PLOS ONE

RESEARCH ARTICLE

Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits

Mateus Teles Vital Gonçalves[1], **Gota Morota**[2], **Paulo Mafra de Almeida Costa**[3], **Pedro Marcus Pereira Vidigal**[4], **Marcio Henrique Pereira Barbosa**[5], **Luiz Alexandre Peternelli**[1]*

1 Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brazil, **2** Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States of America, **3** Instituto Federal Catarinense—Campus Concórdia, Concórdia, SC, Brazil, **4** Centro de Análises de Biomoléculas/NuBioMol, Universidade Federal de Viçosa, Viçosa, MG, Brazil, **5** Departamento de Fitotecnia, Universidade Federal de Viçosa, Viçosa, MG, Brazil

* peternelli@ufv.br

Citation: Gonçalves MTV, Morota G, Costa PMdA, Vidigal P, Barbosa MHP, Peternelli L (2021) Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. PLoS ONE 16(3): e0236853. https://doi.org/10.1371/journal.pone.0236853

Editor: Paulo Eduardo Teodoro, Federal University of Mato Grosso do Sul, BRAZIL

Received: July 13, 2020

Accepted: January 20, 2021

Published: March 4, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0236853

Data Availability Statement: The data underlying the results presented in the study are available from 10.6084/m9.figshare.12635717.

Abstract

The main objectives of this study were to evaluate the prediction performance of genomic and near-infrared spectroscopy (NIR) data and whether the integration of genomic and NIR predictor variables can increase the prediction accuracy of two feedstock quality traits (fiber and sucrose content) in a sugarcane population (*Saccharum* spp.). The following three modeling strategies were compared: M1 (genome-based prediction), M2 (NIR-based prediction), and M3 (integration of genomics and NIR wavenumbers). Data were collected from a commercial population comprised of three hundred and eighty-five individuals, genotyped for single nucleotide polymorphisms and screened using NIR spectroscopy. We compared partial least squares (PLS) and BayesB regression methods to estimate marker and wavenumber effects. In order to assess model performance, we employed random sub-sampling cross-validation to calculate the mean Pearson correlation coefficient between observed and predicted values. Our results showed that models fitted using BayesB were more predictive than PLS models. We found that NIR (M2) provided the highest prediction accuracy, whereas genomics (M1) presented the lowest predictive ability, regardless of the measured traits and regression methods used. The integration of predictors derived from NIR spectroscopy and genomics into a single model (M3) did not significantly improve the prediction accuracy for the two traits evaluated. These findings suggest that NIR-based prediction can be an effective strategy for predicting the genetic merit of sugarcane clones.

Introduction

The strides achieved with improved instruments, laboratory techniques, and bioinformatics tools have allowed the emergence of next-generation sequencing technologies [1]. These technologies can deliver DNA-level information at an ever more cost-effective and high-

Funding: MTVG received a masters degree scholarship (154611/2017-4) from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Grant Number 310503/2015-9 to LAP. We are also thankful for the Inter-University Network for the Development of Sugarcane Industry (RIDESA) for all the field experiment support. 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.

throughput manner, which has boosted the important role genomic prediction (GP) might play in plant breeding [2]. The idea of GP is to fit a regression model using phenotypic records and the entire set of molecular markers concurrently. The model developed enables the prediction of the genetic merit of genotyped but non-phenotyped populations [3].

The application of GP as a breeding strategy is envisioned to reduce costs while saving time and resources [4]. One example is the reduction of generation interval as genitors could be crossed, the resulting progeny have their DNA collected from seeds or juvenile tissues, and then genotyped to have their breeding values predicted, what may result in gains per unit of time [5]. Sugarcane breeding programs are a good case point, which are characterized by long breeding cycles of phenotypic selection, performed over years of evaluation trials across different environments [6]. Another advantage of GP is that it could eventually lead to a reduction in phenotyping costs, especially in situations where traits are difficult to measure and when there are thousands of candidate genotypes to evaluate [7]. Therefore, the adoption of GP over phenotypic selection is expected to augment selection efficiency and to accelerate cultivar release [7,8].

Nevertheless, the implementation of GP is highly dependent on accurate phenotyping records [9]. Moreover, conventional phenotyping is not excluded from GP-based plant breeding schemes as model updates on new training populations would still be necessary [10]. However, besides the forecasts of continuous advances and decreasing costs in molecular breeding, the phenotyping step remains a significant bottleneck [11]. Phenotyping routine traits at breeding stations are commonly performed manually over several crop years and across different environments, which is often a time-consuming, labor-intensive, and high-cost task [12]. Also, conventional phenotyping is prone to human error. Thus, the genetic potential of populations may not be fully exploited [11].

Research efforts to address these constraints are currently being conducted with the development of high-throughput phenotyping (HTP) systems. HTP is an incipient, though a growing area of interest among plant breeders [13]. These novel approaches include a multitude of sensors and imaging techniques mounted on ground-based or uncrewd aerial vehicles (UAV) that can collect phenotypes in a precise, automated, and large-scale fashion [14]. Thus, HTP tools have the potential to improve plant breeding program pipelines significantly [15]. Moreover, HTP technologies can replace standard less effective phenotyping protocols, thus saving much time and resources. For instance, near-infrared (NIR) spectroscopy technology has been successfully applied for many crops, including sugarcane, to screen biological sample compositions and also for breeding purposes [9,16–18]. The attractive features of NIR spectroscopy could potentially aid sugarcane breeders, with the increase of selection accuracy and reduction of costs when compared to conventional phenotyping [11].

The combination of HTP systems with improved genomic tools is heralded to increase genetic gains in plant breeding [11,12]. However, the startling amount of data being generated is outpacing our ability to explore it. Besides, how this information can properly be implemented is still unclear and needs further investigations [14,19]. The integration of HTP information and GP can be performed by the exploitation of HTP platforms to provide phenotypic records that can be either treated as secondary traits (e.g., vegetation indexes) and regressed on molecular markers, or as predictor variables together with molecular markers in a single- or multi-trait analysis [20]. For instance, Crain et al. (2018) [21] investigated different proposals to integrate HTP derived variables into GP models in wheat and found improved prediction accuracies. Other strategies provide modeling alternatives that include interaction effects [22,23].

The main goal of this study was to investigate the performance of the integration of HTP and genomic datasets aiming to increase the accuracy of prediction for two important

sugarcane feedstock quality traits, namely fiber (FIB) and sucrose (PC) content, in a commercial sugarcane (*Saccharum* spp.) population from the sugarcane genetic breeding program of the Universidade Federal de Viçosa (PMGCA-UFV). The population was genotyped for single nucleotide polymorphisms (SNPs) and screened using NIR spectroscopy. We compared three modelling strategies: 1) genome-based prediction, 2) NIR-based prediction, and 3) the integration of SNP markers and NIR wavenumber variables as predictors.

Material and methods

Plant material

In this study, we evaluated a population of 385 clones derived from an originally seedling population of 98 half-sib families. The seedling population, in which each plant is a single genotype, was the result of crosses made at the Serra do Ouro Flowering and Breeding Station, municipality of Murici, Alagoas State, Brazil (09˚13' S, 35˚50' W, 450 m altitude). After processing, seeds were sent to the Sugarcane Genetic Breeding Research Station (CECA) of the Universidade Federal de Viçosa, municipality of Oratórios, Minas Gerais State, Brazil (20˚25' S, 42˚48' W, 494 m altitude) and germinated in a nursery house. Subsequently, seedlings obtained from each family were transplanted to the field and evaluated in first (plant cane) and second (ratoon) crops based on desirable traits in first (T1) and second (T2) clonal trial stages of selection [6].

Experimental design

An augmented block design was initiated in May 2016 at the CECA municipality of Oratórios, Minas Gerais State, Brazil (20˚25' S, 42˚48' W, 494 m altitude). The released cultivars RB867515, RB966928, and RB92579 were included as checks once in each block, and regular unreplicated clones were arranged in 21 blocks [24]. The replicated checks are well-established cultivars, widely grown throughout the country and are often used as parents for crosses. The experimental plots consisted of double-row 3 m long furrows × 1.4 m between rows and clones were cultivated following standard agronomic protocols regarding fertilization, weed control, and pest management [25]. Buffer rows of released cultivars encompassed the whole experiment area.

Phenotypic data

The clones were evaluated in the first ratoon (second crop) 26 months after planting. The method employed to estimate the percentage of FIB content followed recommendations of the CONSECANA manual [26]. Sugarcane breeding programs in Brazil routinely apply these protocols. Harvest and quality analyses were performed in July 2018. To obtain a representative set of the samples, ten randomly selected stalks from each of the double-row plots were cut at ground level with a machete. Green tops, clinging leaves, and leaf sheaths were removed before stalks were bundled and weighted using a dynamometer (S1 Fig). These ten randomly selected stalks in each plot were shredded. A subsample of 500 g from the shredded stalks was collected and pressed with a hydraulic press. After pressing, the remainder fiber cake was collected and taken to the laboratory. We obtained the BRIX% (percentage of soluble solids) and POL% (percentage of sucrose) values from the juice. The BRIX was obtained with a refractometer (HI96801 Model, Hanna® instruments, Woonsocket, USA), while POL was obtained by polarimetry using a saccharimeter (SDA2500 Model, Acatec, Brazil) after clarifying the solution with lead acetate. The remainder fiber cake was weighed (*WC*) and used to derive fiber

content [27]:

$$FIB = 0.08 \times WC + 0.876 \quad (1)$$

The apparent percentage of sucrose in sugarcane (PC%) was derived based on POL% and FIB as follows:

$$PC = POL\% \times (1 - 0.01 \times FIB) \times C \quad (2)$$

where C is the coefficient to convert sucrose of juice into sucrose of cane calculated using the formula $C = 1.0313–0.00575 \times FIB$. The final values were all expressed as the total fresh biomass basis (500 g of shredded stalks).

Sample preparation and NIR spectra acquisition

Another subsample of 100 g from the shredded stalks was collected and immediately taken to dry in a forced-air circulating oven at 50˚C for 24 h or until a constant mass was reached (S1 Fig). Dried samples were then ground, packaged in a plastic zip bag, and stored. The NIR spectra of samples were measured in indoor-conditions at room temperature of 21˚C. The instrument used was a Fourier transform near-infrared (FT-NIR) spectrometer set (Antaris™ II Model, Thermo Scientific Inc., USA), under the following operating conditions: 4 cm^{-1} resolution in an investigated wavenumber range of 10000 to 4000 cm^{-1} and reflectance mode as log (1/R), where R is the measured reflectance. Samples were placed into a powder sampling cup accessory and arranged into the instrument window. At each scan, the accessory was moved to cover different positions of the sample, totaling six positions. A single scan measure was the average result of 32 scans. For each sample, a total of 192 scans were made and then averaged, representing the final spectrum. The final NIR matrix used in the subsequent analyses had a dimension of 385 rows and 3,112 columns.

DNA isolation, sequencing, and genotyping data

Sugarcane DNA samples were isolated using DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) and sent to RAPiD Genomics (Gainesville, Florida, USA) for the construction of probes, sequencing, and identification of molecular markers. Samples were genotyped using single-dose SNP markers based on the Capture-Seq technology (https://www.rapid-genomics.com). Raw sequence reads were mapped, called, and filtered. Reads were anchored to a monoploid reference genome of sugarcane (*Saccharum spp*.) [28] using the BWA-MEM algorithm of the BWA version 0.7.17 [29]. A flag identifying the respective sugarcane genotype was added to each mapping file. The mapping files were processed using SortSam, MarkDuplicates, and BuildBamIndex tools of Picard version 2.18.27 (https://github.com/broadinstitute/picard/). Variants were called using FreeBayes version 1.2.0 (https://github.com/ekg/freebayes) with a minimum mapping quality of 20 (probability of miscalling), minimum base quality of 20 (SNPs with missing data higher than 20% were eliminated), and minimum coverage (how many times a fragment was sequenced) of 20 reads at every position in the reference genome. Thereafter, the SNP marker matrix was coded counting the occurrence of the reference allele A. Thus, considering the genotypes AA, Aa, and aa, the matrix entries would be 2 (homozygosity for the reference allele), 1 (heterozygosity with one reference and one alternative allele), and 0 (homozygosity for the alternative allele), respectively. Further, markers with minor allele frequency lower than 5% were eliminated. Lastly, missing variants were imputed from a binomial distribution density function using the frequency of the non-missing variants. A total of 124,307 SNPs was retained for further analyses.

Statistical analysis

A two-stage analysis was employed. In the first step, we run a mixed model equation with variance components estimated by REML using the SELEGEN-REML/BLUP software [30]. We considered the model

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Xb} + \boldsymbol{Zg} + \boldsymbol{\varepsilon} \quad (3)$$

where **y** is the vector of phenotypes; **1** is a vector of 1s; μ is the overall mean; $\mathbf{b} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_b)$ is the vector of random block effects; $\boldsymbol{g} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_g)$ is the vector of random genetic effects, and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2_\varepsilon)$ is the vector of residuals. The effects of checks were taken as random. The capital letters **X** and **Z** represent the incidence matrices of the respective random effects. In the second step, the best linear unbiased predictors (BLUPs) of each trait were used as dependent variables using three prediction models (M1, M2, and M3) to evaluate the prediction accuracy for FIB and PC.

The first (M1) was a single-trait model using only markers which takes the following form:

$$\boldsymbol{y}^* = \mathbf{1}\mu + \boldsymbol{Wa} + \boldsymbol{\varepsilon} \quad (4)$$

where $\boldsymbol{y}^*$ is the vector of adjusted phenotypic values (BLUPs) for FIB or PC, **1** is a vector of 1s; μ is the overall mean, **W** is the matrix with SNP markers for each individual, ***a*** is the corresponding vector of marker effects, and **ε** is the vector of residuals.

The second (M2) was a single-trait model using NIR wavenumber variables as predictors:

$$\boldsymbol{y}^* = \mathbf{1}\mu + \boldsymbol{Ns} + \boldsymbol{\varepsilon} \quad (5)$$

where ***N*** is the matrix with the spectrum for each individual along the wavelength, and ***s*** is the corresponding vector of wavelength effects. It is a commonplace to apply mathematical transformations to the NIR matrix before analysis to increase the signal to noise ratio [31]. More details can be found elsewhere [32]. We tested different combinations of pre-processing techniques. The pre-processing combination that yielded the best results was Savitzky-Golay smoothing (SGS) (window: 5; polynomial order: 2) followed by multiplicative scatter correction (MSC) and mean centering (MC) for FIB, whereas for PC the best combination was SGS (window: 5; polynomial order: 2) and MC (Fig 1).

In the third model (M3), we combined SNP markers and NIR wavenumber variables as predictors fitting the following linear model:

$$\boldsymbol{y}^* = \mathbf{1}\mu + \boldsymbol{Wa} + \boldsymbol{Ns} + \boldsymbol{\varepsilon} \quad (6)$$

In Eq (6), the pre-processing combination of the NIR matrix that best contributed with the SNP matrix for maximizing the prediction accuracy were SGS (window: 5; polynomial order: 2), MSC and MC for FIB, and SGS (window: 5; polynomial order: 2), 1˚ derivative, MSC and MC for PC. The incidence matrices ***N*** and ***W*** were scaled (centered and standardized) in all prediction models prior to the analyses.

Regression models

The models were tested using two regression methods: The BayesB and partial least squares (PLS). BayesB is a hierarchical Bayesian approach that performs variable selection and shrinkage [3]. We used a multi-layer BayesB by assigning different independent priors for SNP markers and NIR wavenumbers in M3. PLS regression is a dimension reduction method and fundamentally transforms the original collinear predictors into non-correlated variables [33]. In PLS regression, the algorithm identifies the principal components (latent variables) that

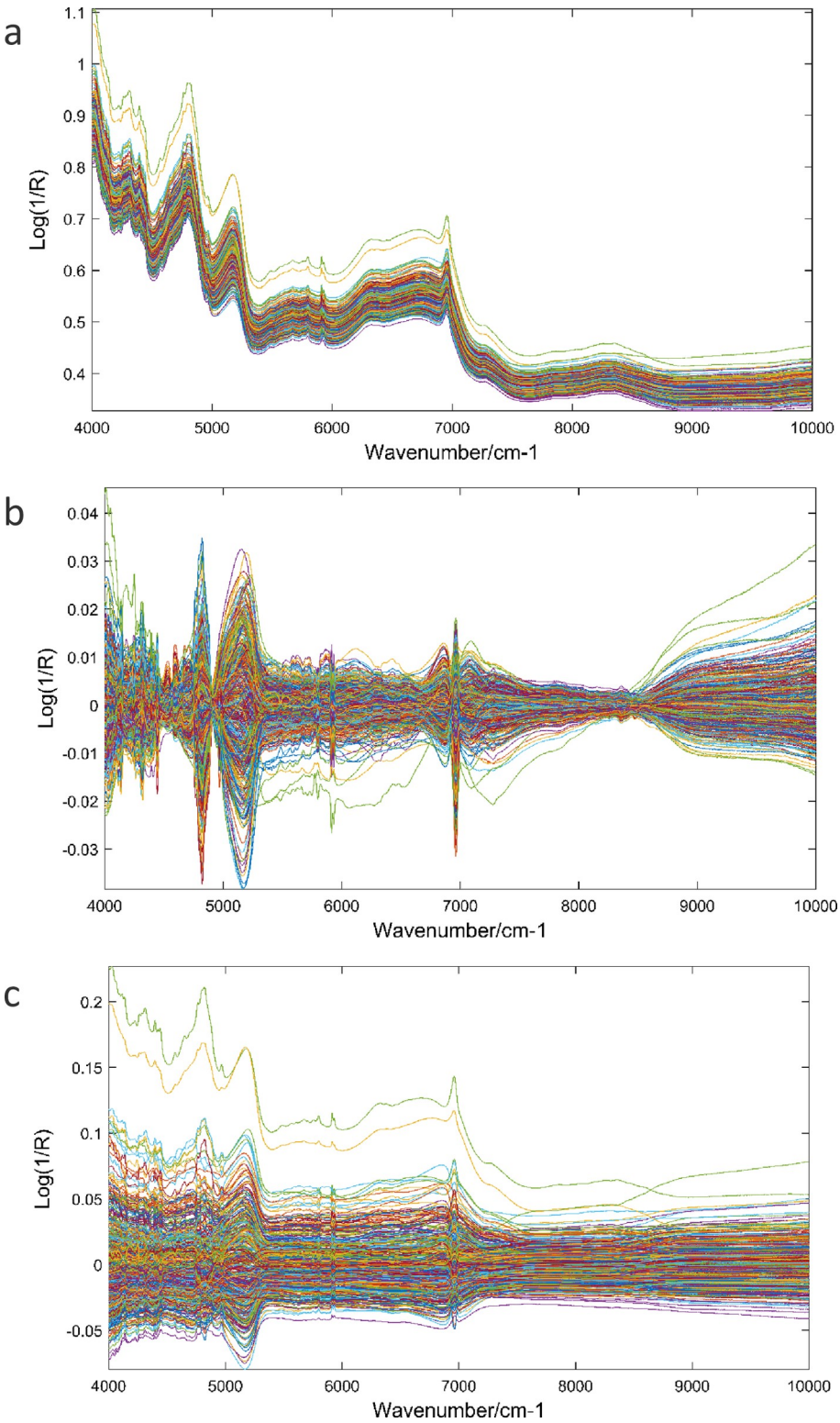

Fig 1. Sugarcane samples raw spectra (a) and pre-processed spectra for fiber content (b), and sucrose content (c).

https://doi.org/10.1371/journal.pone.0236853.g001

best describe the data in terms of variance, and it does so by constructing linear combinations of all predictors. Furthermore, unlike other dimension reduction models such as principal component regression, the fitting procedure of PLS involves finding the latent variables that maximize the covariance between the predictors and phenotypes while minimizing the error [34,35].

The BayesB analyses were carried out using the BGLR package [36]. We run the BayesB for 25,000 samples, with the first 10,000 being discarded (burn-in) with a thinning interval of 10. The PLS regression was performed using the mixOmics package [37]. Both methods were implemented in R [38].

Accuracy of predictions

The prediction accuracy of models M1, M2, and M3 was evaluated by random sub-sampling cross-validation repeated 20 times. The models are fitted using the data of the training set observations and tested to predict unknown samples of the validation set. In this study, the training set contained 80% of the samples (308 clones) and the validation set included the remainder of 20% (77 clones). At each time, the algorithm randomly selected a different subset of observations assigned to the training and validation sets. The results were compared by computing the mean Pearson correlation coefficient between observed and predicted values. A schematic diagram summarizing the whole experimental preparation and processing is depicted in Fig 2.

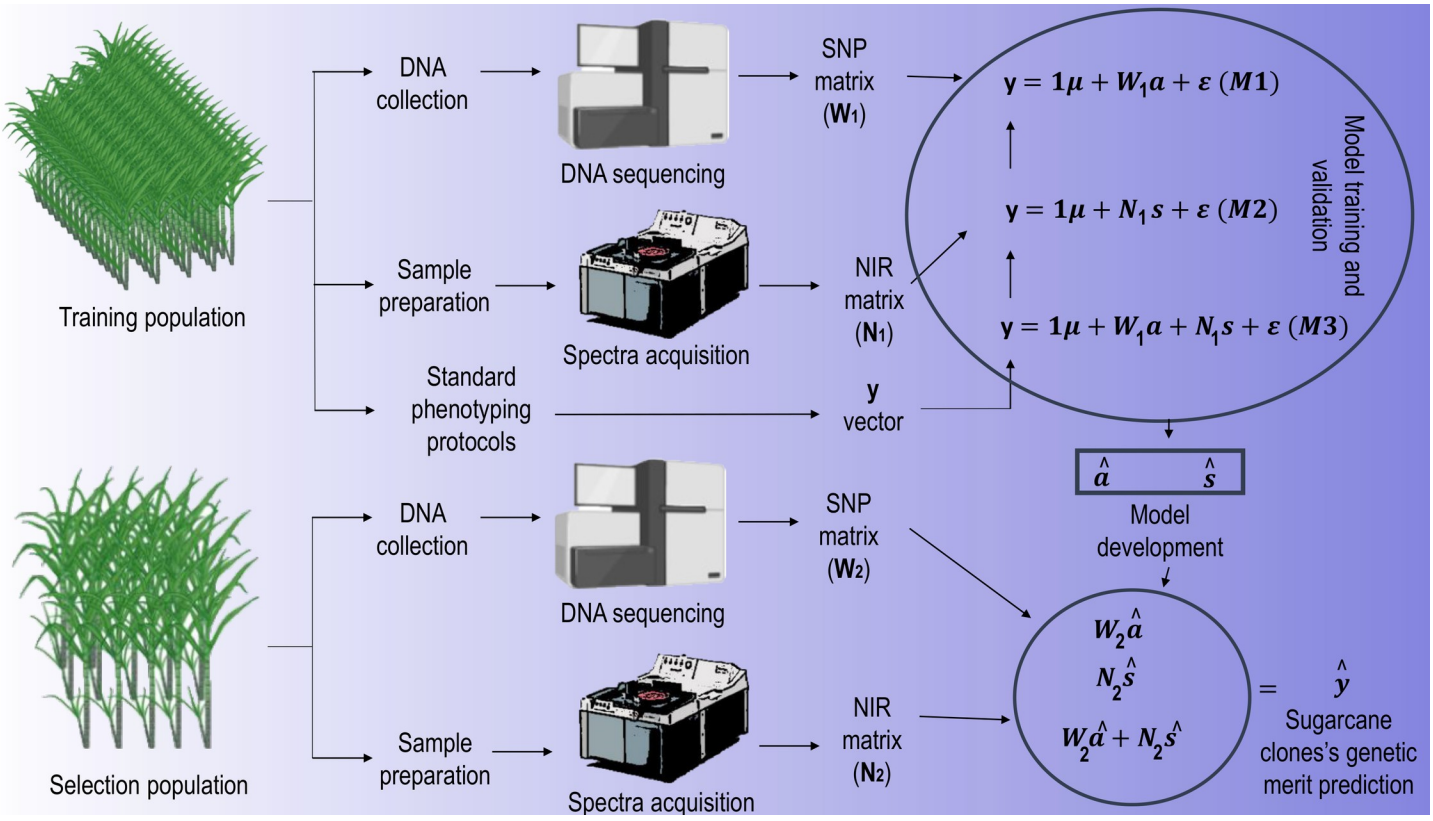

Fig 2. Schematic diagram of the experimental procedure. Created with Biorender.com.

https://doi.org/10.1371/journal.pone.0236853.g002

Results

Phenotypic data

We fitted a bivariate GBLUP (not shown) to estimate genetic correlation between the two traits; however, we found similar results compared to the correlation using univariate BLUPs. Since there was no difference, we decided to show the correlation using BLUPs. Fig 3 shows the correlation analysis using the BLUPs of FIB and PC. The data followed a Gaussian distribution curve. The two traits evaluated are negatively correlated (r = -0.22; $p < 0.001$).

The variance component estimates calculated using the REML/BLUP procedure were used to derive genetic and environmental parameters. Significant values ($p < 0.01$) of genetic variance ($\hat{\sigma}_g^2$) were observed from the deviance analysis for FIB and PC (Table 1). The estimates of individual broad-sense heritability (h^2) for FIB were considerably high. In contrast, the h^2 for PC was low. The heritability values obtained might have been the result of environment variance and the choice of the experimental design.

Prediction models

The BayesB results from cross-validation are presented in Fig 4. We found that M1 resulted in the lowest prediction accuracy for FIB and PC. Additionally, M1 models showed the largest cross-validation uncertainty. The highest prediction accuracy of M2 was obtained for FIB (0.6138), followed by PC (0.5447). The combination of SNP markers and NIR spectra in M3 models yielded an increase in the predictive ability for PC (0.5860) and a marginal improvement for FIB (0.6231) in comparison to the models fitted using only NIR spectra (M2) as predictor variables. However, Tukey's test indicated no significant ($P > 0.05$) difference in predictive performance between M2 and M3 for both traits.

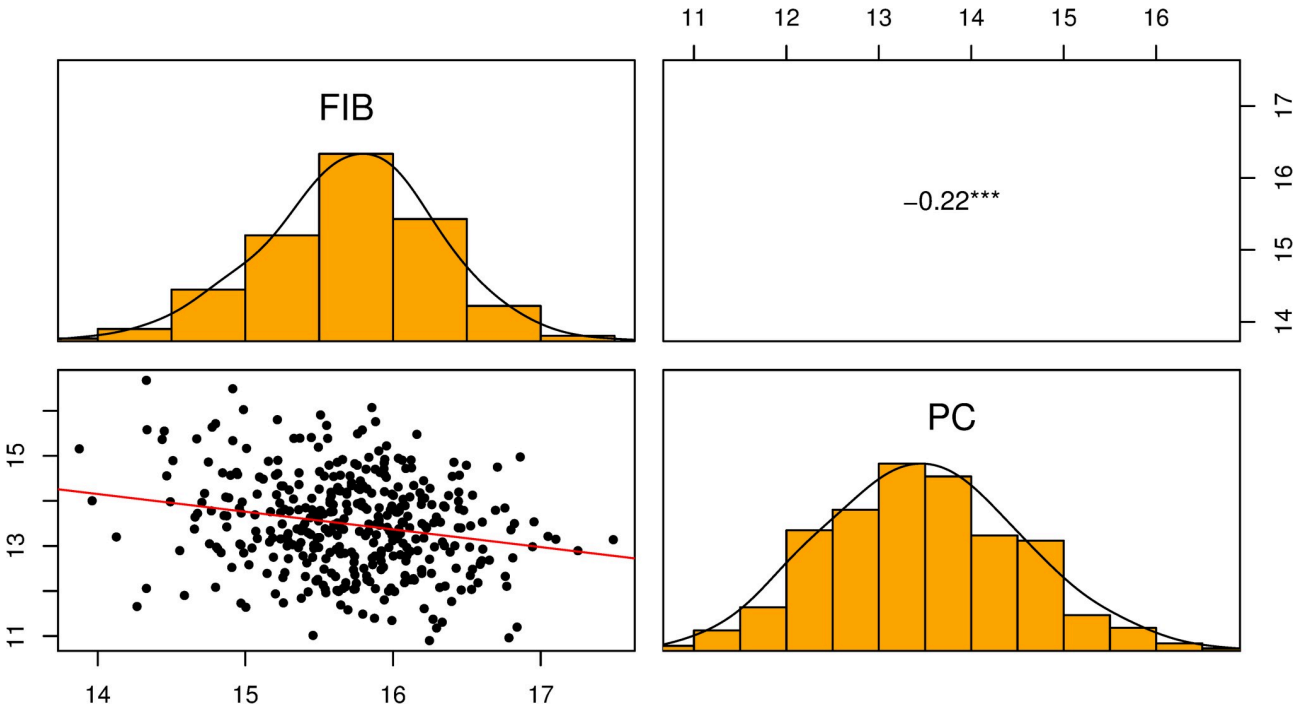

Fig 3. Scatter plot, histogram, and correlation of BLUPs of FIB and PC for 385 sugarcane clones.

https://doi.org/10.1371/journal.pone.0236853.g003

Table 1. Genetic and environmental parameter estimates of the 385 sugarcane clones evaluated.

Parameters	FIB%	PC%
$\hat{\sigma}_g^2$	1.3559*	1.0727*
$\hat{\sigma}_e^2$	0.1861	2.3032
h^2	0.8374	0.3176
General mean	13.46	15.70

$\hat{\sigma}_g^2$: Genetic variance effect; h^2: Individual plots broad sense heritability; FIB%: Fiber content; PC%: Sucrose content; * significant at 1% probability by the analyses of deviance.

https://doi.org/10.1371/journal.pone.0236853.t001

The cross-validation results using PLS regression are shown in Fig 5. Similar to BayesB, M1 models fitted using PLS resulted in the lowest prediction accuracies for both traits. The M2 model for FIB presented the highest prediction accuracy across prediction models (0.3917). Considering M3 models we observed a small increase, although no significant ($P > 0.05$), for PC (0.3942) compared to the M2 model (0.3673). In contrast, no improvement in prediction accuracy was observed for FIB.

Discussion

In the present study, we explored different strategies to incorporate genomic and HTP derived information from NIR spectroscopy into a sugarcane genetic breeding program aiming to improve prediction accuracies. Jannink et al. (2010) [39] reported the similarity of NIR spectroscopy and GP approaches, as they inherently share the same purposes and statistical analysis challenges. For instance, the application of NIR spectroscopy aims to replace demanding and expensive laboratory protocols by developing statistical prediction models using high-

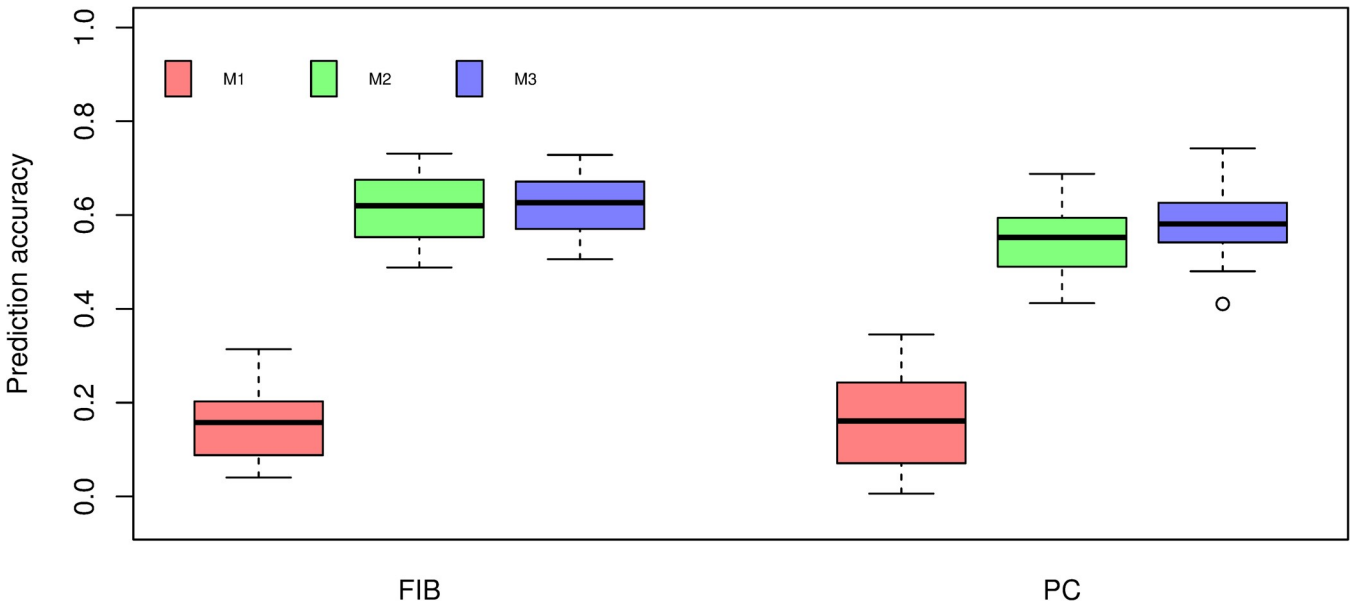

Fig 4. Box-plot of cross-validation prediction accuracy of fiber content (FIB) and sucrose content (PC) using BayesB under three different prediction models. M1: Markers. M2: Near-infrared spectra. M3: The combination of markers and near-infrared spectra. Different lowercase letters denote significant differences with Tukey's test ($P < 0.05$).

https://doi.org/10.1371/journal.pone.0236853.g004

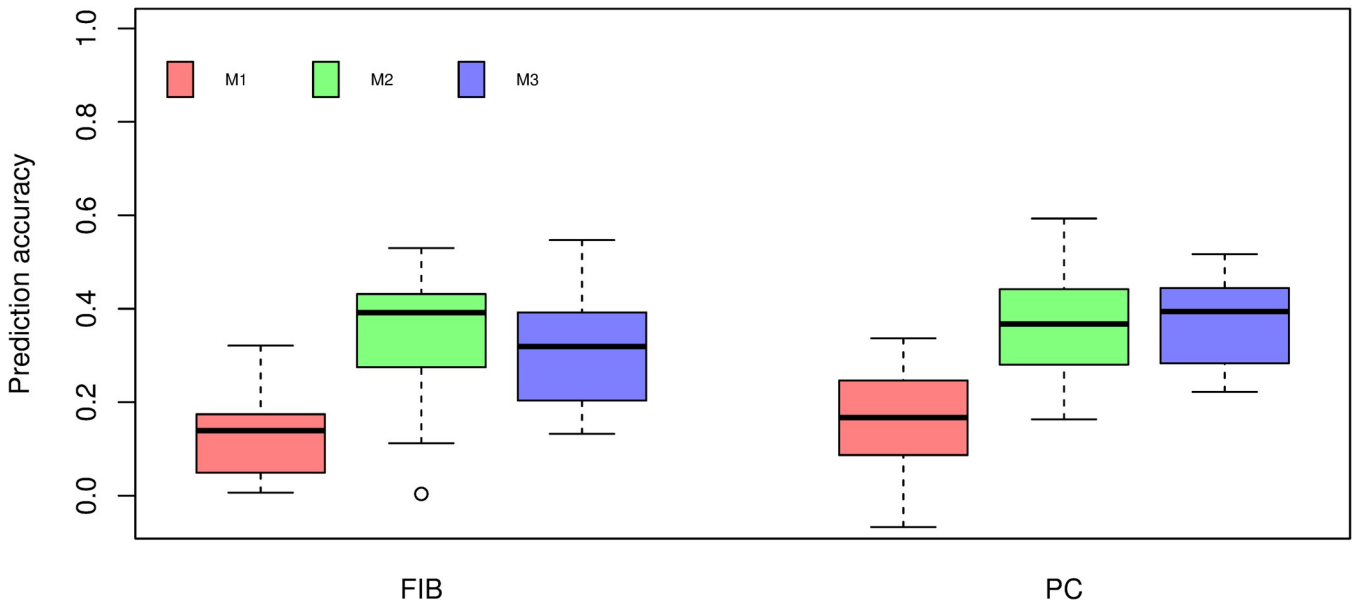

Fig 5. Box-plot of cross-validation prediction accuracy of fiber content (FIB) and sucrose content (PC) using PLS under three different prediction models. M1: Markers. M2: Near-infrared spectra. M3: The combination of markers and near-infrared spectra. Different lowercase letters denote significant differences with Tukey's test ($P < 0.05$).

https://doi.org/10.1371/journal.pone.0236853.g005

dimensional input variables. Likewise, GP is intended to utilizes multivariate statistic methods to build prediction models that associate difficult to record plant phenotypes with easy to measure variables (e.g., SNPs) [40]. Moreover, the existing statistical methods capable of coping with the challenges associated with high-dimensional statistical analysis are used across research fields [21,41–44].

Along with the integration of genomic and spectroscopy datasets, we also considered their application as single predictors in the statistical models and evaluated the impact in the prediction accuracy of two sugarcane feedstock quality traits using a commercial sugarcane population of 385 individuals and tested two regression methods. We assessed the performance of PLS regression, which is arguably the most employed method when dealing with NIR datasets [45] and BayesB for the GP models. For all models, we employed repeated random sub-sampling cross-validation to assess the accuracy of prediction models.

Phenotypic data

The results we found for genetic effects reveal the presence of genetic variance to be exploited for selection purposes. These results are in agreement with the study of Baffa et al. (2014) [46] and Wang et al. (2008) [47] for commercial sugarcane populations.

The estimated broad-sense heritability values indicate that the selection for superior clones based on phenotypic values would be effective for FIB, since environment variance showed little impact, but not for PC [48]. Ramos et al. (2017) [49] reported similar values of h^2 for FIB and PC.

The negative correlation observed between FIB and PC supports the dynamic of carbon partitioning and the antagonistic metabolic pathway of fiber and sucrose synthesis in sugarcane related in other studies, which suggest the trend of PC being negatively impacted by the increase of FIB [50,51].

Prediction models

Overall, models fitted using BayesB were most predictive than PLS models for both traits, with accuracy estimates ranging from 0.5860 to 0.6231 (Fig 5). This result is in agreement with the study of Ferragina et al., (2015) [44], in which Bayesian models outperformed PLS regression for NIR-based prediction of dairy traits. Additionally, Solberg et al. (2009) [52] reported that the prediction accuracy of genome-wide breeding values from PLS regression was lower compared to that of BayesB.

The GP models (M1) showed poor predictive ability and only explained a small portion of the phenotypic variation of the two traits evaluated. The predictive ability of GP models observed was lower than that found by Gouy et al., (2013) [42]. The authors reported accuracies for within cross-validation panels considering bagasse content (an equivalent measure of FIB) and BRIX, which is highly correlated with PC [53], of up to 0.5 and 0.62, respectively. However, they used a DArT low-density molecular marker panel, whereas we used a high-density SNP panel. The extent of coverage is a critical aspect of GP [54,55]. However, Sousa et al. (2019) [56] applied GP to polyploid hybrids of *Coffea* spp. and reported that once the optimal number of SNPs is reached, a plateau is observed followed by a decrease in predictive accuracies. Further, Yang et al. (2017) [57] reported that for sugarcane and other crops with complex genomes, the quality of sequencing might be more important than a large number of SNPs suggesting that sequencing depth is crucial to filter low-quality sequence reads.

Deomano et al. (2020) [58] tested GP in sugarcane and obtained prediction accuracies of up to 0.45 for commercial cane sugar, a trait that is related to PC. By contrast, the highest prediction accuracy we observed in the present study for PC was of 0.1607. One possible explanation for this result might be the fact that the authors used a considerably larger training population than ours. It has been reported that increasing training population size can result in better predictive ability [59]. In the case of sugarcane and other polyploids, this factor is expected to play an even bigger role, given the expected allelic diversity [60–62].

Another factor that could explain the low level of prediction accuracies we observed is that the GP models fitted in this study only considered additive effects. Zeni Neto et al. (2013) [63] reported that additive and non-additive genetic effects are equally important for the determination of complex traits in sugarcane. Hence, the inclusion of non-additive effects in GP models for sugarcane and other clonally propagated species may improve prediction accuracies [64]. Results obtained by Denis et al. (2013) [65] in a simulation study, and de Almeida Filho et al. (2014) [66] using data from a full-sib population of loblolly pine (*Pinus taeda* L.) indicate the improvement of prediction accuracy when accounting for dominance effects. Nevertheless, according to Wei et al. (2016) [67], the slow rate of genetic gains reported for sugarcane yield in recent decades can be partly attributed to the low levels of narrow-sense heritability that most commercial sugarcane breeding populations exhibit [47,68,69]. However, the authors also stated that this could be a poor estimation of additive genetic component based on phenotypic evaluation. Therefore, the genomic models that account only for additive effects would still have usefulness to better estimate genetic values and help identify superior genitors.

In sugarcane and other polyploid crops, the estimation of allele dosage might improve predictions of GP models [70,71]. The utilization of single-dose marker systems in ploypoids is considered less informative [72]. For instance, given the high heterozygosity and varying ploidy levels of sugarcane cultivars, for every cross each locus will segragate into multiple genotype classes. Hence, by using single-dose markers, different heterozygous genotypes will not be distinguished [73]. Recent studies have investigated polyploid parametrization and reported improvment in predicitons [71,74]. However, the big and complex poly-aneuploid genome of modern sugarcane hybrids has hindered the development of tools that provide a reliable

estimation of allele dosage [75]. These pieces of evidence suggest that significant bottlenecks in the application of GP to sugarcane are sequencing and sequencig data processing, as well as proper statistical methods to handle the complex inherent patterns of polyploidy [73,76]. Some research efforts to meet these limitations are underway [73,77,78].

The M3 modelling strategy we evaluated in this study aimed to increase accuracies by combining NIR wavenumbers and SNP markers to predict sugarcane clones' phenotypes for possible release as cultivars. This approach has been tested using different omic predictors (e.g., metabolomic and transcriptomic datasets [79–82]). Riedelsheimer et al. (2012) [80] jointly used metabolites and SNPs to predict general combining ability of maize hybrids. However, the authors found no improvement in predictions. Crain et al. (2018) [21] investigated a similar application we proposed herein and found the same trend. The authors evaluated the effect of including HTP data into GP models in different stressed environmental conditions. The results of their study revealed that models including HTP derived information as single predictors contained most of the predictive ability when compared to models with markers alone in one of the evaluated scenarios, which is consistent with the results we observed. We expected that the combination of M1 and M2 models would bring synergy and thus, improve model performance. However, the combination of predictors produced no significant improvement in comparison to the use of NIR wavenumber predictors alone (M2). Seemingly, most of the variation of the two evaluated traits that were captured by the M3 model came from the NIR spectra. Indeed, according to Rutkoski (2019) [83], the use of molecular markers is rather an indirect form of selection because it is performed based on genotypes. In contrast, selection using NIR spectroscopy is performed directly on phenotypes, which could be a possible reason to explain this result.

A wide range of references regarding the utilization of NIR spectroscopy in agriculture-related topics is available. For instance, studies employing NIR as an analytical tool include post-harvest quality monitoring [84], toxic compounds detection in seeds [35], and grain composition determination [85,86]. Further, Hayes et al. (2017) [87] performed NIR predictions of 19 wheat end-use quality traits using multi-trait analysis and obtained improved accuracies of genomic predictions. In the context of sugarcane, applications are focused on screening biomass sample physical properties and chemical composition [51,88–92]. Nevertheless, examples that make use of NIR spectroscopy for breeding purposes are also available and include the classification of sugarcane clones based on quality parameters [93], resistance to diseases [94], and pests [17]. Moreover, plant breeders have been benefiting from NIR spectroscopy using spectrometer sensors coupled with UAV and ground-based platforms [11,95,96]. Rincent et al., (2018) [97] proposed an approach in which relationship matrices are derived from NIR spectra data and compared the efficiency of predictions with standard GP models considering markers. Krause et al., (2019) [98] extended this concept by using hyperspectral reflectance derived relationship matrices and by modeling genotype × environment interactions. The results found by these authors suggest that models developed using NIR data can outperform GP models. Likewise, we observed with our dataset that NIR-based models alone provided better results. Finally, NIR spectroscopy may offer more opportunities to assist plant breeders with the advent of portable low-cost instruments [99,100].

Genomic prediction and NIR spectroscopy implementation

Sugarcane is cultivated in a semi-perennial scheme and its multiplication is performed by vegetative propagation [101]. The initial step of an ordinary sugarcane breeding program consists of crossing pre-selected elite progenitors [6]. Modern cultivated sugarcane clones feature a complex genome structure, rendering each cross unpredictable [102]. Consequently, large

progeny populations are generated [101]. After crossings, the first stage of clonal selection at the PMGCA-UFV is referred to as T1 [103]. At T1, an array of limitations including physical space and short supply of propagation material precludes the installation of appropriate statistical experimental designs, which contributes to diminishing the selection accuracy, especially regarding low heritability traits [25]. Moreover, most sugarcane traits are believed to be quantitatively inherited [104,105]. Therefore, the subsequent stages of selection involve capital intensive field trials over multiple sites and years. Today, one breeding cycle of phenotypic selection in sugarcane can take up to 13 years [105,106].

The optimal strategy to incorporate GP into breeding schemes is not straightforward and needs consideration, since it can be influenced by many factors [107]. In the context of sugarcane breeding, some authors have argued that the main benefit of applying GP could come from the length reduction of breeding cycles by performing early selection of parental clones [58,105]. Sugarcane breeding programs typically adopt into their pipelines intrapopulational recurrent selection (IRS) schemes [108]. In IRS schemes, clones are reintroduced to the breeding cycle for new hybridizations as candidate superior parents. Presently, the strategy at the PMGCA-UFV is to evaluate candidates for selection in time-consuming field trials with the classical BLUP methodology, using phenotypes and pedigree records for the estimation of breeding values [109]. However, several studies indicate that the utilization of the realized relationship matrix based on markers is preferred because it allows the estimation of the Mendelian sampling term and is less prone to errors [110,111]. Moreover, the combination of genomic and pedigree information could provide a more reliable estimation of breeding values for parental selection in the IRS scheme, thereby improving selection accuracy [6,58,112]. Further, even with low accuracies, gains per unit of cost can be obtained when compared to conventional phenotypic selection with the elimination of seedlings that exhibit poor performance, i.e., lowest genomic estimated breeding values before the installation of early field trials; therefore, saving resources and optimizing the next selection stages [112–114]. Nevertheless, costs need to be thoroughly considered before the routine implementation of GP in this stage because a massive number of seedlings are generated to form the initial base populations.

According to Yadav et al., (2020) [105], the costs associated with high-throughput genotyping can be a major impediment for the adoption of large scale GP in sugarcane breeding. However, the decreasing costs of genotyping platforms and the integration with HTP platforms is encouraging [1,11,40]. The combination of HTP platforms (e.g., NIR spectrcopy) and GP is heralded to promote genetic gains by improving selection accuracy and reducing costs, mainly related to conventional phenotyping sessions [2,11]. From this perspective, NIR spectroscopy could be most useful to train GP models, screening more efficiently large training populations in order to estimate and continuously re-estimate marker effects [10]. In this setting, phenotypes are the predictions based on the NIR spectra. Alternatively, NIR data finds applicability in indirect selection, with the inclusion of phenotypes based on NIR spectra or spectra derived indexes into multi-trait genomic prediction models [87,115]. Further, NIR spectroscopy could likely be used to select superior clones for cultivar development with the replacement of costly and less effective phenotyping protocols in advanced selection stages thus, maximizing resources. The trend of decreasing costs in NIR instrumentation is clear, with the advent of portable instruments [99]. Therefore, it seems likely to be readily available for implementation as a low-cost phenotyping tool.

Conclusion

Our experimental results showed that GP models had the lowest prediction accuracies. In addition, the combination of NIR wavenumbers and SNP markers as predictor variables did

not demonstrate significant improvements in accuracy to predict FIB and PC, when compared to the models solely based on NIR wavenumbers. NIR wavenumber predictors alone achieved high prediction accuracies for the two traits assessed in this study, indicating the potential usefulness of NIR spectroscopy to train GP models and for predicting total phenotypic value of sugarcane clones for possible release as cultivars. We speculate that the combination of GP and NIR spectroscopy has the potential to enable genetic gains and accelerate the release of new cultivars in sugarcane breeding programs by reducing the length of breeding cycles and improving selection efficiency.

Supporting information

S1 Fig. Overview of data acquisition. A: stalks being harvested from double-row plots; B: stationary forage chopper machine used to shred stalks; C: hydraulic press used to extract the fiber cake and juice samples; D: fiber cake being weighted; E: samples being dried at a forced-air circulating oven; F: dried ground samples placed onto the NIR instrument window; G: saccharimeter instrument; H: sample spectrum displayed on the computer screen.
(TIF)

Acknowledgments

The authors thank Professor Dr. Luis Antônio dos Santos Dias for providing the NIR instrument used in this study. Also, we acknowledge Professor Dr. Reinaldo Francisco Teófilo and Dr. Jussara Valente Roque for the helpful assistance regarding NIR spectra collection and downstream analyses. Finally, we acknowledge the numerous co-operators from the Sugarcane Genetic Breeding Research Station (CECA- Minas Gerais State, Brazil) who helped carrying out field trials and to collect phenotypic data.

Author Contributions

Conceptualization: Mateus Teles Vital Gonçalves, Luiz Alexandre Peternelli.

Data curation: Mateus Teles Vital Gonçalves, Paulo Mafra de Almeida Costa, Pedro Marcus Pereira Vidigal, Luiz Alexandre Peternelli.

Formal analysis: Mateus Teles Vital Gonçalves, Pedro Marcus Pereira Vidigal, Luiz Alexandre Peternelli.

Funding acquisition: Marcio Henrique Pereira Barbosa, Luiz Alexandre Peternelli.

Investigation: Mateus Teles Vital Gonçalves, Paulo Mafra de Almeida Costa.

Methodology: Mateus Teles Vital Gonçalves, Gota Morota, Paulo Mafra de Almeida Costa, Pedro Marcus Pereira Vidigal, Luiz Alexandre Peternelli.

Project administration: Marcio Henrique Pereira Barbosa, Luiz Alexandre Peternelli.

Software: Luiz Alexandre Peternelli.

Supervision: Marcio Henrique Pereira Barbosa, Luiz Alexandre Peternelli.

Visualization: Mateus Teles Vital Gonçalves, Luiz Alexandre Peternelli.

Writing – original draft: Mateus Teles Vital Gonçalves, Pedro Marcus Pereira Vidigal.

Writing – review & editing: Gota Morota, Paulo Mafra de Almeida Costa, Luiz Alexandre Peternelli.

References

1. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. Nat Rev Genet. 2016; 17: 333–351. https://doi.org/10.1038/nrg.2016.49 PMID: 27184599
2. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci. 2017; 22: 961–975. https://doi.org/10.1016/j.tplants.2017.08.011 PMID: 28965742
3. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 2001; 157: 1819–1829. PMID: 11290733
4. Meuwissen T, Hayes B, Goddard M. Accelerating Improvement of Livestock with Genomic Selection. Annu Rev ofAnimal Biosci. 2013; 1: 221–237. https://doi.org/10.1146/annurev-animal-031412-103705 PMID: 25387018
5. Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME. Plant breeding with Genomic selection: Gain per unit time and cost. Crop Sci. 2010; 50: 1681–1690. https://doi.org/10.2135/cropsci2009.11.0662
6. Barbosa MHP, Resende MDV, Dias LA dos S, Barbosa GV de S, Oliveira RA, Peternelli LA, et al. Genetic improvement of sugar cane for bioenergy: the Brazilian experience in network research with RIDESA. Crop Breed Appl Biotechnol. 2012; S2: 87–98.
7. Bernardo R. Bandwagons I, too, have known. Theor Appl Genet. 2016; 129: 2323–2332. https://doi.org/10.1007/s00122-016-2772-5 PMID: 27681088
8. Asoro FG, Newell MA, Scott MP. Selection Methods for β-Glucan Concentration in Elite Oat. Crop Sci. 2013; 53: 1894–1906.
9. Cabrera-Bosquet L, Crossa J, von Zitzewitz J, Serret MD, Luis Araus J. High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding Converge. J Integr Plant Biol. 2012; 54: 312–320. https://doi.org/10.1111/j.1744-7909.2012.01116.x PMID: 22420640
10. Heffner EL, Sorrells ME, Jannink JL. Genomic selection for crop improvement. Crop Sci. 2009; 49: 1–12. https://doi.org/10.2135/cropsci2008.08.0512
11. Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS, Cairns JE. Translating High-Throughput Phenotyping into Genetic Gain. Trends Plant Sci. 2018; 23: 451–466. https://doi.org/10.1016/j.tplants.2018.02.001 PMID: 29555431
12. Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet. 2013; 126: 867–887. https://doi.org/10.1007/s00122-013-2066-0 PMID: 23471459
13. Araus JL, Cairns JE. Field high-throughput phenotyping: The new crop breeding frontier. Trends Plant Sci. 2014; 19: 52–61. https://doi.org/10.1016/j.tplants.2013.09.008 PMID: 24139902
14. Zhao C, Zhang Y, Du J, Guo X, Wen W, Gu S, et al. Crop phenomics: Current status and perspectives. Front Plant Sci. 2019;10. https://doi.org/10.3389/fpls.2019.00010 PMID: 30766542
15. Furbank RT, Tester M. Phenomics—technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 2011; 16: 635–644. https://doi.org/10.1016/j.tplants.2011.09.005 PMID: 22074787
16. Montes JM, Melchinger AE, Reif JC. Novel throughput phenotyping platforms in plant genetic studies. Trends Plant Sci. 2007; 12: 433–436. https://doi.org/10.1016/j.tplants.2007.08.006 PMID: 17719833
17. Porto N de A, Roque J V., Wartha CA, Cardoso W, Peternelli LA, Barbosa MHP, et al. Early prediction of sugarcane genotypes susceptible and resistant to Diatraea saccharalis using spectroscopies and classification techniques. Spectrochim Acta—Part A Mol Biomol Spectrosc. 2019; 218: 69–75. https://doi.org/10.1016/j.saa.2019.03.114 PMID: 30954799
18. Valderrama P, Braga JWB, Jesus PR. Variable Selection, Outlier Detection, and Figures of Merit Estimation in a Partial Least-Squares Regression Multivariate Calibration Model. A Case Study for the Determination of Quality Parameters in the. J Agric Food Chem. 2007; 55: 8331–8338. https://doi.org/10.1021/jf071538s PMID: 17927144
19. Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. Plant Phenomics, From Sensors to Knowledge. Curr Biol. 2017; 27: R770–R783. https://doi.org/10.1016/j.cub.2017.05.055 PMID: 28787611
20. Morota G, Jarquin D, Campbell MT, Iwata H. Statistical methods for the quantitative genetic analysis of high-throughput phenotyping data. 2019; 1–47. Available: http://arxiv.org/abs/1904.12341.
21. Crain J, Mondal S, Rutkoski J, Singh RP, Poland J. Combining High-Throughput Phenotyping and Genomic Information to Increase Prediction and Selection Accuracy in Wheat Breeding. Plant Genome. 2018; 11: 1–14. https://doi.org/10.3835/plantgenome2017.05.0043 PMID: 29505641

22. Aguate FM, Trachsel S, González Pérez L, Burgueño J, Crossa J, Balzarini M, et al. Use of hyperspectral image data outperforms vegetation indices in prediction of maize yield. Crop Sci. 2017; 57: 2517–2524. https://doi.org/10.2135/cropsci2017.01.0007

23. Cuevas J, Montesinos-López O, Juliana P, Guzmán C, Pérez-Rodríguez P, González-Bucio J, et al. Deep Kernel for genomic and near infrared predictions in multi-environment breeding trials. G3 Genes, Genomes, Genet. 2019; 9: 2913–2924. https://doi.org/10.1534/g3.119.400493 PMID: 31289023

24. Federer WT. Augmented designs with one-way elimination of heterogeneity. Biometrics. 1961; 17: 447–473.

25. Leite MSDO, Peternelli LA, Barbosa MHP. Effects of plot size on the estimation of genetic parameters in sugarcane families. Crop Breed Appl Biotechnol. 2006; 6: 40–46.

26. Consecana. Manual de instruções (5th ed.) Piracicaba,São Paulo: Conselho doa Produtores de Cana-de-Açúcar, Açúcar e Álcool do Estado de São Paulo. 2006.

27. Melo FDAD. Remuneration system of sugarcane. Sugarcane: Agricultural Production, Bioenergy and Ethanol. Elsevier Inc.; 2015. https://doi.org/10.1016/B978-0-12-802239-9.00019–0

28. 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: 1–10. https://doi.org/10.1038/s41467-017-02088-w PMID: 29317637

29. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26: 589–595. https://doi.org/10.1093/bioinformatics/btp698 PMID: 20080505

30. Resende MDV. Software Selegen-REML/BLUP: A useful tool for plant breeding. Crop Breed Appl Biotechnol. 2016; 16: 330–339. https://doi.org/10.1590/1984-70332016v16n4a49

31. Engel J, Gerretzen J, Szymańska E, Jansen JJ, Downey G, Blanchet L, et al. Breaking with trends in preprocessing? TrAC—Trends Anal Chem. 2013; 50: 96–106. https://doi.org/10.1016/j.trac.2013.04.015

32. Rinnan Å, Berg F Van Den, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. Trends Anal Chem. 2009; 28: 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

33. Wold S, Sjöström M, Eriksson L. PLS-regression: A basic tool of chemometrics. Chemom Intell Lab Syst. 2001; 58: 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

34. James G, Witen D, Hastie T, Tibshirani R. An Introduction to Statistical Learning with Applications in R. 2007. https://doi.org/10.1016/j.peva.2007.06.006

35. Roque J V., Dias LAS, Teófilo RF. Multivariate Calibration to Determine Phorbol Esters in Seeds of Jatropha curcas L. Using Near Infrared and Ultraviolet Spectroscopies. Jorunal Brazilian Chem Soc. 2017; 28: 1506–1516.

36. Pérez P, De Los Campos G. Genome-wide regression and prediction with the BGLR statistical package. Genetics. 2014; 198: 483–495. https://doi.org/10.1534/genetics.114.164442 PMID: 25009151

37. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol. 2017; 13: 1–19. https://doi.org/10.1371/journal.pcbi.1005752 PMID: 29099853

38. Harris R. An Introduction to R. Quantitative Geography: The Basics. 2020. pp. 250–286. https://doi.org/10.4135/9781473920446.n12

39. Jannink JL, Lorenz AJ, Iwata H. Genomic selection in plant breeding: From theory to practice. Briefings Funct Genomics Proteomics. 2010; 9: 166–177. https://doi.org/10.1093/bfgp/elq001 PMID: 20156985

40. Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, et al. Crop Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. Mol Plant. 2017; 10: 1047–1064. https://doi.org/10.1016/j.molp.2017.06.008 PMID: 28669791

41. Boulesteix AL, Strimmer K. Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. oinformaticsBriefings Bioinforma. 2007; 8: 32–44. https://doi.org/10.1093/bib/bbl016 PMID: 16772269

42. Gouy M, Rousselle Y, Bastianelli D, Lecomte P, Bonnal L, Roques D, et al. Experimental assessment of the accuracy of genomic selection in sugarcane. Theor Appl Genet. 2013; 126: 2575–2586. https://doi.org/10.1007/s00122-013-2156-z PMID: 23907359

43. Hernandez J, Lobos GA, Matus I, del Pozo A, Silva P, Galleguillos M. Using ridge regression models to estimate grain yield from field spectral data in bread wheat (Triticum Aestivum L.) grown under three water regimes. Remote Sens. 2015; 7: 2109–2126. https://doi.org/10.3390/rs70202109

44. Ferragina A, de los Campos G, Vazquez AI, Cecchinato A, Bittante G. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. J Dairy Sci. 2015; 98: 8133–8151. https://doi.org/10.3168/jds.2014-9143 PMID: 26387015

45. Pasquini C. Analytica Chimica Acta Near infrared spectroscopy: A mature analytical technique with new perspectives e A review. Anal Chim Acta. 2018; 1026: 8–36. https://doi.org/10.1016/j.aca.2018.04.004 PMID: 29852997

46. Baffa DCF, Costa PM d. A, da Silveira G, Lopes FJF, Barbosa MHP, Loureiro ME, et al. Path analysis for selection of saccharification-efficient sugarcane genotypes through agronomic traits. Agron J. 2014; 106: 1643–1650. https://doi.org/10.2134/agronj13.0576

47. Wang LP, Jackson PA, Lu X, Fan YH, Foreman JW, Chen XK, et al. Evaluation of sugarcane x Saccharum spontaneum progeny for biomass composition and yield components. Crop Sci. 2008; 48: 951–961. https://doi.org/10.2135/cropsci2007.10.0555

48. Dumont T, Thong-Chane A, Barau L, Siegmund B, Hoarau JY. Genetic Variabilities and Genetic Gains for Yield Components in Regional Sugarcane Breeding Programmes on Réunion Island. Sugar Tech. 2019; 21: 868–878. https://doi.org/10.1007/s12355-019-00718-9

49. Ramos RS, Brasileiro BP, Da Silveira LCI, Kist V, Peternelli LA, Barbosa MHP. Selecting parents, families, and clones to obtain energy cane. Agron J. 2017; 109: 762–768. https://doi.org/10.2134/agronj2016.09.0489

50. da Silveira LCI, Brasileiro BP, Kist V, Weber H, Daros E, Peternelli LA, et al. Selection in energy cane families. Crop Breed Appl Biotechnol. 2016; 16: 298–306. https://doi.org/10.1590/1984-70332016v16n4a45

51. Hoang N V, Furtado A, Donnan L, Keeffe EC, Botha FC, Henry RJ. High-Throughput Profiling of the Fiber and Sugar Composition of Sugarcane Biomass. 2017; 400–416. https://doi.org/10.1007/s12155-016-9801-8

52. Solberg TR, Sonesson AK, Woolliams JA, Meuwissen TH. Reducing dimensionality for prediction of genome-wide breeding values. Genet Sel Evol. 2009; 41: 1–8. https://doi.org/10.1186/1297-9686-41-1 PMID: 19284684

53. Mancini MC, Leite DC, Perecin D, Bidóia MAP, Xavier MA, Landell MGA, et al. Characterization of the Genetic Variability of a Sugarcane Commercial Cross Through Yield Components and Quality Parameters. Sugar Tech. 2012; 14: 119–125. https://doi.org/10.1007/s12355-012-0141-5

54. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: Key considerations in genomic analyses. Nature Reviews Genetics. Nature Publishing Group; 2014. pp. 121–132. https://doi.org/10.1038/nrg3642 PMID: 24434847

55. Xu Y, Lu Y, Xie C, Gao S, Wan J, Prasanna BM. Whole-genome strategies for marker-assisted plant breeding. Mol Breed. 2012; 29: 833–854. https://doi.org/10.1007/s11032-012-9699-6

56. Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, Pereira AA, Sakiyama NS, et al. Early selection enabled by the implementation of genomic selection in coffea arabica breeding. Front Plant Sci. 2019; 9: 1–12. https://doi.org/10.3389/fpls.2018.01934 PMID: 30671077

57. Yang X, Song J, You Q, Paudel DR, Zhang J, Wang J. Mining sequence variations in representative polyploid sugarcane germplasm accessions. BMC Genomics. 2017; 18: 1–16. https://doi.org/10.1186/s12864-016-3406-7 PMID: 28049423

58. Deomano E, Jackson P, Wei X, Aitken K, Kota R, Pérez-Rodríguez P. Genomic prediction of sugar content and cane yield in sugar cane clones in different stages of selection in a breeding program, with and without pedigree information. Mol Breed. 2020; 40. https://doi.org/10.1007/s11032-019-1090-4 PMID: 31975784

59. Berro I, Lado B, Nalin RS, Quincke M, Gutiérrez L. Training Population Optimization for Genomic Selection. Plant Genome. 2019; 12: 190028. https://doi.org/10.3835/plantgenome2019.04.0028 PMID: 33016595

60. Udall JA, Wendel JF. Polyploidy and crop improvement. Crop Sci. 2006; 46: 3–14. https://doi.org/10.2135/cropsci2006.07.0489tpg

61. Sverrisdóttir E, Sundmark EHR, Johnsen HØ, Kirk HG, Asp T, Janss L, et al. The value of expanding the training population to improve genomic selection models in tetraploid potato. Front Plant Sci. 2018; 9: 1–14. https://doi.org/10.3389/fpls.2018.00001 PMID: 29410674

62. Song J, Yang X, Resende MFR, Neves LG, Todd J, Zhang J, et al. Natural allelic variations in highly polyploidy Saccharum complex. Front Plant Sci. 2016; 7: 1–18. https://doi.org/10.3389/fpls.2016.00001 PMID: 26858731

63. Zeni Neto H, Daros E, Bespalhok Filho JC, Scapim CA, Vidigal MCG, Vidigal Filho PS. Selection of families and parents of sugarcane (Saccharum spp.) through mixed models by joint analysis of two harvests. Euphytica. 2013; 193: 391–408. https://doi.org/10.1007/s10681-013-0947-6

64. De Almeida Filho JE, Guimarães JFR, Fonsceca E Silva F, De Resende MDV, Muñoz P, Kirst M, et al. Genomic prediction of additive and non-additive effects using genetic markers and pedigrees. G3 Genes, Genomes, Genet. 2019; 9: 2739–2748. https://doi.org/10.1534/g3.119.201004 PMID: 31263059

65. Denis M, Bouvet JM. Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding. Tree Genet Genomes. 2013; 9: 37–51. https://doi.org/10.1007/s11295-012-0528-1

66. De Almeida Filho JE, Guimarães JFR, E Silva FF, De Resende MDV, Muñoz P, Kirst M, et al. The contribution of dominance to phenotype prediction in a pine breeding and simulated population. Heredity (Edinb). 2016; 117: 33–41. https://doi.org/10.1038/hdy.2016.23 PMID: 27118156

67. Wei X, Jackson P. Addressing slow rates of long-term genetic gain in sugarcane. Int Sugar J. 2017; 119: 1923–1930.

68. de Carvalho MP, Gezan SA, Peternelli LA, Barbosa MHP. Estimation of additive and nonadditive genetic components of sugarcane families using multitrait analysis. Agron J. 2014; 106: 800–808. https://doi.org/10.2134/agronj2013.0247

69. Silva FL da, Barbosa MHP, de Resende MDV, Peternelli LA, Pedrozo CÂ. Efficiency of selection within sugarcane families via simulated individual BLUP. Crop Breed Appl Biotechnol. 2015; 15: 1–9. https://doi.org/10.1590/1984-70332015v15n1a1

70. Garcia AAF, Mollinari M, Marconi TG, Serang OR, Silva RR, Vieira MLC, et al. SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. Sci Rep. 2013; 3: 1–10. https://doi.org/10.1038/srep03399 PMID: 24292365

71. Matias FI, Alves FC, Meireles KGX, Barrios SCL, do Valle CB, Endelman JB, et al. On the accuracy of genomic prediction models considering multi-trait and allele dosage in Urochloa spp. interspecific tetraploid hybrids. Mol Breed. 2019;39. https://doi.org/10.1007/s11032-019-1002-7

72. Ferrão LF V., Benevenuto J, Oliveira I de B, Cellon C, Olmstead J, Kirst M, et al. Insights Into the Genetic Basis of Blueberry Fruit-Related Traits Using Diploid and Polyploid Models in a GWAS Context. Front Ecol Evol. 2018;6. https://doi.org/10.3389/fevo.2018.00107

73. Bourke PM, Voorrips RE, Visser RGF, Maliepaard C. Tools for genetic studies in experimental populations of polyploids. Front Plant Sci. 2018; 9: 1–17. https://doi.org/10.3389/fpls.2018.00001 PMID: 29410674

74. Endelman JB, Carley CAS, Bethke PC, Coombs JJ, Clough ME, Silva WL, et al. Genetic Variance Partitioning and Genome-Wide Autotetraploid Potato. Genetics. 2018; 209: 77–87. https://doi.org/10.1534/genetics.118.300685 PMID: 29514860

75. Olatoye MO, Clark L V., Wang J, Yang X, Yamada T, Sacks EJ, et al. Evaluation of genomic selection and marker-assisted selection in Miscanthus and energycane. Mol Breed. 2019;39. https://doi.org/10.1007/s11032-019-1081-5

76. Thirugnanasambandam PP, Hoang N V., Henry RJ. The challenge of analyzing the sugarcane genome. Front Plant Sci. 2018; 9: 1–18. https://doi.org/10.3389/fpls.2018.00001 PMID: 29410674

77. Souza GM, Van Sluys MA, Lembke CG, Lee H, Margarido GRA, Hotta CT, et al. Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. Gigascience. 2019; 8: 1–18. https://doi.org/10.1093/gigascience/giz129 PMID: 31782791

78. Jiao WB, Schneeberger K. The impact of third generation genomic technologies on plant genome assembly. Curr Opin Plant Biol. 2017; 36: 64–70. https://doi.org/10.1016/j.pbi.2017.02.002 PMID: 28231512

79. Li Z, Gao N, Martini JWR, Simianer H. Integrating gene expression data into genomic prediction. Front Genet. 2019; 10: 1–11. https://doi.org/10.3389/fgene.2019.00001 PMID: 30804975

80. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, et al. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet. 2012; 44: 217–220. https://doi.org/10.1038/ng.1033 PMID: 22246502

81. Schrag TA, Westhues M, Schipprack W, Seifert F, Thiemann A, Scholten S, et al. Beyond genomic prediction: Combining different types of omics data can improve prediction of hybrid performance in maize. Genetics. 2018; 208: 1373–1385. https://doi.org/10.1534/genetics.117.300374 PMID: 29363551

82. Guo Z, Magwire MM, Basten CJ, Xu Z, Wang D. Evaluation of the utility of gene expression and metabolic information for genomic prediction in maize. Theor Appl Genet. 2016; 129: 2413–2427. https://doi.org/10.1007/s00122-016-2780-5 PMID: 27586153

83. Rutkoski JE. A practical guide to genetic gain. First edit. Advances in Agronomy. First edit. Elsevier Inc.; 2019. pp. 217–249. https://doi.org/10.1016/bs.agron.2019.05.001

84. Moscetti R, Haff RP, Saranwong S, Monarca D, Cecchini M, Massantini R. Nondestructive detection of insect infested chestnuts based on NIR spectroscopy. Postharvest Biol Technol. 2014; 87: 88–94. https://doi.org/10.1016/j.postharvbio.2013.08.010

85. Melchinger AE, Böhm J, Utz HF, Müller J, Munder S, Mauch FJ. High-throughput precision phenotyping of the oil content of single seeds of various oilseed crops. Crop Sci. 2018; 58: 670–678. https://doi.org/10.2135/cropsci2017.07.0429

86. Ferreira SL, Vasconcellos RS, Rossi RM, de Paula VRC, Fachinello MR, Huepa LMD, et al. Using near infrared spectroscopy to predict metabolizable energy of corn for pigs. Sci Agric. 2018; 75: 486–493. https://doi.org/10.1590/1678-992x-2016-0509

87. Hayes BJ, Panozzo J, Walker CK, Choy AL, Kant S, Wong D, et al. Accelerating wheat breeding for end-use quality with multi-trait genomic predictions incorporating near infrared and nuclear magnetic resonance-derived phenotypes. Theor Appl Genet. 2017; 130: 2505–2519. https://doi.org/10.1007/s00122-017-2972-7 PMID: 28840266

88. Caliari ÍP, Barbosa MHP, Ferreira SO, Teófilo RF. Estimation of cellulose crystallinity of sugarcane biomass using near infrared spectroscopy and multivariate analysis methods. Carbohydr Polym. 2017; 158: 20–28. https://doi.org/10.1016/j.carbpol.2016.12.005 PMID: 28024538

89. Valderrama P, Braga JWB, Poppi RJ. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. J Agric Food Chem. 2007; 55: 8331–8338. https://doi.org/10.1021/jf071538s PMID: 17927144

90. Taira E, Ueno M, Saengprachatanarug K, Kawamitsu Y. Direct sugar content analysis for whole stalk sugarcane using a portable near infrared instrument Direct sugar content analysis for whole stalk sugarcane using a portable near infrared instrument. J NEAR INFRARED Spectrosc. 2013; 21: 281–287. https://doi.org/10.1255/jnirs.1064

91. Assis C, Ramos RS, Silva LA, Kist V, Barbosa MHP, Teófilo RF. Prediction of Lignin Content in Different Parts of Sugarcane Using Near-Infrared Spectroscopy (NIR), Ordered Predictors Selection (OPS), and Partial Least Squares (PLS). Appl Spectrosc. 2017; 71: 2001–2012. https://doi.org/10.1177/0003702817704147 PMID: 28452227

92. Sabatier D, Thuriès L, Bastianelli D, Dardenne P. Rapid prediction of the lignocellulosic compounds of sugarcane biomass by near infrared reflectance spectroscopy: Comparing classical and independent cross-validation. J Near Infrared Spectrosc. 2012; 20: 371–385. https://doi.org/10.1255/jnirs.999

93. Peternelli LA, Gonçalves MTV, Fernandes JG, Brasileiro BP, Teófilo RF. Selection of sugarcane clones via multivariate models using near-infrared (NIR) spectroscopy data. Aust J Crop Sci. 2020; 14: 889–896. https://doi.org/10.21475/ajcs.20.14.06.p2099

94. Purcell DE, O'shea MG, Johnson RA, Kokot S. Near-infrared spectroscopy for the prediction of disease ratings for fiji leaf gall in sugarcane clones. Appl Spectrosc. 2009; 63: 450–457. https://doi.org/10.1366/000370209787944370 PMID: 19366512

95. Hu Y, Knapp S, Schmidhalter U. Advancing high-throughput phenotyping of wheat in early selection cycles. Remote Sens. 2020; 12: 1–10. https://doi.org/10.3390/rs12030574

96. Watanabe K, Guo W, Arai K, Takanashi H, Kajiya-Kanegae H, Kobayashi M, et al. High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. Front Plant Sci. 2017; 8: 1–11. https://doi.org/10.3389/fpls.2017.00001 PMID: 28220127

97. Rincent R, Charpentier JP, Faivre-Rampant P, Paux E, Le Gouis J, Bastien C, et al. Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. G3 Genes, Genomes, Genet. 2018; 8: 3961–3972. https://doi.org/10.1534/g3.118.200760 PMID: 30373914

98. Krause MR, González-Pérez L, Crossa J, Pérez-Rodríguez P, Montesinos-López O, Singh RP, et al. Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. G3 Genes, Genomes, Genet. 2019; 9: 1231–1247. https://doi.org/10.1534/g3.118.200856 PMID: 30796086

99. Teixeira Dos Santos CA, Lopo M, Páscoa RNMJ, Lopes JA. A review on the applications of portable near-infrared spectrometers in the agro-food industry. Appl Spectrosc. 2013; 67: 1215–1233. https://doi.org/10.1366/13-07228 PMID: 24160873

100. Decker SR, Harman-Ware AE, Happs RM, Wolfrum EJ, Tuskan GA, Kainer D, et al. High Throughput Screening Technologies in Biomass Characterization. Front Energy Res. 2018; 6: 1–18. https://doi.org/10.3389/fenrg.2018.00120

101. Cheavegatti-Gianotto A, de Abreu HMC, Arruda P, Bespalhok Filho JC, 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

102. Matsuoka S, Ferro J, Arruda P. The Brazilian Experience of Sugarcane Ethanol Industry The Brazilian experience of sugarcane ethanol industry. Vitr Cell Dev Biol. 2009; 45: 372–381. https://doi.org/10.1007/978-1-4419-7145-6

103. Brasileiro BP, Mendes TO de P, Peternelli LA, da Silveira LCI, de Resende MDV, Barbosa MHP. Simulated individual best linear unbiased prediction versus mass selection in sugarcane families. Crop Sci. 2016; 56: 570–575. https://doi.org/10.2135/cropsci2015.03.0199

104. Balsalobre TWA, Mancini MC, Pereira G da S, 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. https://doi.org/10.2134/agronj2015.0430

105. Yadav S, Jackson P, Wei X, Ross EM, Aitken K, Deomano E, et al. Accelerating genetic gain in sugarcane breeding using genomic selection. Agronomy. 2020; 10: 1–21. https://doi.org/10.3390/agronomy10040585

106. Racedo J, Gutiérrez 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. https://doi.org/10.1186/s12870-015-0696-x PMID: 26759170

107. Voss-Fels KP, Cooper M, Hayes BJ. Accelerating crop genetic gains with genomic selection. Theor Appl Genet. 2018; 132: 669–686. https://doi.org/10.1007/s00122-018-3270-8 PMID: 30569365

108. Barbosa MHP, da Silveira LCI. Breeding Program and Cultivar Recommendations. Sugarcane: Agricultural Production, Bioenergy and Ethanol. 2015. https://doi.org/10.1016/B978-0-12-802239-9.00011–6

109. Barbosa MHP, Resende MDV De, Peternelli LA, Bressiani JA, Silveira LCI Da, Silva FL, et al. Use of REML/BLUP for the selection of sugarcane families specialized in biomass production. Crop Breed Appl Biotechnol. 2004; 4: 218–226.

110. de Bem Oliveira I, Resende MFR, Ferrão LF V., Amadeu RR, Endelman JB, Kirst M, et al. Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. G3 Genes, Genomes, Genet. 2019; 9: 1189–1198. https://doi.org/10.1534/g3.119.400059 PMID: 30782769

111. Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, et al. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity (Edinb). 2014; 112: 48–60. https://doi.org/10.1038/hdy.2013.16 PMID: 23572121

112. Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, et al. Quantitative genetics and genomics converge to accelerate forest tree breeding. Front Plant Sci. 2018; 871: 1–10. https://doi.org/10.3389/fpls.2018.01693 PMID: 30524463

113. Juliana P, Singh RP, Poland J, Mondal S, Crossa J, Montesinos-López OA, et al. Prospects and Challenges of Applied Genomic Selection—A New Paradigm in Breeding for Grain Yield in Bread Wheat. Plant Genome. 2018; 11: 1–17. https://doi.org/10.3835/plantgenome2018.03.0017 PMID: 30512048

114. Edriss V, Gao Y, Zhang X, Jumbo MB, Makumbi D, Olsen MS, et al. Genomic prediction in a large African maize population. Crop Sci. 2017; 57: 2361–2371. https://doi.org/10.2135/cropsci2016.08.0715

115. Rutkoski J, Poland J, Mondal S, Autrique E, Pérez LG, Crossa J, et al. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. G3 Genes, Genomes, Genet. 2016; 6: 2799–2808. https://doi.org/10.1534/g3.116.032888 PMID: 27402362