A first look at the ability to use genomic prediction for improving the ratooning ability of sugarcane

Md. Sariful Islam[1]*, Keo Corak[2], Per McCord[1,3], Amanda M. Hulse-Kemp[2,4] and Alexander E. Lipka[5]

[1]Sugarcane Field Station, USDA-ARS, Canal Point, FL, United States, [2]Genomics and Bioinformatics Research Unit, USDA-ARS, Raleigh, NC, United States, [3]Irrigated Agriculture Research and Extension Center, Washington State University, Prosser, WA, United States, [4]Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC, United States, [5]Department of Crop Sciences, University of Illinois, Urbana-Champaign, IL, United States

The sugarcane ratooning ability (RA) is the most important target trait for breeders seeking to enhance the profitability of sugarcane production by reducing the planting cost. Understanding the genetics governing the RA could help breeders by identifying molecular markers that could be used for genomics-assisted breeding (GAB). A replicated field trial was conducted for three crop cycles (plant cane, first ratoon, and second ratoon) using 432 sugarcane clones and used for conducting genome-wide association and genomic prediction of five sugar and yield component traits of the RA. The RA traits for economic index (EI), stalk population (SP), stalk weight (SW), tonns of cane per hectare (TCH), and tonns of sucrose per hectare (TSH) were estimated from the yield and sugar data. A total of six putative quantitative trait loci and eight nonredundant single-nucleotide polymorphism (SNP) markers were associated with all five tested RA traits and appear to be unique. Seven putative candidate genes were colocated with significant SNPs associated with the five RA traits. The genomic prediction accuracies for those tested traits were moderate and ranged from 0.21 to 0.36. However, the models fitting fixed effects for the most significant associated markers for each respective trait did not give any advantages over the standard models without fixed effects. As a result of this study, more robust markers could be used in the future for clone selection in sugarcane, potentially helping resolve the genetic control of the RA in sugarcane.

1 Introduction

Sugarcane (*Saccharum* sp. hybrid) production is vital to the agricultural economy of tropical and subtropical regions, where 86% of the world's table sugar is produced (FAO, 2020). Sugarcane is also used in the production of bioethanol, energy from the bagasse (sugar production by-product), and paper products. Sugarcane is commonly managed through an agronomic practice known as ratooning in which growers plant the sugarcane once and then harvest the aboveground biomass several times, allowing the plant to recover between harvests. The first-year crop grown from the seed cane after planting is called the plant cane (PC) crop. The subsequent regrowth from the stubble after harvesting is called the ratoon (RT) crop. RT crops can be collected multiple times. In Florida, during the 2020 growing season, 69.8% of the sugarcane cultivation area was under RT crops composed of the first ratoon (FR; 30.5%), second ratoon (SR; 29.5%), third ratoon (7.6%), and fourth ratoon (2.3%) (VanWeelden et al., 2020). Depending on the growing region, the proportion of RT crops ranges from 50% to 75% of the total sugarcane cultivated area in other countries (Xu et al., 2021). Like in every other sector, agricultural producers are struggling with the high cost of labor and associated labor shortages all over the world. This problem is especially acute in western countries like the United States. Ratooning saves growers' money on cultivation costs, and it boosts industry profits because RT crops mature earlier, produce better juice quality, and have better sugar recovery (Xu et al., 2021). Thus, by increasing the proportion of acreage in ratoon crops, sugarcane industries can save on labor costs, increase sugar quality, and decrease soil disturbance by skipping planting operations every year.

Unfortunately, sugarcane yield tends to decline after the first PC crop (Xu et al., 2021). Therefore, sugarcane breeders are currently selecting for a trait they have described as the ratooning ability (RA) in addition to other desired traits in their breeding objectives to make the sugarcane industry more profitable. The RA has various definitions; the most common definition expresses the performance of RT crops (the first ratoon or the average of all ratoons) as a percentage relative to the PC crop as described (Equation 3) below (Milligan et al., 1996; Coto Arbelo et al., 2021). Industry stakeholders desire varietals that can be grown over an increased number of ratoon crops while maintaining their yield. Performance is determined by several morphological and yield traits including root growth of stubble, germination rate after harvesting, tillering rate, stalk number in previous crop, and cane yield, which is a product of stalk population (SP) and stalk weight (SW) (Milligan et al., 1990). Thus, selecting on those traits will generally improve the RA. However, the RA of sugarcane is not determined just by genetics but by a combination of factors including environment (soil, temperature, humidity, and water supply), cultivation technology (Xu et al., 2021), and the interactions between them. Several studies have been reported related to the RA in context of the environment, biotic and abiotic stress, and management practices (Singh et al., 2006; Shukla et al., 2008; Gomathi et al., 2013; Ramburan et al., 2013; Bagyalakshmi et al., 2019; Gao et al., 2019).

Very little is known about the genetics and genetic architecture of RA in sugarcane (Milligan et al., 1996; Zhou and Shoko, 2012; Coto Arbelo et al., 2021). The most cited report on the RA studied

the heritability, correlation between PC and RT crop, and genetic gain in RT crop of six sugar- and yield-related traits using 37 genotypes (Milligan et al., 1996). Another study suggested that the high-yielding genotypes with slightly lower RAs might be viable for commercial production when planted in short ratooning cycles, since higher SP seems to diminish the yield loss in ratoon crop (Zhou and Shoko, 2012). Very recently, although a narrow RA diversity was detected among the 39 tested clones, three checks using the same strategy recommended two clones for commercial release having high yield and good RA potential in a multilocation trial (Coto Arbelo et al., 2021). To the best of our knowledge, no report has been published related to the identification of quantitative trait loci (QTLs) or marker–trait associations for the RA in sugarcane. Investigating the genetic and molecular basis of the RA using a genome-wide association study (GWAS) in sugarcane will assist breeders in better understanding the genetic basis for potentially improving the RA of sugarcane cultivars.

GWASs are widely used to identify genes, mutations, and putative functional markers that are responsible for complex quantitative traits through forward genetics analysis (Vuong et al., 2015; Islam et al., 2016). In comparison with traditional QTL analyses, GWAS has several advantages, including using materials with increased diversity and better downstream utility across diverse germplasms with a shorter development time for new populations (Abdurakhmonov and Abdukarimov, 2008). Thus, GWAS has been successfully utilized in many plants for identifying significant marker–trait associations and putative candidate genes, such as cotton (Islam et al., 2016; Thyssen et al., 2018; Wubben et al., 2019), maize (Pace et al., 2015), barley (Matthies et al., 2014), wheat (Tadesse et al., 2015), sorghum (Morris et al., 2013), and soybean (Vuong et al., 2015). However, many genetic studies using GWAS have been reported in sugarcane (Debibakas et al., 2014; Racedo et al., 2016; Singh et al., 2016; Barreto et al., 2019; Yang et al., 2019a; Yang et al., 2019b; Yang et al., 2020; Pimenta et al., 2021; Chen et al., 2022). To the best of our knowledge, there have been no QTL and/or GWASs related to the RA trait in sugarcane reported so far. Hence, conducting a GWAS to detect markers associated with the RA in sugarcane would be an important contribution to the sugarcane-breeding community. Discovered significant genetic loci could be incorporated into a breeding pipeline, but marker-assisted selection (MAS) approaches have several limitations, for example, many traits are polygenic in nature with small QTL effects and large genotype-by-environment interactions (Xu and Crouch, 2008).

In addition to the insights gained from GWAS approaches, breeders are interested in using genome-wide genotypic data to estimate the genomic estimated breeding value (GEBV) of testing individuals (Meuwissen, 2009) and to use these GEBVs for selection. Numerous animal- and plant-breeding programs have already used genomic selection (GS) successfully since its introduction. Unlike MAS, which typically only considers major effect loci, the use of genome-wide markers allows GS to account for both major and minor effect loci (Gouy et al., 2013). A breeding cycle could be shortened if individuals were selected based on their GEBV in the early phases of the breeding cycle (Jannink et al., 2010). In sugarcane specifically, predicting the RA could allow

breeders to potentially shorten the breeding cycle by reducing ratoon crop cycles during selection.

The utility of GS has increased as new bioinformatics tools and next-generation sequencing have been developed. Thus, GS has been evaluated in many crops (Arruda et al., 2015; Bernal-Vasquez et al., 2017; Gezan et al., 2017; Islam et al., 2020) including 11 studies in sugarcane (Gouy et al., 2013; Olatoye et al., 2019; Deomano et al., 2020; Hayes et al., 2021; Islam et al., 2021; Voss-Fels et al., 2021; Yadav et al., 2021; Aono et al., 2022; Batista et al., 2022; Islam et al., 2022a; O'Connell et al., 2022). However, none of the studies attempted to predict the RA in sugarcane. Hence, it is worthwhile to conduct research on evaluating the genomic prediction accuracy of the RA in sugarcane. It is reported that incorporating major gene or QTL in the GS models as fixed covariates could increase the prediction accuracy most of the time (Bernardo, 2014; Islam et al., 2022a). However, it is not true all the time: it has also been reported that prediction accuracy was decreased and model bias was increased by incorporating the GWAS peak marker as a fixed effect in the GS models depending on the genetic architecture of the trait (Rice and Lipka, 2019; Billings et al., 2022).

We have previously reported two GS studies related to disease resistance, sugar, and yield-related traits using the same population and the same sets of markers as in the current study (Islam et al., 2021; Islam et al., 2022a). The objective of this study is to uncover the genetic basis of the RA and detect the associated markers with five RA traits in sugarcane. We also evaluated four GS models for checking the feasibility of GS in the RA trait in sugarcane using the peak GWAS marker as a fixed covariate in the tested models. This research contributes to understanding the genetic basis of the RA traits, developing more robust trait-associated markers, along with providing information about using the GS approach for sugarcane clone selection and downstream cultivar release to growers.

2 Materials and methods

2.1 Plant material and field trial

Four hundred fourteen sugarcane clones from the second clonal stage of the USDA-ARS Sugarcane Breeding Program at Canal Point, Florida, program and 18 commercial cultivars from the USDA-ARS Sugarcane Breeding Program in Houma, Louisiana, were grown in a replicated field trial (Supplementary Table S1) along with two checks, "CP 00-1101" (Gilbert et al., 2008) and "CP 96-1252" (Edmé et al., 2005). Each tested clone was replicated twice in an augmented row–column experimental design, and each check was repeated 17 times. The details and layout of the field plots are described in our previous studies (Islam et al., 2021; Islam et al., 2022a). In brief, the field trial was established at the USDA-ARS Sugarcane Field Station, Canal Point, Florida, in November 2016. Plots consist of one row (4.6 m in length) with 1.5-m spacing between plots. Each row had 25 plots with an alley between two rows of 6 m between each row. The field was divided into 36 rows. Three crop cycles were evaluated: PC, FR, and SR. The standard protocol was followed for all necessary management practices throughout the trial.

2.2 Data collection and trait measurement

The details of data collection of yield and sugar traits were described in our previous reports (Islam et al., 2022a). SP, SW, total dissolved solids (Brix), juice polarization (Pol), and fiber content data were collected from the field plot directly for PC, FR, and SR crop cycles. In September 2017, 2018, and 2019, the millable stalks per hectare (SP) for each plot were estimated from the number of stalks per plot counted manually for PC, FR, and SR, respectively. We harvested 10 stalks from each plot randomly, removed the top just below the apical meristem, and bundled each plot in February 2018, 2019, and 2020 for PC, FR, and SR, respectively. Those harvested bundles were utilized to collect SW (kg stalk^{-1}), total fiber content (%), Pol (%), Brix (%), and moisture content (%). Sugar content (SC) was then calculated from the corrected Brix and Pol (Legendre, 1992) using the following formula (Islam et al., 2022b):

$$SC \quad (\%) = \frac{Pol \times 26}{[105.811 + (Brix - 15) \times 0.444]} \qquad (1)$$

Cane yield (Mg ha^{-1}) in the form of tonns of cane per hectare (TCH) was calculated as the product of SW (kg stalk^{-1}) and SP (stalks ha^{-1}) and divided by 1,000. Theoretical recoverable sucrose (TRS) was calculated from the juice data and fiber concentration to estimate the sugar yield (Legendre, 1992). All values of TRS were multiplied by a correction factor of 0.86 to approximate the commercial recoverable sugar (CRS; kg ha^{-1}) as suggested by Islam et al. (2022b). Sucrose yield (Mg ha^{-1}) as tonns of sucrose per hectare (TSH) was estimated as:

$$Sucrose \quad yield \quad = \quad (TCH \times CRS) \div 1000 \qquad (2)$$

Following Deren et al. (1995), the economic index (EI) was calculated from the cane yield, sucrose yield, and costs of harvesting, hauling, and milling the cane in Florida. The RA (%) for EI, TCH, TSH, SP, and SW was then estimated as suggested by Coto Arbelo et al. (2021) and Milligan et al. (1996) using the following formula:

$$A = \frac{\sum_{n=1}^{N} R_n}{P * N} * 100 \qquad (3)$$

where A is the RA for a given trait expressed as a percent, R_n is the phenotypic value of the trait in RT crop n, P is the value of the trait in the PC crop, and N is the number of ratoon crops.

2.3 Genotyping and single-nucleotide polymorphism markers

DNA extraction, library preparation, sequencing, single-nucleotide polymorphism (SNP) calling, and filtering protocols were described in detail in our previous reports (Islam et al., 2021; Islam et al., 2022a). After extracting, DNA from young leaves was submitted to RAPiD Genomics LLC (Gainesville, FL, USA) for library preparation, sequencing, and initial bioinformatic analysis. Sequencing was done on an Illumina HiSeq 2 X 100 sequencer with the qualified processed samples combined in

equal amounts. The raw sequence data were demultiplexed using Illumina's bcl2fastq, and then reads were cleaned and trimmed using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Mosaik (Lee et al., 2014) was used to align clean reads against the *Sorghum bicolor* V3.1 reference genome (Paterson et al., 2009). Using FreeBayes (Garrison and Marth, 2012), the SNPs were called and then filtered based on read depth for each SNP ≥35 and minor allele frequency ≥2%. Because of the conversion, marker data were transformed into numerical format, where 0 was assigned to the reference allele at each locus, 2 to the alternate allele, and 1 to the heterozygote. After discarding SNPs with more than two alleles, 10,435 SNPs were remaining for further analysis. The SNP nomenclature was given starting with sorghum reference genome chromosome number followed by the location in the chromosome.

2.4 Variance component, heritability, and correlation

The best linear unbiased predictors (BLUPs) of the RA of the five tested traits for the sugarcane clones were analyzed using a single mixed linear model as described below:

$$y_{aikhj} = \mu + p_j + g_a + \gamma_j + r_{jh} + c_{jk} + \phi_i + \epsilon_{aikhj} \qquad (4)$$

where y_{aikhj} was the vector of phenotypic value of genotype a from population i (a column describing which entry is the tested clone or one of two checks (CP00-1101 and CP96-1252) evaluated from the hth row of the kth column from the jth replicate in trial, γj was the effect of jth complete replicate, r_{jh} was the effect of the hth row and the jth replicate, c_{jk} was the effect of kth column within the jth replicate, ϕ_i is the fixed effect of the ith population (a column describing which entry is the clone or check, g_a is the random effect term for the ath genotype with $g_a \sim N(0, \sigma_g^2)$ and σ_g^2 is the variance due to genotype or line effect, and ϵ_{aikhj} is the residual error with $\epsilon_{aikhj} \sim N(0, \quad R\sigma_\epsilon^2)$, **R** is a diagonal matrix accounting for heterogeneous error variance, and σ_ϵ^2 is the population variance of ϵ_{aikh}. The predict function in ASREML-R 4.0 (Butler et al., 2009) was used to obtain the BLUPs. The variance component estimates from a different model were used to estimate the broad sense heritability (H^2) following entry mean basis as $H^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{e/R}^2)$, where σ_g^2 and σ_e^2 reflect variance associated with genotype and error, respectively, while R indicates the number of replications. The Pearson correlation coefficients among the five tested RA traits and resulting heat map were conducted in GraphPad Prism 9 software (www.graphpad.com).

2.5 Relationship among the clones

An additive relationship matrix was constructed for the 432 clones in the genotypic dataset described above using the AGHmatrix package v2.0.4 in R (Amadeu et al., 2016). Genotypes were assumed to be decaploid (ploidy = 10), but pseudodiploid parameterization was used because the genotypic matrix had only three values (heterozygous, homozygous for reference, and homozygous for alternate). The resulting matrix was centered,

and then singular value decomposition was performed. Principal components were extracted and plotted using the plotly graphing package in R (Li and Bilal, 2021).

2.6 Genome-wide association study

Multivariate mixed-linear models for the five traits of interest were fit using an Average Information Restricted Maximum Likelihood (REML) algorithm (Gilmour et al., 1995) implemented in ASReml-R v 4.1.0.160 (Butler et al., 2009). The multivariate models took the form:

$$y = Xb + Zu + e \qquad (5)$$

where y was a vector of phenotypes from K cropping cycles that is partitioned into $y' = [y'_1, y'_2, y'_3]$ and each vector y'_k has length n_k where n_k was the number of individuals phenotyped in cropping cycle k; $X = \bigoplus_{k=1}^{K} \quad X_k$, where X_k was an incidence vector for each cropping cycle of length n_k; $b' = [b_1, b_2, b_3]'$ was a vector of fixed effects for each cropping cycle; $Z = \bigoplus_{k=1}^{K} \quad Z_k$, where Z_k was an $n_k xc$ incidence matrix of unique clones c in cropping cycle k. and $u' = [u'_1, u'_2, u'_3]$ was an matrix of random clone effects and each vector u'_k has length c; and $e' = [e'_1, e'_2, e'_3]$ was a vector of residuals where each vector e'_k has length n_k. Clone effects were assumed to follow a multivariate normal distribution with mean zero and an unstructured variance–covariance matrix equal to $U \otimes I_c$ where I_c was a $C x C$ identity matrix, U was a K x K matrix in which cropping cycle variances of clone effects σ_{uk}^2 on the diagonal and $\sigma_{u_{kk'}}^2$, the covariance between clone effects in different cropping cycles was on the off-diagonal. Finally, residuals were assumed to be normally distributed with mean zero and correlated among cropping cycled with variance equal to $E \otimes I_k$ where I_k was a $K x K$ identity matrix and E was a K x K matrix with residual variance within a cropping cycle σ_{ek}^2 on the diagonal and residual covariance between cropping cycles $\sigma_{e_{kk'}}^2$ on the off-diagonal. BLUPs for each clone in each cropping cycle were extracted from the model's predicted values, and then the RA for each trait was calculated using the equation from Coto Arbelo et al. (2021) as previously described.

GWAS was performed using the R packages GWASpoly v2.10 (Rosyara et al., 2016) for each trait using the RA metric described above using the BLUPs calculated from Equation 5 and the same genomic dataset used throughout the study. A diploid status was used due to insufficient sequence read depth to estimate allele dosage. The polygenic covariance matrix was computed for each chromosome according to the set.K function in GWASpoly. The maximum genotype frequency was set as 0.988 (1-5/N where N = 432). Both additive and dominance models were tested for each trait. The significance thresholds for each model and trait combination were determined for a false discovery rate (FDR) of 0.1 following the method in Storey (2002). Manhattan and qqplots plots were generated using the qqman v0.1.8 package in R. The LD contour plots for 14 most significant associated SNPs (p ≤ 0.0001) was created from genotypic data of the 432 clones of sugarcane population using JMP genomics 10.0 software and physically aligned to the sorghum reference genome.

2.7 Identification of candidate genes

Sorghum reference genome was used, since it is the closest diploid organism to sugarcane. Genes physically located nearby the significant associated markers with traits were investigated for their possible annotation and function in the literature. The tag sequences containing SNPs associated with traits of interest were also BLASTed against the sugarcane reference genome "R570" using the genomic resource provided at https://sugarcane-genome.cirad.fr/ and a default e-value of 1e-10.

2.8 Genomic selection

For each trait, four different genomic prediction models were evaluated *via* 5-fold cross-validation repeated 25 times. The 5-fold validation was completed by randomly selecting four-fifths of the individuals for the training and the remaining fifth as the validation population. Iteration number was used to set the seed for randomization of genotypes into the test and training sets to ensure the same training set was used to develop all four models. Details of the four models tested, namely, the Ridge regression BLUP [RR-BLUP (Endelman, 2011)], Additive-dominance-epistasis [ADE (Covarrubias-Pazaran, 2016)], Reproducing kernel Hilbert space [RKHS (Gianola and Van Kaam, 2008)), and Bayes A (Perez and De los Campos, 2014), can be found in our previous publication (Islam et al., 2022a). Several evaluation criteria were considered to assess prediction accuracy and model bias: prediction accuracy (Pearson's correlation coefficient r between predicted and observed values), coincidence index (CI) (the proportion of genotypes in the top 20% of observed values that were also in the top 20% of predicted values), and the slope and intercept from a regression of the predicted values on the observed values. For a perfectly unbiased model, the slope and intercept values would be 1 and 0, respectively.

To test the effect of including highly significant GWAS hits as fixed effects in prediction models, the scheme outlined above was repeated for each trait with a slight modification: before fitting each prediction model, a GWAS model testing only additive effects was performed using just the training population from each iteration. GWAS was otherwise performed using the same parameters as described above. The most significant hit was extracted from the GWAS results and the vector corresponding to that SNP from the genotypic marker matrix was included as a fixed effect in the prediction model.

For each trait and model, evaluation metric results for all iterations were summarized in box and whisker plots generated using ggplot v3.3.4.

3 Results

3.1 Distribution, heritability, and correlation

The RA of sugarcane [RA (%)] for five traits, EI, SP, SW, TCH, and TSH, was estimated. The distribution, mean values, ranges, and variance components of these values are included in Figure 1 and Table 1. The RA for SP appears to be highest over PC performance (average 104.45%, range 23.8%–216.8%), followed by TCH (average 67.5%, range 5.4%–146.3%). Likewise, the genetic variance for SP was highest (477.94) followed by TCH (224.14). The broad-sense heritabilities (H^2) calculated from variance components were moderate to high (Table 1). Once again, the highest heritability for the RA was observed in SP (0.71) followed by TSH (0.66), TCH (0.65), EI (0.62), and SW (0.58).

Highly significant correlations were observed among the five tested RA traits in sugarcane (Figure 2). As expected, the highest positive correlation coefficient (0.99) was observed between EI and TSH, while the lowest negative correlation coefficient (-0.13) was associated between SP and SW. The RA of SP has a higher correlation coefficient value than that of SW with the other three tested traits (TCH, TSH, and EI).

3.2 Population structure

A principal component biplot was created using the SNP data to examine the population structure among the tested clones (Figure 3). Based on the pseudo-diploidized SNP data, we found little evidence for population structure among the tested clones; the first two principal components explained 1.44% and 1.19% of the total variation, respectively. The tested clones did separate into two groups based on first principal components. One group fell in both the third and fourth quadrants and the majority of the tested CP clones and almost all of the Louisiana clones were placed in this cluster.

3.3 Marker−trait association

The RA of the five traits (EI, SP, SW, TCH, and TSH) were subjected to GWAS analysis and results of marker−trait associations were included in Figure 4, Table 2, and Supplementary Table S2.
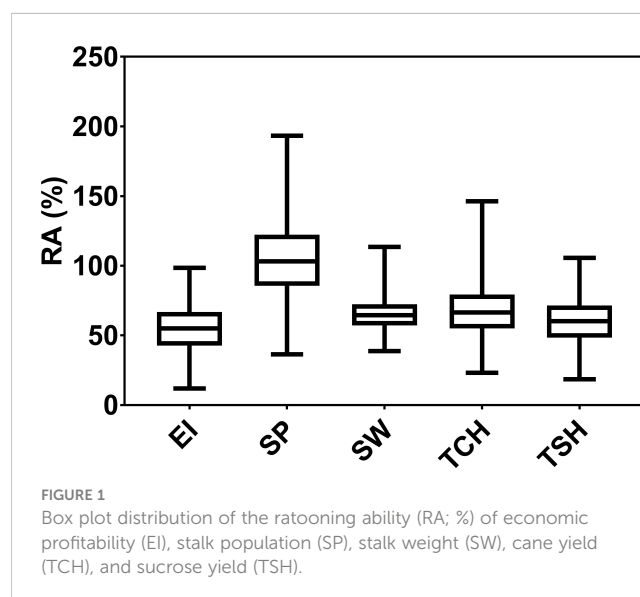

FIGURE 1
Box plot distribution of the ratooning ability (RA; %) of economic profitability (EI), stalk population (SP), stalk weight (SW), cane yield (TCH), and sucrose yield (TSH).

TABLE 1 The mean, range, variance components, and broad-sense heritability (H²) of the ratooning ability (RA; %) of stalk weight (SW), stalk population (SP), cane yield (TCH), sucrose yield (TSH), and economic profitability (EI).

Traits	BLUPs mean	Mean	Max	Min	Genetic variance ($\hat{\sigma}_g^2$)	Residual ($\hat{\sigma}_e^2$)	Heritability (H²)
EI	55.48	55.38	99.50	9.83	172.69	210.40	0.62
SP	104.47	104.48	216.84	23.82	477.94	396.11	0.71
SW	65.25	65.13	131.14	28.44	78.04	113.99	0.58
TCH	67.67	67.49	146.32	15.38	224.14	242.30	0.65
TSH	60.67	60.47	105.64	14.90	168.40	176.49	0.66

A mixed linear model estimated the best linear unbiased prediction (BLUP), variance components, and heritability (entry mean basis).

Three GWAS models (additive, 1-dominant reference, and 1-dominant alternative) were explored, and FDR-adjusted p-values were determined for each SNP. Adjusted p-values greater than 0.001 are reported in Supplementary Table S2 with an indication of whether they are statistically significant according to a genome-wide FDR of 0.1.

Since many observed p-values deviated from the uniform distribution of expected and observed probability of obtaining an association with the respective traits, the quantile-quantile (Q-Q) plots demonstrated a potential for false-positive associations. Results revealed that most of the observed p-values follow a uniform distribution, but the few in linkage disequilibrium (LD) with a causal polymorphism have significant p-values in the tail (Supplementary Figure S1). A total of 14 significant marker–trait associations were detected for the RA of the five tested traits (Table 2). Of the 14 associations, two, four, five, and three were linked with traits EI, SW, TCH, and TSH, respectively. No

significant marker–trait association was found for SP. There were eight nonredundant SNPs associated with all tested traits. One SNP (S10_53811870), located on sorghum chromosome 10, was commonly associated with three traits, EI, TCH, and TSH. Two SNPs located on sorghum chromosomes 5 and 7 (S05_10359269 and S07_59440142) were significantly associated with two traits: TCH, TSH and EI, TSH, respectively. One SNP, S03_8476715, located on sorghum chromosome 3, was significantly associated with SW for all three models tested.

A single putative QTL was considered when multiple loci associated with a trait were within a 5-Mb interval. The QTL nomenclature was according to McCouch et al. (1997). A total of six QTLs associated with the RA of the four tested traits (EI, SW, TCH, and TSH) were identified (Table 2; Supplementary Table S3). The highest number (three) of QTL was associated with the RA of TCH and located on three different chromosomes. The extremely significant (p value ≤ 0.000001) loci (*qRATCH-cS10*) associated
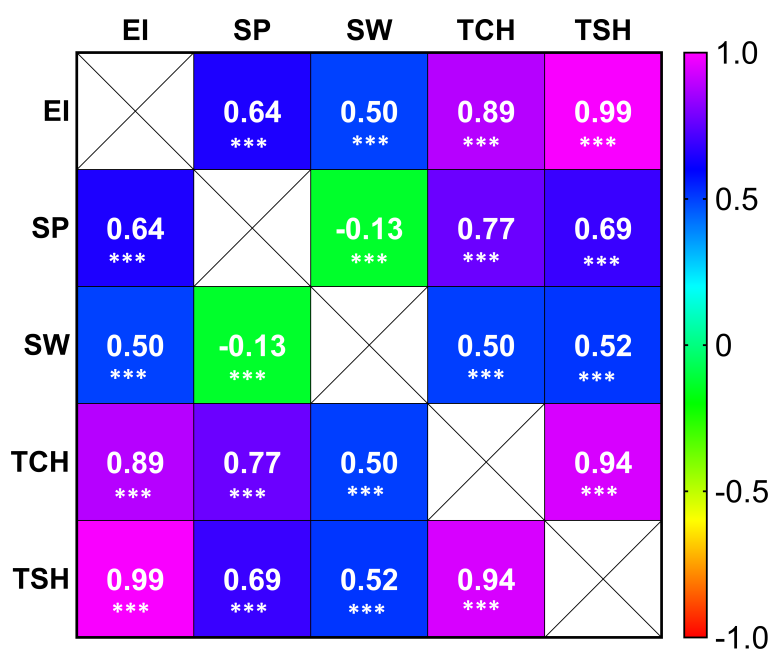

FIGURE 2
Correlation coefficient heat map among the tested traits [ratooning ability (RA; %) of stalk weight (SW), stalk population (SP), cane yield (TCH), sucrose yield (TSH), and economic profitability (EI)]. The three asterisks (***) denote the correlation coefficient (R) significant at p ≤ 0.001.
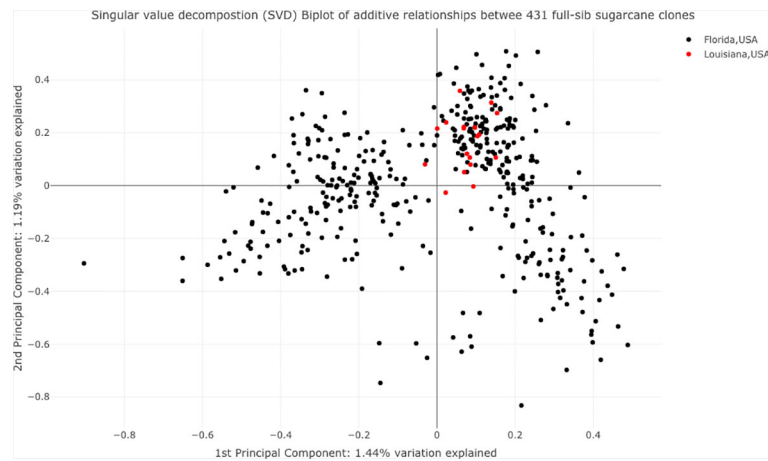
FIGURE 3
Principle component analysis of 432 tested sugarcane clones. The black and red dots represent clones belonging to the sugarcane breeding programs at Canal Point, Florida, and Houma, Louisiana, USA, respectively.

with the RA of TCH comprises three SNPs (*S10_51068605, S10_53811870,* and *S10_55391413*) covering 4.3 Mb on chromosome 10. The SNP (*S10_53811870*) with the strongest association with the RA of TCH (p value = 8.55E-6) is located at position 53,811,870 bp on chromosome 10.

3.4 Candidate genes

SNPs significantly associated with the RA of sugarcane were aligned with the sorghum genome, and those SNPs (7) with closely located genes (<20 kb from the SNP) were included in Table 3 and
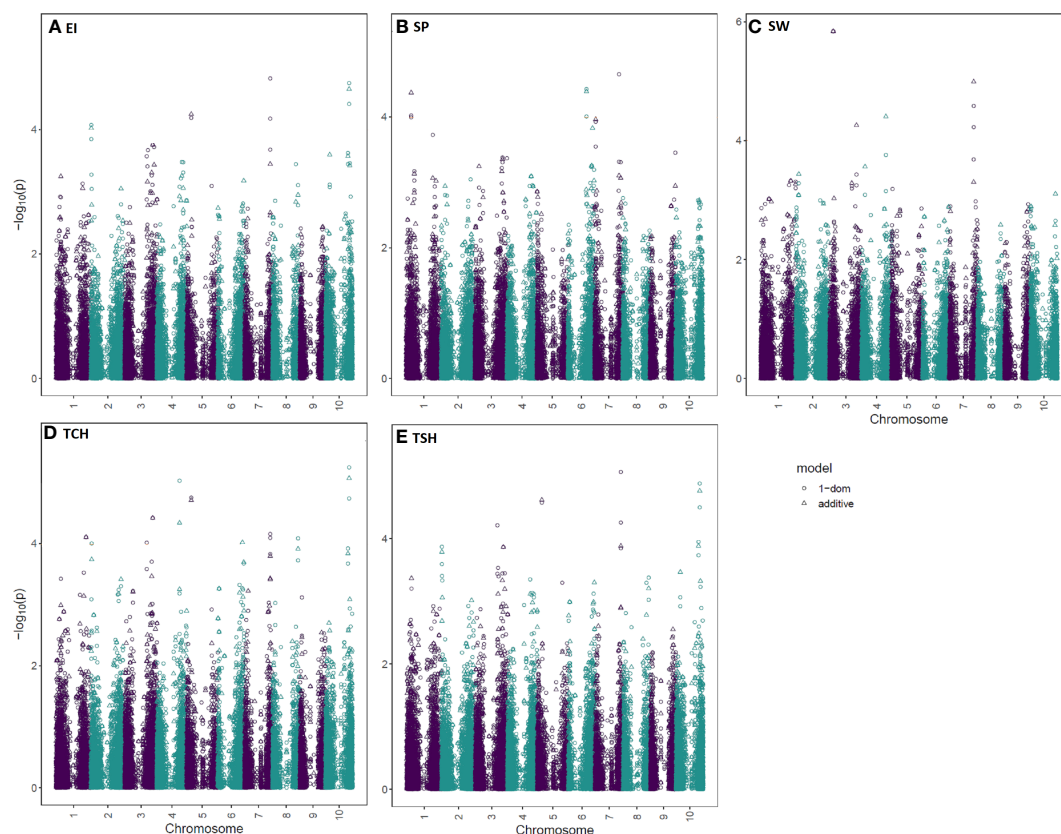

FIGURE 4
Manhattan plots for the sugarcane ratooning ability (RA; %) of economic profitability (EI), stalk population (SP), stalk weight (SW), cane yield (TCH), and sucrose yield (TSH) using two models (additive and 1-dominant). The negative $\log_{10}$ converted p-values were plotted against the marker positions on the 10 chromosomes of the sorghum reference genome.

Supplementary Table S4. We also BLASTed the tag sequences of those seven SNPs against the monoploid sugarcane "R570" genome (Garsmeur et al., 2018) using an e-value cutoff of 1e-10, and associated homologous genes were listed in Supplementary Table S4. Gene annotation of the sorghum genome suggested that seven putative genes were colocated with those significant SNPs associated with the five RA traits in sugarcane. The high proportion of genes located 0.0 kb from significant SNPs indicates that many SNPs occurred within the coding sequence of genes. Out of seven genes, the annotation of five genes were found to have homologs among *Arabidopsis*, rice, and "R570" (sugarcane monoploid genome) gene families and function. The LD contour plot near the most significant SNP (*S03_8476715*) associated with the RA of SW showed that there are several small LD blocks within a large one, indicated by red color in the plot and most of the SNPs have linked each other as indicated by the strong LD (Figure 5). This SNP is colocated with gene Sobic003G096000 in the sorghum reference genome, which is functionally annotated as aldolase superfamily protein. The LD contour plots for the other 14 significant SNPs also exhibited similar results (Supplementary Figure S2).

3.5 Genomic selection prediction

The prediction accuracy for the five RA traits was evaluated using four GS methods with and without highly significant SNPs as fixed effects (total eight models), and four metrics were used to evaluate the GS models (prediction accuracy, CI, slope, and intercept). Results of genomic prediction and CI for the RA of EI, SP, SW, TCH, and TSH are presented in Figure 6. The overall prediction accuracies attained for EI were greater than that of the other tested traits for the RA. The highest and lowest accuracies (0.36 and 0.21) were observed for EI with the ADE model without fixed effects and SW with RR-BLUP model with fixed effect, respectively. In general, the standard genomic prediction models that did not include significant markers as fixed-effect covariates yielded the highest prediction accuracies; however, the differences in prediction accuracies between models were slight and not statistically significant in any case. Slope and intercept of the regression lines were used to measure the prediction bias of the models. By comparing between the slope and intercept distribution for standard models and models, the results showed that models

TABLE 2 Summary of marker−trait associations and quantitative trait loci (QTLs) detected at different threshold levels of p-values for the ratooning ability of five traits in sugarcane.

Trait[a]	Marker−trait association				QTL[b]			
	p ≤ 0.001	p ≤ 0.0001	Total	# Chr.	p ≤ 0.001	p ≤ 0.0001	Total	# Chr.
EI	17	5	22	8	3	0	3	3
SP	15	3	18	5	2	0	2	2
SW	14	4	18	6	1	0	1	1
TCH	24	11	35	9	5	1	5	5
TSH	17	5	22	9	2	0	2	2
Total	87[c]	28[c]	115[c]	9[c]	13	1	13	5[c]

[a] Traits are economic index (EI), stalk population (SP), stalk weight (SW), cane yield (TCH), and sugar yield (TSH).
[b] Marker loci mapped within 5-Mb intervals in the same chromosome were considered as a putative QTL.
[c] Some marker loci were associated with more than one trait.

TABLE 3 Candidate genes associated with significant single-nucleotide polymorphisms (SNPs; p ≤ 0.0001) linked with the ratooning ability (RA; %) of economic index (EI), stalk population (SP), stalk weight (SW), cane yield (TCH), and sugar yield (TSH) along with their annotation from the sorghum reference genome.

| SNP	Chr.	Candidate gene (Sobic. #)	Distance from SNP (kb)	Gene annotation	Trait
S03_64270241	3	003G314700	0.0	Vps51/Vps67 family (components of vesicular transport) protein	EI, TCH, TSH
S03_8476715	3	003G096000	0.0	Aldolase superfamily protein	SW
S04_51880407	4	004G168200	0.0	heat shock protein 101	TCH, TSH
S05_10359269	5	005G079201	4.8	F-box/RNI-like superfamily protein	EI, TCH, TSH
S07_56501497	7	007G137300	3.7	Homeodomain-like superfamily protein	SW
S07_59440142	7	007G160050	0.0	dicarboxylate carrier 2	EI, TCH, TSH
S10_53811870	10	010G195000	0.0	SEC7-like guanine nucleotide exchange family protein	EI, TCH, TSH
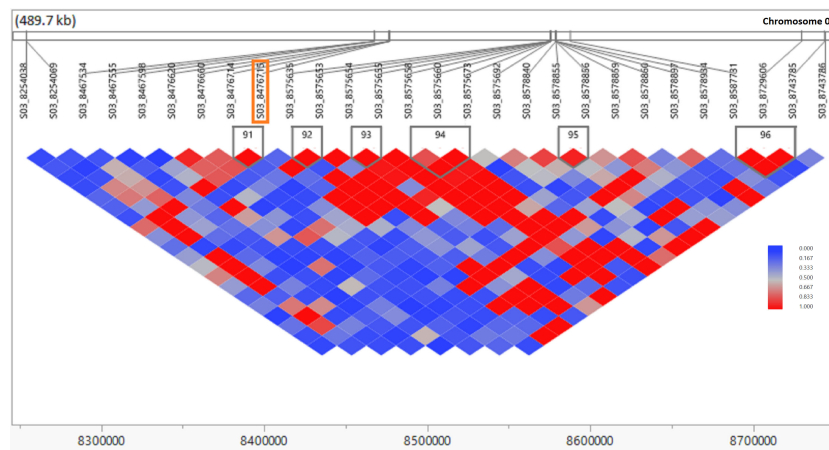
FIGURE 5
Linkage disequilibrium (LD) observed in the genomic region of the most significant SNP (p ≤ 0.00001) associated with ratooning ability of stalk weight (SW) in sugarcane. The LD contour plot is physically located in the sorghum genomic region of 8.2−8.8 Mb on chromosome 03. The LD contour was created from genotypic data of the 432 clones of the sugarcane population using JMP genomics 10.0 software. The X axis is physical distance in Mb and r^2 (CorrCoeff) between marker pairs is shown in different color blocks as per legend. The most significant SNP (S03_8476715) is indicated inside the orange box.

with fixed effects were severely deviated from the expected values of 1 and 0, respectively (Supplementary Figure S3).

The name and frequency of SNPs utilized as fixed effect in the four GS models for the five RA traits were included in Figure 7. The SNP *S03_8476715* located on sorghum chromosome 3 was used most frequent 60 times as fixed effect during predicting the RA of SW followed by *S10_53811870* SNP located on chromosome 10 was used as fixed effect for predicting the GEBV of three (EI, TCH, and TSH) RA traits. Both the SNPs showed significant association during the GWAS analysis using GWASpoly in this study.

4 Discussion

The profitability of sugarcane production is highly dependent on the RA of sugarcane cultivation. A genomic analysis of the sugarcane RA for the five traits was conducted through a combined GWAS and GS approach. GWAS results discovered a total of 14 significant marker–trait associations. A total of seven putative candidate genes associated with the RA in sugarcane were identified. Thus, the findings of the genomic analysis conducted in this study provide new information and a potential avenue for improving the RA traits in sugarcane through molecular breeding.

Worldwide, the sugarcane industry is facing increased planting costs and shortages of laborers. Thus, improvement of the RA is one of the most important target traits in breeding because replanting costs can be reduced by planting varieties that have a high RA, thereby increasing profitability. There are three main methods that have been reported to calculate the RA in sugarcane (Dunckelman, 1982; Milligan et al., 1996; Zhou and Shoko, 2012). Although those methods have some differences and performed differently, we measured the RA following the methods described by Milligan et al. (1996) and Coto Arbelo et al. (2021), since it is easiest to

calculate. The box plot distribution of phenotypic data of all five RA traits was continuous (Figure 1) and normal, signifying that several QTLs likely control the inheritance of those complex traits. Although the traits that contribute to the RA are very complex, broad-sense heritability estimates were moderate to high for each trait studied, which is congruent with previous studies (Todd et al., 2017; Abu-Ellail et al., 2019). As expected, the RA for SW was negatively correlated with the SP. This is because clones with heavier stalks tend to produce fewer tillers in the ratoon crop. The correlation coefficients SP with TCH, TSH, and EI were higher than that of SW. This implies that SP is the greatest contributor to the RA of sugar and yield traits in sugarcane. These results are consistent with few previous studies (Zhou and Shoko, 2012; Coto Arbelo et al., 2021).

The molecular and genetic basis of the RA variation has received little attention in scientific literature. Thus, it has been unknown what molecular mechanism accounts for sugarcane's ability to ratoon. To break through this impasse, we have conducted GWAS analysis for the five RA traits, detected putative candidate genes, and explored GS for the first time. In GWAS, mutations and potential functional markers statistically associated with complex quantitative traits are identified as underlying causal genes, mutations, and genetic variants (Islam et al., 2016). However, our efforts were not easy and further complicated for several reasons. At first, estimating the RA is complicated by difficulty in accounting for genotype by year interactions. Using multivariate mixed modeling in conjunction with the methods described by Milligan et al. (1996) and including data from two ratoon cycles, we estimated the RA of five traits while accounting for covariance between ratoon crops. Like all other genomic studies in sugarcane, our study faced numerous challenges such as a large genome size (10 Gb), complex genome architecture, types and level of ploidy, method of propagation, highly heterozygous genetic makeup, and
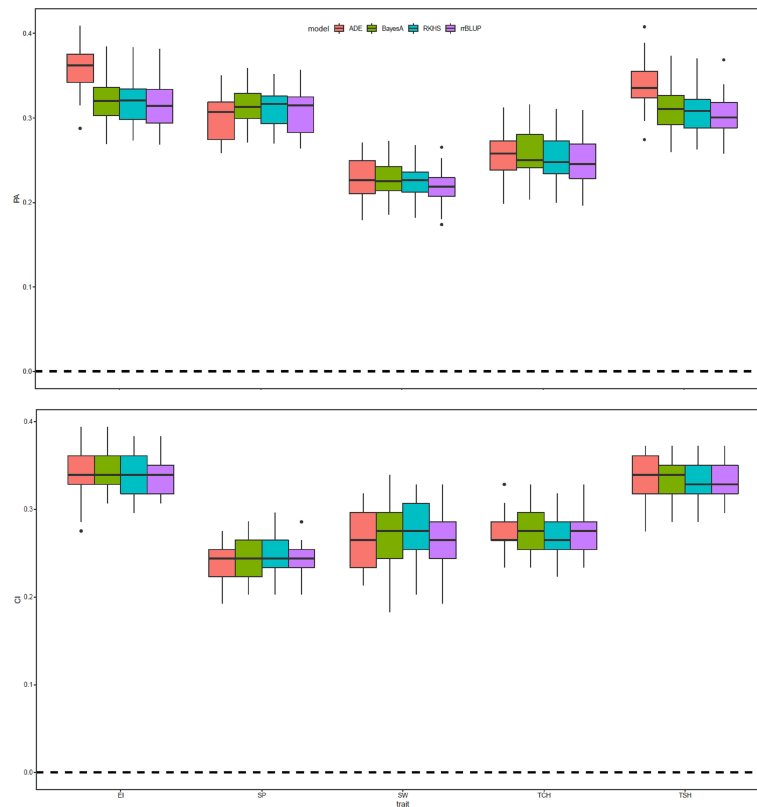
FIGURE 6
Prediction accuracy (PA) and coincidence index (CI) of the genomic estimated breeding value (GEBV) of sugarcane ratooning ability (%) of economic profitability (EI), stalk population (SP), stalk weight (SW), cane yield (TCH), and sucrose yield (TSH) for 5-fold cross-validation (5-fold CV) of four genomic selection (GS) methods with and without most significant peak marker from genome-wide association analysis as a fixed covariate for the respective traits. The 5-fold CV was completed by randomly selecting four-fifths of the individuals for the training and the remaining fifth as the validation population.
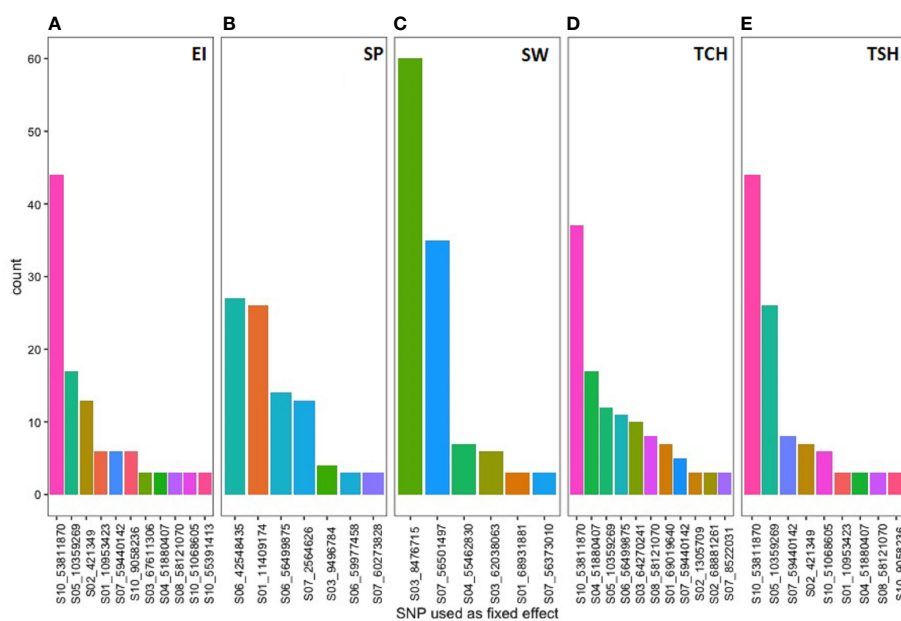

FIGURE 7
The single-nucleotide polymorphism (SNP) markers associated with the most significant QTL linked to tested ratooning ability traits [(A) EI, (B) SP, (C) SW, (D) TCH, and (E) TSH] and their frequency were used as a fixed effect in the prediction models.

nonadditive genetic variation. When we began this study, a sugarcane genome reference was not openly available. Therefore, we aligned our sequence data to the sorghum genome, likely missing genomic regions unique to the sugarcane genome. Furthermore, while software packages such as GWASpoly are designed to model the effect of allelic dosage in polyploids, doing so requires sufficient sequence read depth to estimate ploidy. Given the read depth of our sequence data, we modeled sugarcane as a pseudo-diploid. We used GWASpoly, an R package, to quantify three different types of polyploid gene action, including additive, simplex dominant, and duplex dominant (Rosyara et al., 2016; Yang et al., 2019b). By using three models (additive, 1-dom-alternate, and 1-dom-reference), eight nonredundant SNPs were found significantly associated with four RA traits (EI, SW, TCH, and ISH) (Figure 4, Table 2; Supplementary Table S2). Out of six putative unique QTLs associated with the RA of the five tested traits, one QTL located on 10 was associated with three traits (EI, TCH, and TSH). This is not surprising, since EI and TSH were estimated from TCH data as well as highly correlated with each other (Figure 2). Thus, these three traits are interdependent and closely related genetically, suggesting that the improvement of one trait will improve the others, since they share common QTLs. The outcome of these results could not be compared, since no other study has been reported. Thus, all of the marker–trait associations and QTLs related to the RA identified in this study are believed to be novel.

The probe sequences of the significant seven associated markers with the RA of the five traits were utilized to pinpoint the putative candidate genes possibly regulating the RA in sugarcane. Sugarcane stubble morphology studies have found that deep (long) roots, a large number of buds, a large number of live buds, and a large number of permanent roots (Chumphu et al., 2019), along with a reasonable leaf size and a reasonable tillering ability (Gomathi et al., 2013), are associated with a strong RA. Photosynthetic parameters such as chlorophyll fluorescence and stomatal conductance associated with ratoon sugarcane yield are significantly correlated with these variables (Chumphu et al., 2019). A C4 plant, such as sugarcane, has superior photosynthesis over a C3 plant because it concentrates CO_2 around Rubisco and uses the NADP malic enzyme to increase the sugar yield (Faralli and Lawson, 2020). The most significant associated SNP ($S03\_8476715$) with the RA of SW is physically colocated with the gene Sobic.003G096000 in the sorghum reference genome, which was annotated as an aldolase superfamily protein (Table 3; Supplementary Table S4). Surprisingly, this SNP was also most frequently picked up as a fixed-effect covariate during genomic prediction analysis. When the probe sequence of this SNP was compared with the sugarcane monoploid genome "R570" using BLAST, it was aligned with a similar homologous gene (Supplementary Table S4). This nuclear encoded chloroplast gene regulates several pathways that produce energy through photosynthesis, glycolysis, gluconeogenesis, and the Calvin cycle in the plant (Lu et al., 2012; Mininno et al., 2012; Carrera et al., 2021). It also reported that this gene governs root growth through helping the transportation of metabolites into the root (Mininno et al., 2012). An additional critical SNP associated with several target traits ($S07\_59440142$) with the RA of EI, TCA,

and TSH was linked with gene Sobic.007G160050 and $Sh07\_t012030$ in the sorghum and sugarcane monoploid "R570" reference genome, respectively. These genes are functionally annotated as dicarboxylate carrier 2 and regulate the transportation of metabolites such as malic acid and abscisic acid (ABA) in the cell for controlling several energy-related pathways glycolysis, gluconeogenesis, and the TCA cycle in the plant (Narsai et al., 2011; Shingaki-Wells et al., 2014; Barreto et al., 2022). It has been found that ABA regulates the withering of ineffective tillers through molecular signals in the form of hormonal interactions in sugarcane (Qiu et al., 2018). Another significant SNP ($S10\_53811870$) associated with three RA traits (EI, TCH, and TSH) and picked up most frequently as a fixed-effect covariate during genomic prediction was linked with sorghum gene Sobic.010G195000. This gene is actively participating in plant development and growth. It has been shown that a double mutant of this type of gene reduced plant growth specially root length (Suo et al., 2021). It may be possible to gain greater insights into the mechanisms that underlie genetic control of the RA in sugarcane by further investigating these genomic locations and genes.

GS was experimentally evaluated in this study using four different GS methods in conjunction with the GWAS results incorporated as fixed effects to determine the prospect for future sugarcane hybrid breeding. The RA of five yield and sugar traits (EI, SP, SW, TCH, and TSH) were evaluated, and 5-fold cross-validated prediction accuracy differed by traits and models. The overall prediction accuracies were moderate (0.21–0.36). We are unable to compare our results, since there are no GS reports published on the RA traits. However, prediction accuracies on other yield, sugar, and disease traits in sugarcane were found to be low to moderate in several previous reports (Gouy et al., 2013; Hayes et al., 2021; Islam et al., 2021; Yadav et al., 2021; Islam et al., 2022a). The nonparametric model ADE performed slightly better than other models for some traits. Studies have found that nonparametric models performed better in the presence of dominant and epistatic effects, reducing prediction error and increasing prediction accuracy while detecting SNP–SNP and SNP–covariate interactions (Lubke et al., 2013; Waldmann, 2016). Hence, our results suggest that the genetic makeup of the complex RA traits in sugarcane contains nonadditive effects in this testing population (Howard et al., 2014; Li et al., 2018). Another study in sugarcane also advised that additive and nonadditive gene effects play a big role during predicting the GEBV in the GS study (Hayes et al., 2021). Overall, the models fitting fixed effects for the most significant associated markers for each respective trait performed poorer than the standard models without fixed effect, although this difference was not statistically significant. The genetic architecture of the tested traits has a greater effect on the performance of a fixed-effect model during genomic prediction (Bernardo, 2014; Rice and Lipka, 2019). Models that include QTL with large effects, such as those for disease resistance, tend to improve the prediction accuracy (Poland and Rutkoski, 2016) while including fixed marker effects for complex traits controlled by many small effect QTL reduces the performance of genomic prediction and increases model bias. A fixed-effect covariate included in the validation set may offer little or no

advantage over a standard model in terms of prediction accuracy, as it may have a substantially weaker association in the validation set (Rice and Lipka, 2019).

This is the first study of its kind for the RA, and while prediction accuracies are low to moderate, these are likely good enough to begin testing GS practically for the RA. There is promise on the ability to increase prediction accuracies by addressing some of the known issues in marker development and genome complexities as mentioned above, as the heritability of the five RA traits is quite high. The findings of this study lay the groundwork for opening a new avenue for improving RA of sugar- and yield-related traits in sugarcane through breeding.

Data availability statement

The genotype data presented in the study are deposited in the Ag Data Commons repository : https://doi.org/10.15482/USDA.ADC/1528986.

Author contributions

MI contributed to the conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, visualization, writing—original draft, and writing—review and editing. KC contributed to the formal analysis, software, visualization, and writing—review and editing. PM contributed to the conceptualization, resources, and writing—review and editing. AH-K contributed to the writing—review and editing and resources. AL contributed to the writing—review and editing. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The mention of trade names or commercial products in this article provide specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture, an equal opportunity provider and employer.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1205999/full#supplementary-material

References

Abdurakhmonov, I. Y., and Abdukarimov, A. (2008). Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int. J. Plant Genomics* 2008, 574927. doi: 10.1155/2008/574927

Abu-Ellail, F. F. B., Abd El-Azez, Y. M., and Bassiony, N. A. (2019). Assessment of ratooning ability and genetic variability of promising sugarcane varieties under middle Egypt conditions. *Electronic J. Plant Breed.* 10 (1), 143–154. doi: 10.5958/0975-928X.2019.00017.6

Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A. F., Resende, M. F. R.Jr., and Muñoz, P. R. (2016). AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. *The Plant Genome* 9 (3). doi: 10.3835/plantgenome2016.01.0009

Aono, A. H., Ferreira, R. C. U., Moraes, A.d. C.L., Lara, L.A.d. C., Pimenta, R. J. G., Costa, E. A., et al. (2022). A joint learning approach for genomic prediction in polyploid grasses. *Sci. Rep.* 12 (1), 12499. doi: 10.1038/s41598-022-16417-7

Arruda, M. P., Brown, P. J., Lipka, A. E., Krill, A. M., Thurber, C., and Kolb, F. L. (2015). Genomic selection for predicting head blight resistance in a wheat breeding program. *The Plant Genome* 8 (3). doi: 10.3835/plantgenome2015.01.0003

Bagyalakshmi, K., Viswanathan, R., and Ravichandran, V. (2019). Impact of the viruses associated with mosaic and yellow leaf disease on varietal degeneration in sugarcane. *Phytoparasitica* 47 (4), 591–604. doi: 10.1007/s12600-019-00747-w

Barreto, P., Arcuri, M. L. C., Lima, R. P. M., Marino, C. L., and Maia, I. G. (2022). Comprehensive In Silico Analysis and Transcriptional Profiles Highlight the Importance of Mitochondrial Dicarboxylate Carriers (DICs) on Hypoxia Response in Both Arabidopsis thaliana and Eucalyptus grandis. *Plants* 11 (2), 181. doi: 10.3390/plants11020181

Barreto, F. Z., Rosa, J. R. B. F., Balsalobre, T. W. A., Pastina, M. M., Silva, R. R., Hoffmann, H. P., et al. (2019). A genome-wide association study identified loci for yield

component traits in sugarcane (Saccharum spp.). *PloS One* 14 (7), e0219843. doi: 10.1371/journal.pone.0219843

Batista, L. G., Mello, V. H., Souza, A. P., and Margarido, G. R. A. (2022). Genomic prediction with allele dosage information in highly polyploid species. *Theor. Appl. Genet.* 135 (2), 723–739. doi: 10.1007/s00122-021-03994-w

Bernal-Vasquez, A. M., Gordillo, A., Schmidt, M., and Piepho, H. P. (2017). Genomic prediction in early selection stages using multi-year data in a hybrid rye breeding program. *BMC Genet.* 18 (1), 51. doi: 10.1186/s12863-017-0512-8

Bernardo, R. (2014). Genomewide selection when major genes are known. *Crop Sci.* 54 (1), 68–75. doi: 10.2135/cropsci2013.05.0315

Billings, G. T., Jones, M. A., Rustgi, S., Bridges, W. C., Holland, J. B., Hulse-Kemp, A. M., et al. (2022). Outlook for implementation of genomics-based selection in public cotton breeding programs. *Plants* 11 (11), 1446. doi: 10.3390/plants11111446

Butler, D., Cullis, B. R., Gilmour, A., and Gogel, B. (2009). *ASReml-R reference manual* (Brisbane, Australia: The State of Queensland, Department of Primary Industries and Fisheries).

Carrera, D.Á., George, G. M., Fischer-Stettler, M., Galbier, F., Eicke, S., Truernit, E., et al. (2021). Distinct plastid fructose bisphosphate aldolases function in photosynthetic and non-photosynthetic metabolism in Arabidopsis. *J. Exp. Bot.* 72 (10), 3739–3755. doi: 10.1093/jxb/erab099

Chen, X., Huang, Z., Fu, D., Fang, J., Zhang, X., Feng, X., et al. (2022). Identification of genetic loci for sugarcane leaf angle at different developmental stages by genome-wide association study. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.841693

Chumphu, S., Jongrungklang, N., and Songsri, P. (2019). Association of physiological responses and root distribution patterns of ratooning ability and yield of the second ratoon cane in sugarcane elite clones. *Agronomy* 9 (4), 200. doi: 10.3390/agronomy9040200

Coto Arbelo, O., Sandhu, H., Momotaz, A., Davidson, W., Sandoya, G., Baltazar, M., et al. (2021). Selection *via* cane yield and ratooning ability of sugarcane genotypes in sandy soils in Florida. *Sugar Tech* 23 (5), 986–998. doi: 10.1007/s12355-021-00973-9

Covarrubias-Pazaran, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PloS One* 11 (6), e0156744. doi: 10.1371/journal.pone.0156744

Debibakas, S., Rocher, S., Garsmeur, O., Toubi, L., Roques, D., D'Hont, A., et al. (2014). Prospecting sugarcane resistance to Sugarcane yellow leaf virus by genome-wide association. *Theor. Appl. Genet.* 127 (8), 1719–1732. doi: 10.1007/s00122-014-2334-7

Deomano, E., Jackson, P., Wei, X., Aitken, K., Kota, R., and Pérez-Rodríguez, P. (2020). 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.* 40 (4). doi: 10.1007/s11032-020-01120-0

Deren, C. W., Alvarez, J., and Glaz, B. (1995). Use of economic criteria for selecting clones in a sugarcane breeding program. *Proc. Int. Soc Sugar Cane Technol.* 21 (2), 437–447.

Dunckelman, J. W. (1982). Breeding behavior of ratooning ability in sugarcane. *LSU Historical Dissertations and Theses.* pp. 3714.

Edmé, S. J., Tai, P. Y. P., Glaz, B., Gilbert, R. A., Miller, J. D., Davidson, J. O., et al. (2005). Registration of 'CP 96-1252' sugarcane. *Crop Sci.* 45, 423. doi: 10.2135/cropsci2005.0423

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome J.* 4 (3). doi: 10.3835/plantgenome2011.08.0024

F. A. O (2020). *OECD-FAO Agricultural Outlook 2020-2029.* (accessed on 14 March 2023). Available online: https://www.oecd-ilibrary.org/sites/3736a600-en/index.html?itemId=/content/component/3736a600-en.

Faralli, M., and Lawson, T. (2020). Natural genetic variation in photosynthesis: an untapped resource to increase crop yield potential? *Plant J.* 101 (3), 518–528. doi: 10.1111/tpj.14568

Gao, X., Wu, Z., Liu, R., Wu, J., Zeng, Q., and Qi, Y. (2019). Rhizosphere bacterial community characteristics over different years of sugarcane ratooning in consecutive monoculture. *BioMed. Res. Int.* 2019, 4943150. doi: 10.1155/2019/4943150

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* 1207, 3907. doi: 10.48550/arXiv.1207.3907

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* 9 (1), 2638. doi: 10.1038/s41467-018-05051-5

Gezan, S. A., Osorio, L. F., Verma, S., and Whitaker, V. M. (2017). An experimental validation of genomic selection in octoploid strawberry. *Hortic. Res.* 4, 16070. doi: 10.1038/hortres.2016.70

Gianola, D., and Van Kaam, J. B. (2008). Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178 (4), 2289–2303. doi: 10.1534/genetics.107.084285

Gilbert, R. A., Comstock, J. C., Glaz, B., Edmé, S. J., Davidson, R. W., Glynn, N. C., et al. (2008). Registration of 'CP 00-1101' Sugarcane. *J. Plant Registrations* 2, 95–101. doi: 10.3198/jpr2007.12.0660crc

Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995). Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51 (4), 1440–1450. doi: 10.2307/2533274

Gomathi, R., Rao, P., Rakkiyappan, P., Sundara, B., and Shiyamala, S. (2013). Physiological studies on ratoonability of sugarcane varieties under tropical indian condition. *Am. J. Plant Sci.* 4 (2), 8. doi: 10.4236/ajps.2013.42036

Gouy, M., Rousselle, Y., Bastianelli, D., Lecomte, P., Bonnal, L., Roques, D., et al. (2013). Experimental assessment of the accuracy of genomic selection in sugarcane. *Theor. Appl. Genet.* 126 (10), 2575–2586. doi: 10.1007/s00122-013-2156-z

Hayes, B. J., Wei, X., Joyce, P., Atkin, F., Deomano, E., Yue, J., et al. (2021). Accuracy of genomic prediction of complex traits in sugarcane. *Theor. Appl. Genet.* 134 (5), 1455–1462. doi: 10.1007/s00122-021-03782-6

Howard, R., Carriquiry, A. L., and Beavis, W. D. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3 (Bethesda)* 4 (6), 1027–1046. doi: 10.1534/g3.114.010298

Islam, M. S., Fang, D. D., Jenkins, J. N., Guo, J., McCarty, J. C., and Jones, D. C. (2020). Evaluation of genomic selection methods for predicting fiber quality traits in Upland cotton. *Mol. Genet. Genomics* 295 (1), 67–79. doi: 10.1007/s00438-019-01599-z

Islam, M. S., McCord, P. H., Olatoye, M. O., Qin, L., Sood, S., Lipka, A. E., et al. (2021). Experimental evaluation of genomic selection prediction for rust resistance in sugarcane. *Plant Genome* 14 (3), e20148. doi: 10.1002/tpg2.20148

Islam, M. S., McCord, P., Read, Q. D., Qin, L., Lipka, A. E., Sood, S., et al. (2022a). Accuracy of genomic prediction of yield and sugar traits in saccharum spp. *Hybrids. Agric.* 12 (9), 1436. doi: 10.3390/agriculture12091436

Islam, M. S., Sandhu, H., Zhao, D., Sood, S., Momotaz, A., Davidson, R. W., et al. (2022b). Registration of 'CP 13-1223' sugarcane for Florida organic soils. *J. Plant Registrations* 16 (1), 54–63. doi: 10.1002/plr2.20186

Islam, M. S., Thyssen, G. N., Jenkins, J. N., Zeng, L., Delhom, C. D., McCarty, J. C., et al. (2016). A MAGIC population-based genome-wide association study reveals functional association of GhRBB1_A07 gene with superior fiber quality in cotton. *BMC Genomics* 17 (1), 903. doi: 10.1186/s12864-016-3249-2

Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief Funct. Genomics* 9 (2), 166–177. doi: 10.1093/bfgp/elq001

Lee, W. P., Stromberg, M. P., Ward, A., Stewart, C., Garrison, E. P., and Marth, G. T. (2014). MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PloS One* 9 (3), e90581. doi: 10.1371/journal.pone.0090581

Legendre, B. L. (1992). The core/press method for predicting the sugar yield from cane for use in cane payment. *Sugar J.* 54 (9), 2–7.

Li, R., and Bilal, U. (2021). Interactive web-based data visualization with R, plotly, and shiny (Carson Sievert). *Biometrics* 77 (2), 776–777. doi: 10.1111/biom.13474

Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A., and Li, Y. (2018). Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Front. Genet.* 9, 237. doi: 10.3389/fgene.2018.00237

Lu, W., Tang, X., Huo, Y., Xu, R., Qi, S., Huang, J., et al. (2012). Identification and characterization of fructose 1,6-bisphosphate aldolase genes in Arabidopsis reveal a gene family with diverse responses to abiotic stresses. *Gene* 503 (1), 65–74. doi: 10.1016/j.gene.2012.04.042

Lubke, G., Laurin, C., Walters, R., Eriksson, N., Hysi, P., Spector, T., et al. (2013). Gradient boosting as a SNP filter: an evaluation using simulated and hair morphology data. *J. Data Min. Genomics Proteomics* 4, 143. doi: 10.4172/2153-0602.1000143

Matthies, I. E., Malosetti, M., Röder, M. S., and van Eeuwijk, F. (2014). Genome-wide association mapping for kernel and malting quality traits using historical European barley records. *PloS One* 9 (11), e110046. doi: 10.1371/journal.pone.0110046

McCouch, S., Cho, Y., Yano, P., Blinstrub, M., Morishima, H., and Kinoshita, T. (1997). Report on QTL nomenclature. *Rice Genet. Newsl* 14, 11–13.

Meuwissen, T. H. (2009). Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genet. Selection Evol.* 41 (1), 35. doi: 10.1186/1297-9686-41-35

Milligan, S. B., Gravois, K. A., Bischoff, K. P., and Martin, F. A. (1990). Crop effects on genetic relationships among sugarcane traits. *Crop Sci.* 30 (4), cropsci1990. doi: 10.2135/cropsci1990.0011183X003000040034x

Milligan, S. B., Martin, F. A., and Gravois, K. A. (1996). Inheritance of sugarcane ratooning ability and the relationship of younger crop traits to older crop traits. *Crop Sci.* 36 (1), 45–60. doi: 10.2135/cropsci1996.0011183X003600010008x

Mininno, M., Brugière, S., Pautre, V., Gilgen, A., Ma, S., Ferro, M., et al. (2012). Characterization of chloroplastic fructose 1,6-bisphosphate aldolases as lysine-methylated proteins in plants*. *J. Biol. Chem.* 287 (25), 21034–21044. doi: 10.1074/jbc.M112.359976

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 110 (2), 453–458. doi: 10.1073/pnas.1215985110

Narsai, R., Rocha, M., Geigenberger, P., Whelan, J., and van Dongen, J. T. (2011). Comparative analysis between plant species of transcriptional and metabolic responses to hypoxia. *New Phytol.* 190 (2), 472–487. doi: 10.1111/j.1469-8137.2010.03589.x

O'Connell, A., Deo, J., Deomano, E., Wei, X., Jackson, P., Aitken, K. S., et al. (2022). Combining genomic selection with genome-wide association analysis identified a large-

effect QTL and improved selection for red rot resistance in sugarcane. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1021182

Olatoye, M. O., Clark, L. V., Wang, J., Yang, X., Yamada, T., Sacks, E. J., et al. (2019). Evaluation of genomic selection and marker-assisted selection in Miscanthus and energycane. *Mol. Breed.* 39 (12), 171. doi: 10.1007/s11032-019-1081-5

Pace, J., Yu, X., and Lubberstedt, T. (2015). Genomic prediction of seedling root length in maize (Zea mays L.). *Plant J.* 83 (5), 903–912. doi: 10.1111/tpj.12937

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 457 (7229), 551–556. doi: 10.1038/nature07723

Perez, P., and De los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198 (2), 483–495. doi: 10.1534/genetics.114.164442

Pimenta, R. J. G., Aono, A. H., Burbano, R. C. V., Coutinho, A. E., da Silva, C. C., dos Anjos, I. A., et al. (2021). Genome-wide approaches for the identification of markers and genes associated with sugarcane yellow leaf virus resistance. *Sci. Rep.* 11 (1), 15730. doi: 10.1038/s41598-021-95116-1

Poland, J., and Rutkoski, J. (2016). Advances and challenges in genomic selection for disease resistance. *Annu. Rev. Phytopathol.* 54 (1), 79–98. doi: 10.1146/annurev-phyto-080615-100056

Qiu, L. H., Fan, Y. G., Luo, H. M., Huang, X., Chen, R. F., Yang, R. Z., et al. (2018). Advances of regulation study on tillering formation and stem forming from available tillers in sugarcane (Saccharum officinarum). *Zhiwu Shengli Xuebao/Plant Physiol. J.* 54, 192–202. doi: 10.13592/j.cnki.ppj.2017.0464

Racedo, J., Gutierrez, L., Perera, M. F., Ostengo, S., Pardo, E. M., Cuenya, M. I., et al. (2016). Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biol.* 16 (1), 142. doi: 10.1186/s12870-016-0829-x

Ramburan, S., Wettergreen, T., Berry, S. D., and Shongwe, B. (2013). Genetic, environmental and management contributions to ratoon decline in sugarcane. *Field Crops Res.* 146, 105–112. doi: 10.1016/j.fcr.2013.03.011

Rice, B., and Lipka, A. E. (2019). Evaluation of RR-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *Plant Genome* 12 (1), 180052. doi: 10.3835/plantgenome2018.07.0052

Rosyara, U. R., De Jong, W. S., Douches, D. S., and Endelman, J. B. (2016). Software for genome-wide association studies in autopolyploids and its application to potato. *The Plant Genome* 9 (2). doi: 10.3835/plantgenome2015.08.0073

Shingaki-Wells, R., Millar, A. H., Whelan, J., and Narsai, R. (2014). What happens to plant mitochondria under low oxygen? An omics review of the responses to low oxygen and reoxygenation. *Plant Cell Environ.* 37 (10), 2260–2277. doi: 10.1111/pce.12312

Shukla, S. K., Yadav, R. L., Suman, A., and Singh, P. N. (2008). Improving rhizospheric environment and sugarcane ratoon yield through bioagents amended farm yard manure in udic ustochrept soil. *Soil Tillage Res.* 99 (2), 158–168. doi: 10.1016/j.still.2008.02.007

Singh, R. K., Banerjee, N., Khan, M. S., Yadav, S., Kumar, S., Duttamajumder, S. K., et al. (2016). Identification of putative candidate genes for red rot resistance in sugarcane (Saccharum species hybrid) using LD-based association mapping. *Mol. Genet. Genomics* 291 (3), 1363–1377. doi: 10.1007/s00438-016-1190-3

Singh, A., Chauhan, S. S., Singh, A., and Singh, S. B. (2006). Deterioration in sugarcane due to pokkah boeng disease. *Sugar Tech* 8 (2), 187–190. doi: 10.1007/BF02943659

Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B-Statistical Method.* 64, 479–498. doi: 10.1111/1467-9868.00346

Suo, Y., Hu, F., Zhu, H., Li, D., Qi, R., Huang, J., et al. (2021). BIG3 and BIG5 redundantly mediate vesicle trafficking in arabidopsis. *Biomolecules* 11 (5), 732. doi: 10.3390/biom11050732

Tadesse, W., Ogbonnaya, F. C., Jighly, A., Sanchez-Garcia, M., Sohail, Q., Rajaram, S., et al. (2015). Genome-wide association mapping of yield and grain quality traits in winter wheat genotypes. *PloS One* 10 (10), e0141339. doi: 10.1371/journal.pone.0141339

Thyssen, G. N., Jenkins, J. N., McCarty, J. C., Zeng, L., Campbell, B. T., Delhom, C. D., et al. (2018). Whole genome sequencing of a MAGIC population identified genomic loci and candidate genes for major fiber quality traits in upland cotton (Gossypium hirsutum L.). *Theor. Appl. Genet* 132, 989–999. doi: 10.1007/s00122-018-3254-8

Todd, J. R., Kimbeng, C., Hale, A. L., Gravois, K., and Dufrene, E. O.Jr. (2017). GGE analysis of ratooning ability in Louisiana sugarcane breeding. *J. Am. Soc. Sugar Cane Technol.* 37, 29–30.

VanWeelden, M., Swanson, S., Davidson, W., Baltazar, M., and Rice, R. (2020). Sugarcane variety census: Florida 2019. *Sugar J.* 82 (7), 8–20.

Voss-Fels, K. P., Wei, X., Ross, E. M., Frisch, M., Aitken, K. S., Cooper, M., et al. (2021). Strategies and considerations for implementing genomic selection to improve traits with additive and non-additive genetic architectures in sugarcane breeding. *Theor. Appl. Genet.* 134 (5), 1493–1511. doi: 10.1007/s00122-021-03785-3

Vuong, T. D., Sonah, H., Meinhardt, C. G., Deshmukh, R., Kadam, S., Nelson, R. L., et al. (2015). Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* 16, 593. doi: 10.1186/s12864-015-1811-y

Waldmann, P. (2016). Genome-wide prediction using Bayesian additive regression trees. *Genet. Sel Evol.* 48 (1), 42. doi: 10.1186/s12711-016-0219-8

Wubben, M. J., Thyssen, G. N., Callahan, F. E., Fang, D. D., Deng, D. D., McCarty, J. C., et al. (2019). A novel variant of Gh_D02G0276 is required for root-knot nematode resistance on chromosome 14 (D02) in Upland cotton. *Theor. Appl. Genet.* 132 (5), 1425–1434. doi: 10.1007/s00122-019-03289-1

Xu, Y., and Crouch, J. H. (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* 48 (2), 391–407. doi: 10.2135/cropsci2007.04.0191

Xu, F., Wang, Z., Lu, G., Zeng, R., and Que, Y. (2021). Sugarcane ratooning ability: research status, shortcomings, and prospects. *Biol. (Basel)* 10 (10), 1052. doi: 10.3390/biology10101052

Yadav, S., Wei, X., Joyce, P., Atkin, F., Deomano, E., Sun, Y., et al. (2021). Improved genomic prediction of clonal performance in sugarcane by exploiting non-additive genetic effects. *Theor. Appl. Genet.* 134 (7), 2235–2252. doi: 10.1007/s00122-021-03822-1

Yang, X., Luo, Z., Todd, J., Sood, S., and Wang, J. (2020). Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (Saccharum spp.). *Plant Genome* 13 (1), e20006. doi: 10.1002/tpg2.20006

Yang, X., Sood, S., Luo, Z., Todd, J., and Wang, J. (2019a). Genome-wide association studies identified resistance loci to orange rust and yellow leaf virus diseases in sugarcane (Saccharum spp.). *Phytopathology* 109 (4), 623–631. doi: 10.1094/PHYTO-08-18-0282-R

Yang, X. P., Todd, J., Arundale, R., Binder, J. B., Luo, Z. L., Islam, M. S., et al. (2019b). Identifying loci controlling fiber composition in polyploid sugarcane (Saccharum spp.) through genome-wide association study. *Ind. Crops Products* 130, 598–605. doi: 10.1016/j.indcrop.2019.01.023

Zhou, M. M., and Shoko, M. D. (2012). Simultaneous selection for yield and ratooning ability in sugarcane genotypes using analysis of covariance. *South Afr. J. Plant Soil* 29 (2), 93–100. doi: 10.1080/02571862.2012.717639