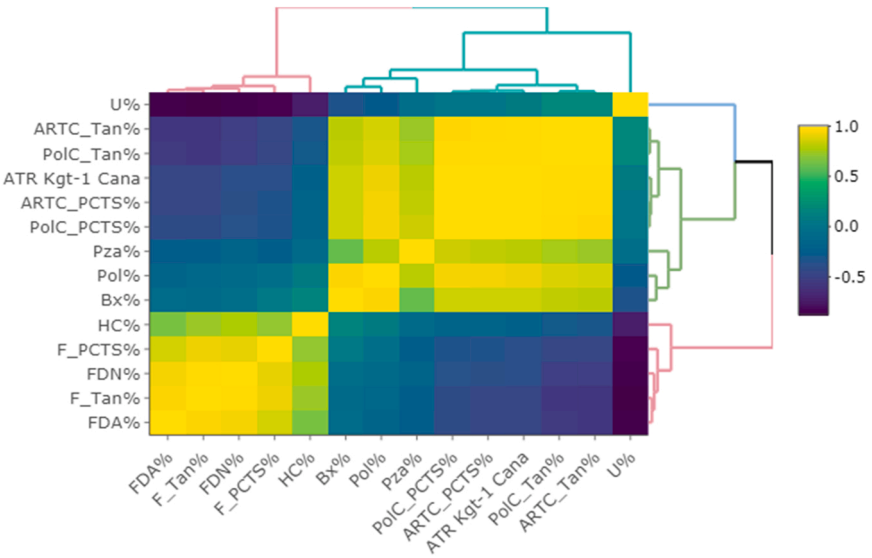

Fig. 2. Heatmap of the hierarchical clustering. The branch length of the dendrogram corresponds to the degree of similarity between the composition traits. The colors of the heatmap represent the color code of Pearson's correlated coefficients (see the histogram).

considering the six environments (Conchal, Jaboticabal, Taciba, Rolândia, Montividiu and Carpina) and two harvests (2015 and 2016) (Supplementary Table 3 - Material C). Then, the pairwise correlations were calculated and illustrated in Fig. 2 and Supplementary Table3 - Material D. Overall, significant genotypic correlations ($P < 0.05$) occurred between the 14 evaluated nutritional composition traits.

Fig. 2 indicates that the nutritional composition traits were clustered into three major groups: a) a group of traits associated with fiber (FDN %, FDA%, F_{PCTS}%, F_{Tan}%, and HC%); b) a group of traits associated with sugar (Bx%, Pol%, Pza%, $PolC_{PCTS}$%, $PolC_{Tan}$%, $ARTC_{PCTS}$%, $ARTC_{Tan}$%, and ATR Kgt^{-1} Cane); and c) a group with only the moisture content (U %) trait.

For the fiber group, the correlation values were all positive, and the values varied from moderate to high. If one considers the traits F_{Tan}%, F_{PCTS}%, FDA%, and FDN%, the correlation varied from 0.86 (FDA% - F_{PCTS}%) to 0.98 (FDN% - F_{Tan}%). Moderate values of correlation were observed when these values were associated with HC%, i.e., from 0.66 (HC% - FDA%) to 0.78 (HC% - FDN%).

Considering the group of traits associated with sugar, the correlations between variables were all positive. The lowest correlation values were obtained when Pza% was involved (0.61 – Pza%-Bx%, 0.74 Pza %-$ARTC_{Tan}$%, 0.77 Pza%-$PolC_{Tan}$%). For the other traits, two patterns were observed: i) correlations with values from 0.90 to 1.0, i.e., Bx% – Pol%; Pol% – $PolC_{PCTS}$%; $PolC_{PCTS}$% – ART%; $PolC_{PCTS}$% – $ARTC_{PCTS}$%; $PolC_{PCTS}$% – $ARTC_{Tan}$%; $PolC_{PCTS}$% – $PolC_{Tan}$%; Pol% – $ARTC_{PCTS}$%; Pol % – $ARTC_{PCTS}$%; Pol% – $ARTC_{Tan}$%; and Pol% – $PolC_{Tan}$%; and ii) correlation values that varied from 0.80 to 0.89, i.e., Bx% – $PolC_{PCTS}$%; Bx% – $PolC_{Tan}$%; Bx% – ART%; Bx% – $ARTC_{PCTS}$%; and Bx% – $ARTC_{Tan}$%. Only three pairwise correlation values between the sugar and fiber trait groups were positive and weak (0.08 - Pol%-HC%, 0.09 - Bx%-F_{PCTS}%, and 0.17 - Bx%-HC%); the others were negative and varied from − 0.02 (Pol% - F_{PCTS}%) to − 0.59 ($ARTC_{Tan}$-F_{Tan}%). In the present work, the strong genotypic correlation between the traits associated with sugar (Bx% – $PolC_{PCTS}$%; Bx% – Pol% and $PolC_{PCTS}$% – Pol%) (Fig. 2) shows that the selection practiced by breeding programs aims to increase the amount of sugar, considering that the parents of the commercial varieties evaluated in this study have different genetic backgrounds. When moisture content (U%) was considered with the fiber group, negative correlations with moderate to high magnitude, i.e., from − 0.73 (U%-HC %), − 0.88 (U%-F_{Tan}%), were observed; when U% was considered with the sugar group, negative values were observed for Bx% and Pol% (−0.35 and −0.28, respectively), positive values were observed for $PolC_{Tan}$% and $ARTC_{Tan}$% (0.18 and 0.20, respectively), and others varied from − 0.03 (U%-Pza%) and 0.07 (U%-ATR).

Considering that the cultivars used here can be clustered according to their maturity period, the next step was to verify whether these values would be similar if we recalculated the correlations for each of the maturing groups. Here, the cultivars were split into early- and late-maturing cultivars (8 and 12, respectively), and the correlation values were re-estimated (Supplementary Table 3- Material E) and compared with values considering all the cultivars. Based on 91 correlations between centesimal components for the overall mean and early- and late-maturing periods, a (general) correlation was calculated between the overall mean and early-maturing period cultivars (cor = 0.98), overall mean and late-maturing period cultivars (cor = 0.81), and early- and late-maturing period cultivars (cor=0.71). These results indicate that the correlation obtained here (Fig. 2) is stable when different maturation periods are considered.

Based on a similar approach, inferences about the correlations were made when the environmental conditions were changed. Here, all pairwise comparisons were also calculated between traits for each location and compared them against the overall values. Again, a general correlation was calculated to summarize the data based on the comparison between the correlation table of each location (Supplementary Table 3-Material F) and the overall correlations (Supplementary Table 3-Material D). This comparison provided a general correlation of 0.96. The highest values were obtained for Rolândia/PR (0.98) and Taciba/SP (0.98), and for Carpina/PE, Montividiu/GO, and Conchal/SP, the correlations were 0.95, 0.92 and 0.76, respectively. If this approach was used to compare between locations, the lowest value was found between 3 (Rolândia) and 6 (Carpina), with cor = 0.68; all the other values were above 0.80.

Also, the adjusted means were used to cluster the environments. First, the Tucker3 method with four (genotypes), three (location), and three (traits) components captured 77% of all dataset variation. Regarding the variability by components: the genotype captured 32.81% (g1), 19.67% (g2), 13.38% (g3), and 11.13% (g4); the location captured 50% (q1), 15.85% (q2), and 11.15% (q3); the traits captured 47.54% (r1), 28.70% (r2), and 0.75% (r3). The q1 component q1 represents an average between locations; q2 represents a comparison between L2 (Conchal) and L5 (Montevidiu) versus L3 (Rolandia) and L6 (Carpina); q3 represents L2 (Conchal) and L3 (Rolandia) versus L5 (Montevidiu) and L6 (Carpina). For the trait components, one verifies that r1 majorly represents the sugar traits (Bx%, Pol%, $PolC_{PCTS}$%,

$PolC_{Tan}$%, $ARTC_{PCTS}$%, $ARTC_{Tan}$%, ATR Kgt-1 Cana), r2 major represents the Fiber traits (F_{Tan}%, F_{PCTS}%, HC%, FDA%, FDN%), and r3 represents Pza%, U%, and HC% (Supplementary Table 3 - Material G).

The r1 and r2 parameters were used to generate two linear combinations used for the GE study (Fig. 3). For the first linear combination, there were two major clusters: the one for the northern locations (Montividiu/GO and Carpina/PE) and another for the southern locations (Jaboticabal/SP, Conchal/SP, Rolândia/PR, Taciba/SP). For the second linear combination, two other groups were indicated: one composed of only Carpina/PE and another with the five locations. The last one could be split into two subgroups: i) Montividiu/GO, Jaboticabal/SP; Conchal/SP; ii) Taciba/SP and Rolândia/PR.

4. Discussion

Sugarcane (a complex hybrid of *Saccharum* spp.) is an important worldwide crop in which the understanding of sugarcane's nutritional and technological composition is essential for agriculture, industry, nutrition, food trade, and biosafety assessment of GM (McMillan, 1997; Harrison, 2016; Mutton et al., 2020; Rey et al., 2021). However, this information is not fully detailed in the literature (Giuntini et al., 2006; Bressan et al., 2020). This study presented a database of 20 sugarcane nutritional composition traits evaluated for the main Brazilian cultivated genotypes in six production environments over two harvests. Here, the raw data contains 14,300 experimental data points (Supplementary Table 2), and the statistical analysis provided genetic predictions and insights into the genetic variability of this sample (Tables 1 and 2; Supplementary Table 3 - Material B and C). We believe these results can be used as a reference for future studies involving substantial equivalence.

For the GM regulatory process, for example, the critical factors for the database interpretation are the genetic and GE interaction effects. The ladder is an essential aspect that influences the relative performance of genotypes (Jackson et al., 1991; Bajpai and Kumar, 2005). Several methodologies are available to study GE interaction (Kang, 2002). The first approach used in this study assumed a statistical model that considers the same assumptions as the classical analysis of variance (Steel and Torrie, 1980). Seven nutritional traits (AR%, ARC_{PCTS}%, ARC_{Tan}%, Lp%, Ash%, and Pt%) did not present genetic variation in this step. The absence of genetic variation could reflect a high demanding standard of the industry, in which the new cultivars do not genetically vary for these traits. On the other hand, if these cultivars are considered for germplasm bank inclusion, strategies for broadening the genetic basis should be considered.

Therefore, linear mixed models or LMM (Henderson, 1984) have advantages such as flexibility for modeling different structures of variance-covariance that can indicate how genetic and nongenetic variability can vary over the experiment (Balzarini, 2002; Gouy et al., 2013; Pastina et al., 2012; Siraree et al., 2017; Hoarau et al., 2021). For the remaining traits, GE interaction was verified due to the change of genetic variability (Bernardo, 2010); for example, Bx% showed a genetic variance varying from 0.17 (Plant cane in Taciba – location 4) to 1.17 (first ratoon in Conchal – location 2). Consequently, the cultivar predictions will be more precise (Supplementary Table 3 - Material B), and the values and ranges can provide insights. Considering the varietal profiles of conventional sugarcane (except energy cane), the mean values for the traits Bx%, Pol%, $PolC_{PCTS}$% and $PolC_{Tan}$%, and fiber (F_{Tan}% and F_{PCTS}%) were maintained independently of the genetic background (pedigree) or harvest time (early or late varieties) (Table 2). Generally, the means obtained in this study are similar to values observed in the literature (Balsalobre et al., 2016; Barreto et al., 2021; Menandro et al., 2017; Todd et al., 2018). However, the range of values (Table 2) suggests that the GE interaction substantially influences cultivar performance.

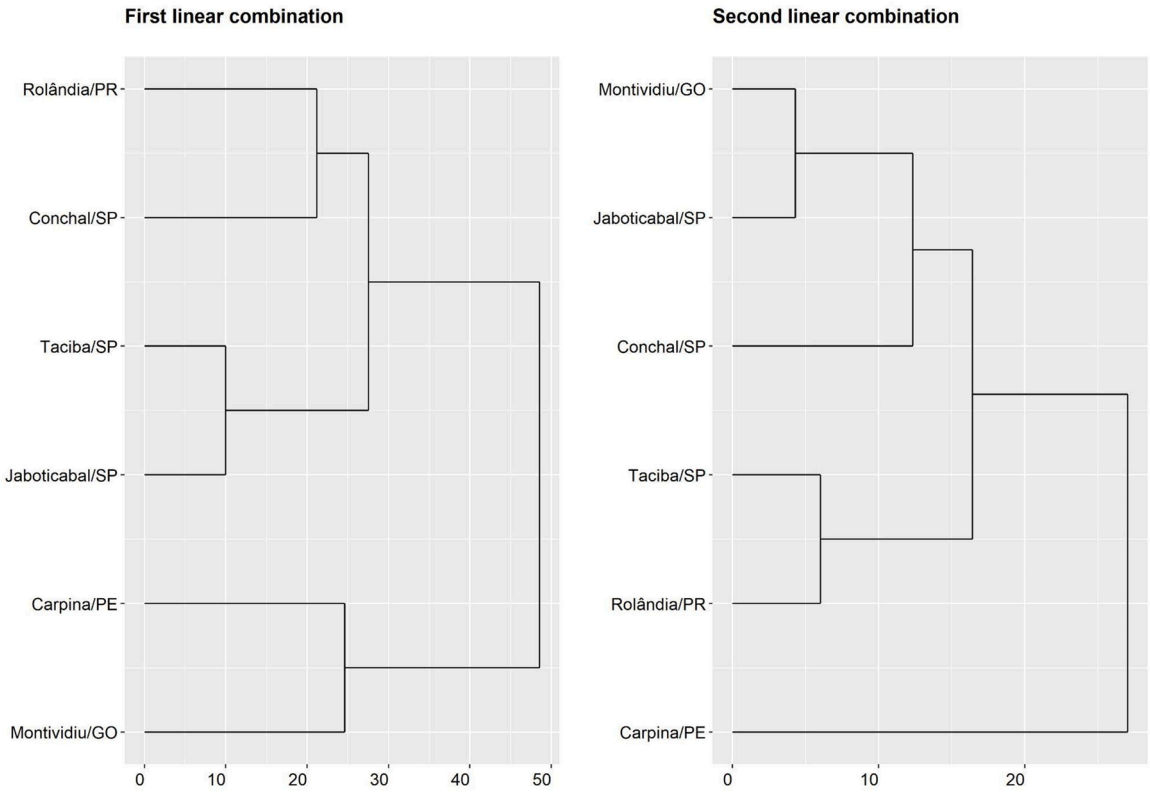

Fig. 3. Cluster analysis of the six sugarcane-producing environments where the experiment was conducted (1- Jaboticabal-SP, 2- Conchal-SP, 3- Rolândia-PR, 4- Taciba-SP, 5- Montividiu-GO and 6-Carpina-PE). The dendrogram was based on the unweighted pair group method with arithmetic means (UPGMA), considering 20 sugarcane cultivars. As 14 evaluated traits were available, the Tucker3 approach was used to obtain two linear combinations among traits.

The harvesting period in the Brazilian Central-South region, which is responsible for 90% of sugarcane production, is approximately from April to November. The sugarcane cultivars vary in maturation time and can be classified into early- and medium/late-maturing groups. Here, the cultivars were harvested according to their maturation times, allowing them to express their potential in relation to the environmental conditions (Cursi et al., 2021; de Morais et al., 2015). Except for U%, FDA%, and HC%, all the traits were significant for maturation time under the Wald test, e.g., ATR showed a difference of 5.4 Kgt^{-1} cane (150.2 for early- vs. 144.8 for late-maturing varieties), indicating that correct management at harvest is essential for maximizing sugar production and consequently industry profits. However, the minimum and maximum of ATR were similar for both maturation groups (early: 117.8–175.6; late: 121.0–175.0) (Supplementary Table 3 - Material C), indicating that the cultivar means differed between the early- and late-maturing groups, but the ranging limits were similar.

Two levels of environmental variation were studied: the microenvironment (residual) and the macroenvironment (combinations of location and crop year). The first level can be verified with CV_R, and the results of this work indicated good experimental accuracy in the field trials and good reliability of the data, i.e., from all the CV_R, only four values were higher than 0.20 (all of them for HC%) (Supplementary Table 3 - Material B). An efficient experimental control combined with a statistical approach that can integrate data at different locations and from multiple harvests can provide higher heritability estimates (Sadras et al., 2013; Schmidt et al., 2019); this pattern was observed when heritability estimates were compared between approaches (Table 1 versus Table 2 and Supplementary Table 3 - Material B). Assuming higher heritability values are associated with higher accuracy, insights using the mean predictions with LMM are preferable to the raw data. Also, the fiber-associated traits (F_{PCTS}%, F_{Tan}%, FDN% and FDA%) had higher values when compared to sugar-associated traits (Bx%, Pol%, $PolC_{PCTS}$, $PolC_{Tan}$%, $ARTC_{Tan}$ and ATR). This result reflects the effect of breeding efforts selecting sugar-related traits instead of fiber-related traits, i.e., Brazilian breeding programs focused on the increase in sugar as a priority (Cursi et al., 2021), and fiber-related traits were neglected by selection, resulting in this cultivar sample having more genetic variability available for fiber-related than for sugar-related traits.

Differences between production areas (macroenvironmental conditions) also influence the phenotypic values helping to increase the data range (Balsalobre et al., 2016). In this study, the (macro)environmental effect was significant for all the traits, using the Wald test (Table 2), indicating the environment changed the nutritional composition traits. These changes can be attributed to several factors, such as climate, soil fertility, and agronomic practices (irrigation and fertilization). In Brazil, one criterion for classification of the sugarcane production area is based on the expected production of tons of cane per hectare, considering mainly the soil properties (physical and chemical) and water management (natural or supplemented by irrigation) (Prado et al., 2008); this classification is present in the Supplementary Table 4 for the six environments. However, the same environmental classification may not be applied to the nutritional composition traits.

In this context, the cluster analysis was used to identify similarities between environments. Here, 14 traits were available, and a priori, all of them are equally important. In this case, a multivariate approach was applied to obtain two linear combinations for the traits that maximize the data variability (Supplementary Table 3 - Material G) for all the 20 genotypes in the six environments. Then clustering was applied for each linear combination (Fig. 3). The first one represents the sugar-related traits and clustered the environments in two clusters, the northern and southern locations. The second linear combination that represents fiber-related traits grouped the environments in three clusters according to the Koppen climate classification, i.e., Tropical zone with dry summer (As) composed by Carpina; Tropical or subtropical climate with dry winter composed by Montividiu (Aw), Jaboticabal (Cwa), and Conchal (Cwa); Humid subtropical zone, with fully humid and hot summer (Cfa) composed by Taciba and Rolândia. For future studies, we suggest that different climate (Koppen) classification and latitude coordinates were considered to investigate extreme contrasts for discriminating the cultivars for sugar or fiber traits.

The relationship between variables that can be investigated with the pairwise correlation index. Correlations between traits may reflect biological processes that are of considerable evolutionary interest and are the result of genetic, functional, physiological or developmental factors (Jamoza et al., 2014; Soomro et al., 2006). In Fig. 2, the genotypic correlations were calculated over six locations (Conchal, Jaboticabal, Taciba, Rolândia, Montividiu, and Carpina), two harvests (2015 and 2016), and 20 cultivars (early- and late-maturing time), but these associations varied when the factors changed. Since the environmental conditions were variable, both the genetic variability and the predicted means varied, but the association between traits did not tend to vary significantly. This result was also verified when early- and late-cultivars were compared. The simplest explanation for these correlations may involve the process of analysis used to determine the nutritional composition of sugarcane. For example, the variable POL (Juice) % was calculated as a function of the Brix value, which generates an association between these traits (Lopes et al., 2011).

On the other hand, the chemical composition of sugarcane is highly variable and depends on climatic conditions, physical, chemical and microbiological properties of the soil, type of cultivation, variety, stage of maturation and age, among other factors. Generally, sugarcane consists of fiber and juice. Fiber, defined as the set of substances insoluble in water, is mainly composed of cellulose, lignin and pentosans. Its content depends on variety, age and many other factors and can range from 10% to 16%. Broth, defined as an impure and dilute sucrose solution, consists of water (80%) and soluble solids (20%). Soluble solids (Brix) are grouped into organic and inorganic sugars and no sugars. Sugars are mainly represented by sucrose, glucose and fructose. Sucrose, the most important component, has an average value of 17%, while fructose and glucose have average values of 0.2% and 0.4%, respectively (Kim and Day, 2011; Lavanholi, 2010).

In the present study, considering the dendrogram in Fig. 2, three large subgroups of correlated nutritional composition characteristics were observed: i) fiber-related traits (FDA%, F_{Tan}%, FDN%, F_{PCTS}%, and HC%); ii) sugar-related traits (Bx%, Pol%, Pza%, $PolC_{PCTS}$%, $ARTC_{PCTS}$, ATR, $PolC_{Tan}$%, and $ARTC_{Tan}$); and iii) moisture (U%). These groups are concordant with the ones obtained with the Tucker3 method, i.e., the two linear combinations captured sugar and fiber variability. The quality of the raw sugarcane material can be affected by two sets of factors: a) intrinsic factors related to the composition of the cane (sucrose content, reducing sugars, fibers, phenolic compounds, starch, aconitic acid and minerals), which are affected according to the variety of sugarcane, climate variations (temperature, relative humidity, and rain), soil and cultural treatments; and b) extrinsic factors related to materials foreign to the stalk (earth, stone, crop residues, and invasive plants) or compounds produced by microorganisms due to their action on sugars in the stalk (Ripoli and Ripoli, 2004). In this case, the water content of the sugarcane stalk is an intrinsic factor that can vary very quickly in response to environmental aspects, in particular, low soil moisture, due to the occurrence of rain or drought (Dinardo-Miranda et al., 2008). As sugars (especially sucrose) present in sugarcane make up the water-soluble fraction, it was expected that there would be a positive and low correlation between the components associated with sugar and moisture (Fig. 2). Thus, it is natural to group moisture in the subgroup of associated sugar traits. However, a negative correlation was observed between moisture and both Pol% and Bx%. This behavior is expected since the analyses of these two variables directly considered sugarcane stalk juice, which is mostly composed of water. The variation in soil moisture in different environments at the time of sugarcane harvesting directly influences the estimated values of Pol% and Bx% (Dinardo-Miranda et al., 2008). Thus, according to the correlation

values obtained in the present work, it can be concluded that Bx% and POL% were influenced by the moisture present in the sugarcane stalk. Additionally, Pza% is an interesting trait because although it is not a measure that can be directly selected in sugarcane breeding, it is nevertheless an important variable for industrial decisions in sugar mills regarding the quality of the technological composition of raw sugarcane and can be associated with all sugar-related traits.

The present work showed a high genotypic correlation between traits associated with sugar (specially between Pol%-Bx%, Pol%-$PolC_{PCTS}$%, Pol%-$ARTC_{PCTS}$%, Pol%-ATR) (Fig. 2). Considering the main characteristics of the nutritional composition associated with sugar (Bx%, PolC %, and Pol%) throughout the selection period of a breeding program, the cultivars can satisfy the expectations of high sucrose accumulation. The evaluated cultivars came from different breeding programs and genetic backgrounds, despite that, all programs were equally efficient in producing cultivars with high sugar levels in the selection process.

Finally, the understanding of the sugarcane nutritional components was not fully detailed until now, but we hope the generated database as well the insights can provide enough information for future studies, especially for GM regulatory processes, and for breeders in terms of GE interaction aspects.

5. Conclusions

The centesimal composition nutrients are essential for fully characterizing a commercial variety. However, they are neglected over the breeding selection process. Consequently, there is not enough genetic characterization of these traits. The present study contributes significantly to the sugarcane community, providing more than 14,000 novel data points and insights about the genetic architecture of nutritional components and their GE interactions. We estimated genetic parameters, GE interaction, and genetic correlations between centesimal components traits based on twenty nutritional components evaluated for 20 commercial varieties in six locations and two years. We verified that AR %, ARC_{PCTS}%, ARC_{Tan}%, Lp%, Ash%, and Pt% did not show genetic variance, probably, due to the sugar industry standards that led to the breeding achievements. For the other traits (Bx%, Pol%, $PolC_{PCTS}$%, $PolC_{Tan}$%, Pza%, F_{PCTS}%, F_{Tan} %, $ARTC_{PCTS}$%, $ARTC_{Tan}$%, ATR (Kgt^{-1}Cane), U%, FDN%, FDA%, and HC%), genetic variability was available, as well the GE interaction.

The proper GE modeling through mixed models approach capitalized the genetic estimates. We modeled the GE interaction considering heteroscedasticity and covariances (unstructured models). The correlations between traits indicated three groups: one for the sugar-related traits, the other for fiber-related traits, and moisture as an intermediate one. The correlations were stable for locations, i.e., GE interaction did not affect the associations between centesimal components. To understand the similarity between locations, it first used the Tucker3 method to find linear combinations that maximize the dataset variability; then it also used the clustering algorithm. It was noticed that some locations showed high similarity due to their geographical and climate conditions. It is suggested that, if repeated, some locations can be discarded to optimize resources or replaced with new locations.

Funding

This research was co-funded by Fundação de Amparo à Pesquisa de São Paulo - FAPESP (2013/50195-2) and CropLife Brazil (at the time AgroBio - Associação das Empresas de Biotecnologia na Agricultura e Agroindústria).

Role of the funding source

We declare that the sponsors were not involved in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

All relevant data are within the paper and its Supplementary Materials.

Acknowledgements

We gratefully acknowledge CropLife Brazil associates (at the time AgroBio) for structuring and supporting the project.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fcr.2022.108678.

References

Alvares, A.A., Stape, J.L., Sentelhas, P.C., Gonçalves, J.L.M., Sparovek, G., 2014. Koppen's climate classification map for Brazil. Meteorol. Z. 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507.

Azevedo, J.A.G., Pereira, J.C., de Queiroz, A.C., Carneiro, P.C.S., de Paula Lana, R., Barbosa, M.H.P., Fernandes, A.M., Rennó, F.P., 2003. Composição químico-bromatológica, fracionamento de carboidratos e cinética da degradação in vitro da fibra de três variedades de cana-de-açúcar (*Saccharum spp.*). Rev. Bras. De. Zootec. 32, 1443–1453. https://doi.org/10.1590/S1516-35982003000600019.

Bajpai, P., Kumar, R., 2005. Comparison of methods for studying genotype x environment interaction in sugarcane. Sugar Tech. 7, 129–135. https://doi.org/10.1007/BF02950597.

Balsalobre, T.W.A., Mancini, M.C., Pereira, G.D.S., Anoni, C.O., Barreto, F.Z., Hoffmann, H.P., De Souza, A.P., Garcia, A.A.F., Carneiro, M.S., 2016. Mixed modeling of yield components and brown rust resistance in sugarcane families. Agron. J. 108, 1824–1837. https://doi.org/10.2134/agronj2015.0430.

Balzarini, M., 2002. Applications of mixed models in plant breeding. In: Kang, M. (Ed.), Quantitative Genetics, Genomics and Plant Breeding. CABI Publisher, New York, NY, pp. 353–363.

Barreto, F.Z., Balsalobre, T.W., Chapola, R.G., Garcia, A.A., Souza, A.P., Hoffmann, H.P., Gazaffi, R., Carneiro, M.S., 2021. Genetic variability, correlation among agronomic traits, and genetic progress in a sugarcane diversity panel. Agriculture 11, 533. https://doi.org/10.3390/agriculture11060533.

Bernardo, R.N., 2010. Breeding for Quantitative Traits in Plants, second ed.,. Stemma Press, Woodbury, MN.

Braga Jr., R., Landell, M., Da Silva, D., Bidoia, M., Da Silva, T., Da Silva, V., Luz, A., Dos Anjos, I., 2021. Censo varietal IAC: de cana-de-açúcar no Brasil - Safra 2018/19 e na região Centro-Sul - Safra 2019/20. Instituto Agronômico (IAC), Campinas, Brazil.

Bressan, E.A., De Carvalho, I.A.S., Borges, M., Carneiro, M.S., Da Silva, E.F., Gazaffi, R., Shirasuna, R.T., Abreu, V., Popin, R.V., Figueira, A., Oliveira, G.C.X., 2020. Assessment of gene flow to wild relatives and nutritional composition of sugarcane in Brazil. Front. Bioeng. Biotechnol. 8, 598. https://doi.org/10.3389/fbioe.2020.00598.

Cheavegatti-Gianotto, A., De Abreu, H.M., Arruda, P., Bespalhok Filho, J.C., Burnquist, W.L., Creste, S., Di Ciero, L., Ferro, J.A., De Oliveira Figueira, A.V., De Sousa Filgueiras, T., Grossi-de-Sá, M.D., Guzzo, E.C., Hoffmann, H.P., De Andrade Landell, M.G., Macedo, N., Matsuoka, S., De Castro Reinach, F., Romano, E., Da Silva, W.J., De Castro Silva Filho, M., César Ulian, E., 2011. Sugarcane (Saccharum X officinarum): a reference study for the regulation of genetically modified cultivars in Brazil. Trop. Plant Biol. 4, 62–89. https://doi.org/10.1007/s12042-011-9068-3.

Cursi, D.E., Hoffmann, H.P., Barbosa, G.V.S., Bressiani, J.A., Gazaffi, R., Chapola, R.G., Fernandes Jr., A.R., Balsalobre, T.W.A., Diniz, C.A., Santos, J.M., Carneiro, M.S., 2021. History and current status of sugarcane breeding, germplasm development and molecular genetics in Brazil. Sugar Tech. https://doi.org/10.1007/s12355-021-00951-1.

Dal-Bianco, M., Carneiro, M.S., Hotta, C.T., Chapola, R.G., Hoffmann, H.P., Garcia, A.A. F., Souza, G.M., 2012. Sugarcane improvement: how far can we go? (https://doi). Curr. Opin. Biotechnol. 23, 265–270. https://doi.org/10.1016/j.copbio.2011.09.002.

D'Hont, A., 2005. Unravelling the genome structure of polyploids using FISH and GISH: examples in sugarcane and banana. Cytogenet. Genome Res. 109, 27–33. https://doi.org/10.1159/000082378.

Dinardo-Miranda, L.L., de Vasconcellos, A.C.M., Landell, M.G.A., 2008. Cana-de-açúcar (Ed.). Instituto Agronômico, Campinas, p. 882. https://doi.org/10.3389/fpls.2018.00616 (Ed.).

Furlan, F.F., Tonon Filho, R., Pinto, F.H., Costa, C.B.B., Cruz, A.J.G., Giordano, R.L.C., Giordano, R.C., 2013. Bioelectricity versus bioethanol from sugarcane bagasse: is it

worth being flexible? Biotechnol. Biofuels 6, 1–12. https://doi.org/10.1186/1754-6834-6-142.
Gianotto, A.C., Rocha, M.S., Cutri, L., Lopes, F.C., Dal'Acqua, W., Hjelle, J.J., Lirette, R. P., Oliveira, W.S., Sereno, M.L., 2019. The insect-protected CTC91087-6 sugarcane event expresses Cry1Ac protein preferentially in leaves and presents compositional equivalence to conventional sugarcane. GM Crops Food 10, 208–219. https://doi.org/10.1080/21645698.2019.1651191.
Giuntini, E.B., Lajolo, F.M., de Menezes, E.W., 2006. Composição de alimentos: um pouco de história. Arch. Latinoam. Nutr. 56, 295–303.
Gouy, M., Rousselle, Y., Bastianelli, D., Lecomte, P., Bonnal, L., Roques, D., Efile, J.C., Rocher, S., Daugrois, J., Toubi, L., Nabeneza, S., Hervouet, C., Telismart, H., Denis, M., Thong-Chane, A., Glaszmann, J.C., Hoarau, J.Y., Nibouche, S., Costet, L., 2013. Experimental assessment of the accuracy of genomic selection in sugarcane. Theor. Appl. Genet. 126, 2575–2586. https://doi.org/10.1007/s00122-013-2156-z.
Gouy, M., Luquet, D., Rouan, L., Martiné, J.F., Thong-Chane, A., Costet, L., Nibouche, S., Gozé, E., 2015. Site and *Saccharum spontaneum* introgression level drive sugarcane yield component traits and their impact on sucrose yield in contrasted radiation and thermal conditions in La Réunion. Field Crops Res 171, 99–108. https://doi.org/10.1016/j.fcr.2014.11.002.
Harrison, M., 2016. Sugarcane-derived animal feed. In: O'Hara, I.M., Mundree, S.G. (Eds.), Sugarcane-Based Biofuels and Bioproducts. John Wiley & Sons, Hoboken, NJ, pp. 281–310.
Henderson, C., 1984. Applicatıons of lınear models ın animal breedıng. Univ. Guelph Press Guelph 11, 652–653.
Hiloidhari, M., Araújo, K., Kumari, S., Baruah, D.C., Ramachandra, T.V., Kataki, R., Thakur, I.S., 2018. Bioelectricity from sugarcane bagasse co-generation in India - an assessment of resource potential, policies and market mobilization opportunities for the case of Uttar Pradesh. J. Clean. Prod. 182, 1012–1023. https://doi.org/10.1016/j.jclepro.2018.02.087.
Hoang, N.V., Furtado, A., Botha, F.C., Simmons, B.A., Henry, R.J., 2015. Potential for genetic improvement of sugarcane as a source of biomass for biofuels. Front. Bioeng. Biotech. 3, 1–15. https://doi.org/10.3389/fbioe.2015.00182.
Hoarau, J.Y., Dumont, T., Wei, X., Jackson, P., D'Hont, A., 2021. Applications of quantitative genetics and statistical analyses in sugarcane breeding. Sugar Tech. https://doi.org/10.1007/s12355-021-01012-3.
Jackson, P.A., 2005. Breeding for improved sugar content in sugarcane. Field Crops Res. 92, 277–290. https://doi.org/10.1016/j.fcr.2005.01.024.
Jackson, P.A., Horsley, D., Foreman, J., Hogarth, D.M., Wood, A.W., 1991. Genotype × environment (GE) interaction in sugarcane variety trials in the Herbert. Proc. Aust. Soc. Sugar Cane Technol. 13, 103–109.
Jamoza, J., Owuoche, J., Kiplagat, O., Opile, W., 2014. Broad-sense heritability estimation and correlation among sugarcane (Saccharum spp. hybrids) yield and some agronomic traits in western Kenya. Int. J. Agric. Policy Res 2, 16–25.
Kang, M.S., 2002. Genotype-environment interaction: progress and prospects. In: Kang, M.S. (Ed.), Quantitative Genetics, Genomics, and Plant Breeding. CABI Publisher, Wallingford, England, pp. 221–243.
Kennedy, R., Cheavegatti Gianotto, A., Oliveira, W., Lirette, R., Hjelle, J., 2018. A General Safety Assessment for Purified Food Ingredients Derived From Biotechnology Crops: Case Study of Brazilian Sugar and Beverages Produced From Insect-Protected Sugarcane. Front. Bioeng. Biotechnol. 6 https://doi.org/10.3389/fbioe.2018.00045.
Kim, M., Day, D.F., 2011. Composition of sugar cane, energy cane, and sweet sorghum suitable for ethanol production at Louisiana sugar mills. J. Ind. Microbiol. Biotechnol. 38, 803–807. https://doi.org/10.1007/s10295-010-0812-8.
Lavanholi, M.G.D.P., 2010. In: Dinardo-Miranda, L.L., Vasconcelos, A.C.M., Landell, M. G.A. (Eds.), Qualidade da cana-de-açúcar como matéria-prima para produção de açúcar e álcool. Instituto Agronômico, Campinas.
Lopes, C.H., Gabriel, A.V.M.D., Borges, M.T.M.R. 2011. Produção de Etanol a partir da cana-de-açúcar. São Carlos: EDUFSCAR. 129p.
McMillan, J.D., 1997. Bioethanol production: status and prospects. Renew. Energy 10, 295–302. https://doi.org/10.1016/0960-1481(96)00081-X.
Menandro, L.M.S., Cantarella, H., Franco, H.C.J., Kölln, O.T., Pimenta, M.T.B., Sanches, G.M., Rabelo, S.C., Carvalho, J.L.N., 2017. Comprehensive assessment of sugarcane straw: implications for biomass and bioenergy production. Biofuels Bioprod. Bioref. 11, 488–504. https://doi.org/10.1002/bbb.1760.
de Morais, L.K., De Aguiar, M.S., De Albuquerque e Silva, P., Câmara, T.M.M., Cursi, D. E., Júnior, A.R.F., Chapola, R.G., Carneiro, M.S., Bespalhok Filho, J.C., 2015. Breeding of sugarcane. In: Cruz, V.M.V., Dierig, D.A. (Eds.), Industrial Crops: Breeding for BioEnergy and Bioproducts. Springer, New York, NY, pp. 29–42.
Mutton, M.J.R., Garcia, G., Teixeira, V., Silva, A.F., Costa, G.H.G., Ferreira, O.E., 2020. The clarification of sugarcane juice and the use of CA-11 yeast produces better quality cachaça1. Rev. Ciênc. Agron. 51, e20197021 https://doi.org/10.5935/1806-6690.20200067.
OECD. Consensus document on compositional considerations for new varieties of sugarcane (Saccharum ssp. hybrids): key food and feed nutrients, anti-nutrients and toxicants. Series on the safety of novel foods and feeds no. 23. Paris, France: OECD Environment Dictorate; 2011. [accessed 2021Mar]. ⟨https://www.oecd.org/env/ehs/biotrack/48962816.pdf⟩.
Pastina, M.M., Malosetti, M., Gazaffi, R., Mollinari, M., Margarido, G.R., Oliveira, K.M., Pinto, L.R., Souza, A.P., Van Eeuwijk, F.A., Garcia, A.A., 2012. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. Theor. Appl. Genet. 124, 835–849. https://doi.org/10.1007/s00122-011-1748-8.
Payne, R.W., D.A. Murray, S.A. Harding, D.B. Baird, D.M. Soutar. 2009. GenStat for Windows. 12th ed. VSN Int., Hemel Hempstead, UK.
Piperidis, N., D'Hont, A., 2020. Sugarcane genome architecture decrypted with chromosome-specific oligo probes. Plant J. 103, 2039–2051. https://doi.org/10.1111/tpj.14881.
Prado, H., Junior, A.L.P., Garcia, J.C., Moraes, J.F.L. de, Carvalho, J.P. de, Donzeli, P.L. Solos e, 2008. ambientes de produção. In: Dinardo-Miranda, L.L., Vasconcelos, A.C. Mde, Landell, M.G. de A. (Eds.), Cana-de-açúcar. Instituto Agronômico, Campinas, pp. 179–204.
Rey, J.R.C., Pacheco, J.J.T., Tarelho, L.A.C., Silva, V., Cardoso, J.S., Silveira, J.L., Tuna, C.E., 2021. Evaluation of cogeneration alternative systems integrating biomass gasification applied to a Brazilian sugar industry. Renew. Energy 178, 318–333. https://doi.org/10.1016/j.renene.2021.06.053.
Ripoli, T.C.C., Ripoli, M.L.C., 2004. Biomassa de cana-de-açúcar: colheita, energia e ambiente. Barros & Marques, Piracicaba, p. 302 (Ed. Eletrônica).
Sadras, V.O., Rebetzke, G.J., Edmeades, G.O., 2013. The phenotype and the components of phenotypic variance of crop traits. Field Crops Res. 154, 255–259. https://doi.org/10.1016/j.fcr.2013.10.001.
Saini, J.K., Saini, R., Tewari, L., 2015. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: Concepts and recent developments. 3 Biotech 5, 337–353. https://doi.org/10.1007/s13205-014-0246-5.
Santos, R.V., Evangelista, A.R., Pinto, J.C., Couto-Filho, C.C.C., Souza, R.M., 2006. Composição química da cana-de-açúcar (*Saccharum* spp.) e das silagens com diferentes aditivos em duas idades de corte. Ciência e Agrotecnologia 30, 1184–1189. https://doi.org/10.1590/S1413-70542006000600022.
Schmidt, P., Hartung, J., Bennewitz, J., Piepho, H.P., 2019. Heritability in plant breeding on a genotype-difference basis. Genetics 212, 991–1008. https://doi.org/10.1534/genetics.119.302134.
Silveira, M.H.L., Vanelli, B.A., Chandel, A.K., 2018. Second Generation Ethanol Production: Potential Biomass Feedstock, Biomass Deconstruction, and Chemical Platforms for Process Valorization. In: Chandel, A.K., Silveira, M.H.L. (Eds.), Advances in Sugarcane Biorefinery Technologies, Commercialization, Policy Issues and Paradigm Shift for Bioethanol and By-Products. Elsevier, Cambridge, pp. 135–152. https://doi.org/10.1016/B978-0-12-804534-3.00006-9.
Siraree, A., Banerjee, N., Kumar, S., Khan, M.S., Singh, P.K., Kumar, S., Sharma, S., Singh, R.K., Singh, J., 2017. Identification of marker-trait associations for morphological descriptors and yield component traits in sugarcane. Physiol. Mol. Biol. Plants 23, 185–196. https://doi.org/10.1007/s12298-016-0403-x.
Soomro, A.F., Junejo, S., Ahmed, A., Aslam, M., 2006. Evaluation of different promising sugarcane varieties for some quantitative and qualitative attributes under Thatta (Pakistan) conditions. Int. J. Agric. Biol. 8, 195–197.
de Souza França, A.F., Queiroz, S.S.M., Rosa, B., Reyes, A. de los B., Pereira, S.M., Ferreira, Magalhães, M.R.M., Aparecido, T.R. de M., e Gonçalves, J. dos R., 2004. Avaliação do potencial produtivo e das características químico-bromatológicas de nove variedades de cana-de-açúcar irrigada. Livest. Res. Rural Dev. 17, 1. ⟨https://lrrd.cipav.org.co/lrrd17/1/souz17007.htm⟩.
Steel, R.G.D., Torrie, J.H., 1980. Principles and Procedures of Statistics. McGraw-Hill, New York.
Thirugnanasambandam, P.P., Hoang, N.V., Henry, R.J., 2018. The challenge of analyzing the sugarcane genome. Front. Plant Sci. 9, 616. https://doi.org/10.3389/fpls.2018.00616.
Todd, J.R., Sandhu, H., Binder, J., Arundale, R., Gordon, V., Song, J., Glaz, B., Wang, J., 2018. Fiber composition of a diversity panel of the world collection of sugarcane (Saccharum spp.) and related grasses. Bragantia 77, 48–61. https://doi.org/10.1590/1678-4499.2016525.
Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31, 279–311.
Yadav, S., Jackson, P., Wei, X., Ross, E.M., Aitken, K., Deomano, E., Atkin, F., Hayes, B.J., Voss-Fels, K.P., 2020. Accelerating genetic gain in sugarcane breeding using genomic selection. Agronomy 10 (4), 585. https://doi.org/10.3390/agronomy10040585.