



Genomic Selection in Sugarcane: Current Status and Future Prospects

Channappa Mahadevaiah^{1*}, Chinnaswamy Appunu^{1*}, Karen Aitken²,
Giriyapura Shivalingamurthy Suresha³, Palanisamy Vignesh¹,
Huskur Kumaraswamy Mahadeva Swamy¹, Ramanathan Valarmathi¹,
Govind Hemaprabha¹, Ganesh Alagarasan¹ and Bakshi Ram¹

¹ Division of Crop Improvement, ICAR-Sugarcane Breeding Institute, Coimbatore, India, ² CSIRO (Commonwealth Scientific and Industrial Research Organization), St. Lucia, QLD, Australia, ³ Division of Crop Production, ICAR-Sugarcane Breeding Institute, Coimbatore, India

OPEN ACCESS

Edited by:

Marcelino Perez De La Vega,
Universidad de León, Spain

Reviewed by:

Youxiang Que,
Fujian Agriculture and Forestry
University, China
Mulatu Geleta,
Swedish University of Agricultural
Sciences, Sweden

*Correspondence:

Channappa Mahadevaiah
C.Mahadevaiah@icar.gov.in
Chinnaswamy Appunu
C.Appunu@icar.gov.in

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 11 May 2021

Accepted: 24 August 2021

Published: 27 September 2021

Citation:

Mahadevaiah C, Appunu C, Aitken K,
Suresha GS, Vignesh P, Mahadeva
Swamy HK, Valarmathi R,
Hemaprabha G, Alagarasan G and
Ram B (2021) Genomic Selection in
Sugarcane: Current Status and Future
Prospects.
Front. Plant Sci. 12:708233.
doi: 10.3389/fpls.2021.708233

Sugarcane is a C4 and agro-industry-based crop with a high potential for biomass production. It serves as raw material for the production of sugar, ethanol, and electricity. Modern sugarcane varieties are derived from the interspecific and intergeneric hybridization between *Saccharum officinarum*, *Saccharum spontaneum*, and other wild relatives. Sugarcane breeding programmes are broadly categorized into germplasm collection and characterization, pre-breeding and genetic base-broadening, and varietal development programmes. The varietal identification through the classic breeding programme requires a minimum of 12–14 years. The precise phenotyping in sugarcane is extremely tedious due to the high propensity of lodging and suckering owing to the influence of environmental factors and crop management practices. This kind of phenotyping requires data from both plant crop and ratoon experiments conducted over locations and seasons. In this review, we explored the feasibility of genomic selection schemes for various breeding programmes in sugarcane. The genetic diversity analysis using genome-wide markers helps in the formation of core set germplasm representing the total genomic diversity present in the *Saccharum* gene bank. The genome-wide association studies and genomic prediction in the *Saccharum* gene bank are helpful to identify the complete genomic resources for cane yield, commercial cane sugar, tolerances to biotic and abiotic stresses, and other agronomic traits. The implementation of genomic selection in pre-breeding, genetic base-broadening programmes assist in precise introgression of specific genes and recurrent selection schemes enhance the higher frequency of favorable alleles in the population with a considerable reduction in breeding cycles and population size. The integration of environmental covariates and genomic prediction in multi-environment trials assists in the prediction of varietal performance for different agro-climatic zones. This review also directed its focus on enhancing the genetic gain over time, cost, and resource allocation at various stages of breeding programmes.

Keywords: sugarcane, genomic selection, pre-breeding, genetic base-broadening, multi-environment trials, genomic estimated breeding value

INTRODUCTION

Sugarcane is an important agro-based industrial crop cultivated in tropical and sub-tropical regions; it serves as a raw material for the production of sugar, bioethanol, and bioenergy (Hoang et al., 2015). Globally, sugarcane is cultivated in 28.19 million hectares which produces 2059.74 million tonnes of canes annually with average productivity of 72.80 t ha⁻¹ (FAOSTAT, 2019). As a C4 plant, it is a high biomass producing crop generating around 279 million tonnes of lignocellulosic biomass residues, i.e., leaves and bagasse, per year worldwide (Chandel et al., 2012). The adoption of new varieties has enhanced the cane and sugar yield in India, Australia, Brazil, and South Africa (Burnquist et al., 2010; Ram and Hemaprabha, 2020; Schmitz et al., 2020). Modern breeding approaches, such as genomic selection involving inter and multi-disciplinary approaches (Crossa et al., 2017), are required to further enhance the genetic gain for cane yield and sugar yield.

Sugarcane is an exemplary crop that exploits the heterosis from the wild genetic resources. The modern cultivars are interspecific and intergeneric hybrids of *Saccharum officinarum*, *Saccharum spontaneum*, and other related genera (Bhat and Gill, 1985; Sreenivasan and Ahloowalia, 1987; Grivet and Arruda, 2002; Lekshmi et al., 2017). The basic clones of *Saccharum* species significantly contributed to the heterosis and genetic base diversification in sugarcane germplasm. The in-depth pedigree analysis of the modern sugarcane cultivars and breeding lines revealed that a limited number of germplasm and basic species were used in the varietal development programmes (Jackson, 2005). The pedigree analysis of Indian 'Co' varieties exhibited that only two *S. spontaneum* accessions viz., *S. spontaneum* CBE and *S. spontaneum* Java were used in the breeding programmes (Kumar et al., 2012). The pedigree analysis of notified thirteen varieties in India (Figure 1) depicts the narrow genetic base in cultivated varieties and these varieties are derived from the clones of a limited basic *Saccharum* species such as 17 accessions of *S. officinarum* and its derivatives, one *Saccharum barberi*, two *S. spontaneum*, and one *Erianthus arundinaceus* derived clone and two foreign clones. Though the diverse wild genetic resources were used in the pre-breeding and many breeding lines or genetic stocks were identified (Moharaj and Nair, 2014; Nair et al., 2017), it might not have selected or captured all the desirable genes in the background of undesirable linkage drags. Hence, there is a requirement to take the advantage of the genome-wide markers and genomic selection for effective utilization of genetic resources in the sugarcane breeding programme.

Sugarcane breeding has been performed in three phases (Jackson, 2005). First is the crossing and selecting among the *S. officinarum* clones. Second is the interspecific hybridization and nubilization of canes. Third is the intermating or recurrent selection among the nubilized cane to develop the commercial

canes. The genetic variability in the present working sugarcane germplasm is not fully exhausted or no yield plateau was observed in sugarcane (Edm et al., 2005), but there is evidence of slow genetic gain (Wei and Jackson, 2017). Additionally, new variability is continuously created through pre-breeding in many countries (da Silva, 2017; Nair et al., 2017; Cursi et al., 2021), yet the application of genomic selection in sugarcane is capable of capturing the entirety of the allelic variability and enhancing the genetic gain (Deomano et al., 2020; Hayes et al., 2021; Voss-Fels et al., 2021; Yadav et al., 2021). The recent review on genomic selection in sugarcane highlighted the recurrent selection schemes (Yadav et al., 2020) and we reviewed the feasibility of genomic selection at various stages of sugarcane breeding programmes, such as germplasm evaluation and formation of core germplasm, pre-breeding and genetic base-broadening programmes, and multi-environmental trials of breeding lines.

GENOMIC SELECTION IN SUGARCANE

The genetic gain for commercial sugar yield can be improved either by enhancing the cane yield or sucrose content, and both traits were reported to have a poor genetic correlation (Jackson, 2005). Both cane yield and sucrose were governed by many quantitative trait loci (QTL) or genomic regions (Hoarau et al., 2002; Ming et al., 2002; Aitken et al., 2008). The narrow-sense heritability and very high non-additive genetic variance for cane yield (Brown et al., 1968; Hogarth, 1971; Hogarth et al., 1981; Barbosa et al., 2005; Wei and Jackson, 2017), the selection of parents solely based on breeding values do not improve the genetic gain for cane yield (Wei and Jackson, 2017). The utility of genome-wide markers such as single nucleotide polymorphism (SNP) for prediction of breeding values of parents and genotypes through genomic selection approaches were demonstrated in other crops (Unterseer et al., 2014; Yu et al., 2014; Aitken et al., 2016; Pandey et al., 2017; Roorkiwal et al., 2018; Saxena et al., 2018; Li et al., 2019). The Diversity Array Technology (DArT), Axiom 345K SNP, and Affymetrix Axion 100K SNP arrays were now available (Heller-Uszynska et al., 2011; Berkman et al., 2014; Aitken et al., 2016; Song et al., 2016; You et al., 2018) and utilized for genomic prediction studies in sugarcane (Gouy et al., 2013; Deomano et al., 2020; Hayes et al., 2021; Voss-Fels et al., 2021; Yadav et al., 2021). The accuracy of genomic prediction or selection studies depends on many parameters such as robust statistical models, genome-wide markers, and parameters related to training and testing populations such as population structure, trait heritability, trait architecture, genetic diversity, and genetic relatedness (Meuwissen et al., 2001; Zhang et al., 2017; Berro et al., 2019).

Genomic Prediction Models

The genomic prediction models are statistical models, which combines the pedigree data, genotypic data, phenotypic data, and environmental covariates to estimate the genomic estimated breeding values (GEBV) and enhance the prediction accuracies (Meuwissen et al., 2001; Rincent et al., 2012; Heslot et al., 2014;

Abbreviations: CCS, Commercial cane sugar; QTL, Quantitative trait loci; NGS, Next-generation sequencing technologies; GBS, Genotyping by sequencing; SNP, Single nucleotide polymorphism; GWAS, Genome-wide association studies; BLUP, Best linear unbiased prediction; RR-BLUP, Ridge regression BLUP; LASSO, Least absolute shrinkage and selection operator; GBLUP, Genomic BLUP; RKHS, Reproducing kernel Hilbert space; DArT, Diversity arrays technology; taBLUP, Trait associated BLUP; sBLUP, Super BLUP; cBLUP, Compressed BLUP.

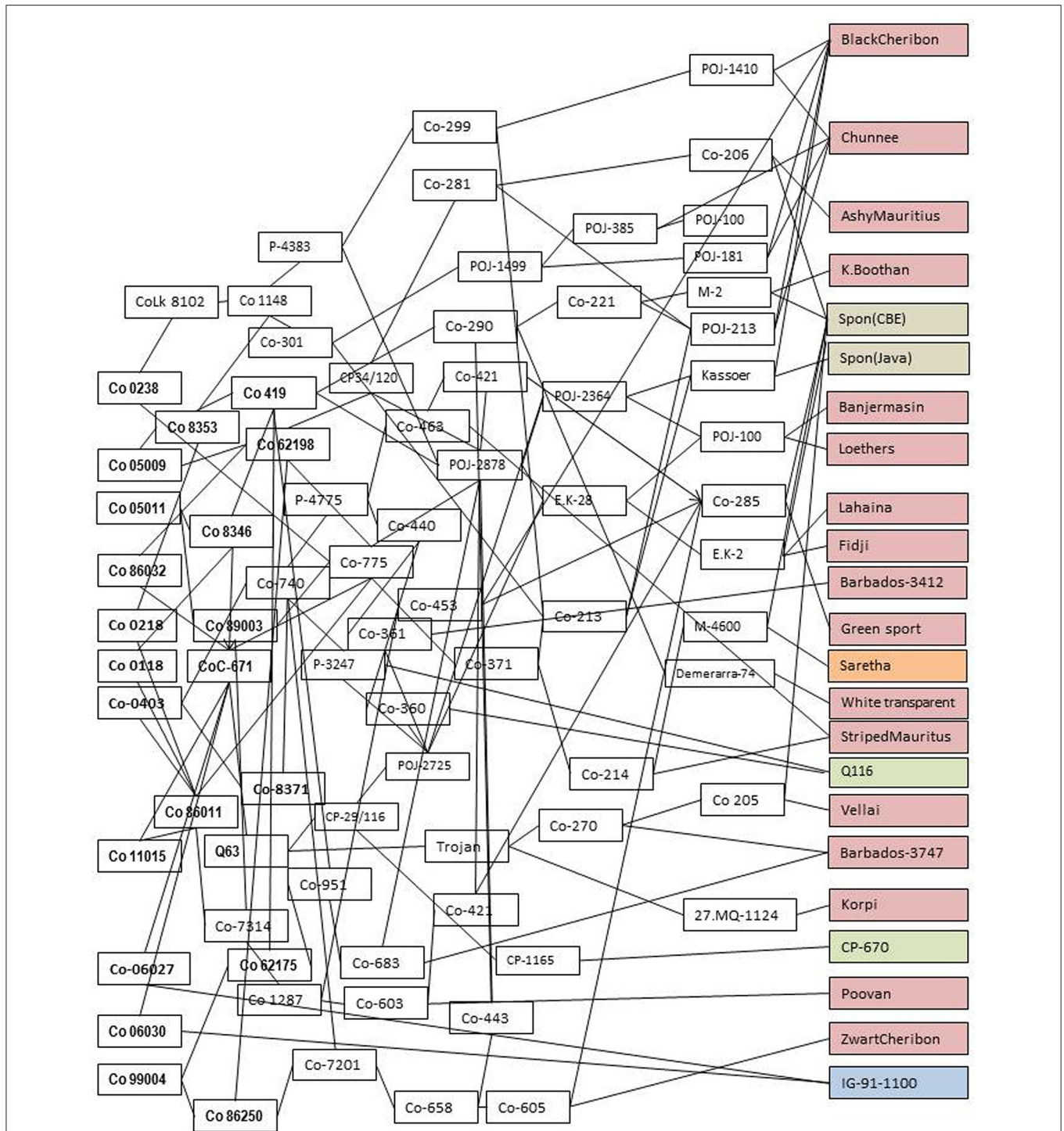


FIGURE 1 | In-depth pedigree analysis of notified sugarcane varieties in India stipulating the total genetic variability present in the field. They are derived from the limited germplasm viz., 17 clones of *Saccharum officinarum* and its derivatives (pink color), one clone of *Saccharum barberi* (orange), two clones of *Saccharum spontaneum* (gray), one genetic stock IG91-1100 derived from intergeneric hybridization between sugarcane (CoC 772) and *Erianthus arundinaceus* (blue), and two foreign clones (green). The implementation of genomic prediction in sugarcane germplasm characterization and genomic selection in pre-breeding aids in introgression and augmentation of more favorable alleles, base-broadening of working germplasm, and finally deploying more number of favorable genes into the target environments.

Malosetti et al., 2016; Monteverde et al., 2019). This is a multi-disciplinary approach involving computer science, genetics, quantitative genetics, statistics, bioinformatics, mathematics, physics, and machine learning to deduce the estimated breeding values (Crossa et al., 2017). The genomic prediction models also involve cross-validation to determine their accuracies (Burgueño et al., 2012; Rincent et al., 2012, 2017). Most of the statistical models, such as ridge-regression best linear unbiased prediction (RR-BLUP), BayesA, BayesB, BayesC, least absolute shrinkage and selection operator (LASSO), and genomic BLUP (GBLUP), estimate the additive genetic variance or breeding values (Wang et al., 2018b). In sugarcane genomic prediction studies, genomic models such as BayesA, BayesB, Bayesian LASSO, Bayesian GBLUP, and reproducing kernel Hilbert space (RKHS) were used to predict the commercial cane sugar (CCS) and cane yield in early and advance breeding experiments in Australia. The prediction accuracy (0.25–0.45) was highly encouraging under additive genomic models. The prediction accuracies for CCS in advanced clonal selection trials which was higher than the early clonal trials and, vigorous selection in early clonal trials did not affect the prediction accuracies of advanced selection trials. The reverse trend was observed for cane yield, wherein the prediction accuracies in advanced trials were less than that of early clonal trials (Deomano et al., 2020) and it was attributed to the competition effects in smaller plots (Jackson and McRae, 2001). Three additive genomic models namely GBLUP, genomic single step, and BayesR were tested against 3,984 breeding lines genotyped with 26K SNP, achieving satisfactory prediction accuracies (0.30–0.44) with 5% improvement for CCS and fiber without any improvement for cane yield due to competition effects in early breeding trials (Hayes et al., 2021).

The non-additive genetic variance significantly contributes to the heterosis in sugarcane and the low narrow-sense heritability is one of the causes for slow genetic gain to cane yield (Wei and Jackson, 2017). RKHS and extended GBLUP genomic models captures the non-additive genetic variance (Endelman, 2011; Jiang and Reif, 2015; Momen and Morota, 2018; Momen et al., 2018; Varona et al., 2018) and RKHS genomic model was used in sugarcane genomic selection (Gouy et al., 2013; Yadav et al., 2021). In simulation studies, genetic gain due to additive genomic model was almost double than the phenotypic selection and the genomic selection enhanced the rapid increase in frequencies of favorable alleles. Whereas, under the non-additive model, the genetic gain under genomic selection was still higher than the phenotypic selection and the success of the genomic selection depends on genomic breeding schemes, clonal population and stage of breeding programmes (Voss-Fels et al., 2021). In another study, extended GBLUP incorporated with additive, dominance, epistatic, and average genome-wide heterozygosity components to dissect the non-additive variance for cane yield, CCS, and fiber. The large portion of variance was attributed to dominance variance for cane yield and epistatic genetic variance for both cane yield and CCS. The epistatic genetic variance (additive \times additive genetic variance) and variance due to genome-wide heterozygosity had accounted significant portion of genetic variance for both cane yield and CCS under additive-dominance-epistatic

and additive-dominance-epistatic-genome-wide heterozygosity genomic models (Yadav et al., 2021).

Genomic prediction models are evolving science, and several modifications of the genetic relationship matrices were reported to enhance the prediction accuracies. The trait associated BLUP (taBLUP) make use of the trait-specific relationship matrix which captures the genetic variance of a locus, whose elements were derived based on the probability of a marker locus identical by descent and genetic variance of a trait. The genetic architecture was incorporated into the kinship matrix by assigning the suitable weight to the markers identified through genome-wide association studies (GWAS) and additionally improved by removing the duplicate trait-specific markers associated with the same (Zhang et al., 2010; Wang et al., 2018a). Further, the elements of kinship matrices were modified by assigning the markers to bins or real quantitative trait nucleotides identified in GWAS. In super BLUP (sBLUP), the maximum likelihood association between the traits and markers were enhanced by incorporating the elements of kinship matrices by the markers identified within a bin or real quantitative trait nucleotide identified in GWAS. Whereas, in compressed BLUP (cBLUP), the kinship matrices were derived by clustering the genotypes into groups and each group was treated as a random effect (Wang et al., 2018a). Further studies are required to understand the impact of sBLUP and cBLUP in the background of additive, dominance, epistatic and genome-wide heterozygosity models for improving the prediction accuracies in sugarcane.

Genome-Wide Markers

One of the most commonly raised issues in the sugarcane genome analysis is that it exhibits hyperploidy. Hyperploidy is a major problem in SNP identification among polyploids, as it may lead to false SNP calling in crops like sugarcane and strawberry. The challenges in SNP identification and SNP selection criteria in polyploids were recently reviewed (You et al., 2018; Manimekalai et al., 2020). In recently formed polyploid genomes, such as sugarcane, the problem is 2-fold for SNP calling. On the other hand, there is a paucity of high-end algorithms and computational methods, as large numbers of the available SNP analyzing tools are not suitable for the polyploid genome. Two relevant kinds of technology gaps have been identified in the sugarcane SNP genotyping. The first is that of homoeologous SNPs required to differentiate from the true allelic SNPs in polyploid crops. For read alignment, Bowtie-2 and Burrows-Wheeler Aligner - Maximal Exact Matches (BWA-mem) identify the read alignments based on the number of mismatch scores (Clevenger et al., 2015) and Universal Network-Enabled Analysis Kit (UNEAK) SNP discovery pipeline is also included in the Trait Analysis by aSSociation, Evolution and Linkage - Genotyping by sequencing (TASSEL-GBS) (Elshire et al., 2011; Lu et al., 2013). The various advanced sequencing platforms have demonstrated their role in diploid/polyploid crops. The Illumina platform presents the most robust and cost-effective method of polyploid genome analysis through short-read sequencing, even though a longer read length is preferred to sequence genomes like sugarcane, which possess high levels of repetitive regions (You et al., 2018). Currently, PacBio offers long-read sequencing on

an average read length of 15 kb through single-molecule real-time sequencing (Manimekalai et al., 2020). However, researchers from developing nations are not able to use PacBio or any other advanced platforms due to the higher sequencing costs. Cheaper sequencing costs would facilitate in-depth sequence coverage, which is a preferred strategy in outcrossing species. In diploids, 7.7x coverage is sufficient, whereas the tetraploid, autotetraploid, and autooctoploid require 15x, 48x, and 100x, respectively, for accurate SNP genotyping by GBS. For example, the suggested sequencing depth in potatoes is 60x to avoid type I and type II errors in SNP genotyping (Clevenger et al., 2015).

The second problem for the generation of robust SNP data is the raw data analysis of the sequence reads that are generated. Data duplications, such as PCR duplication, is a common problem faced by many computational biologists (Clevenger et al., 2015). It is very difficult to categorize the natural duplication from PCR duplication, which results in incorrect InDel/SNP calling even in diploids (Li, 2014). Despite severe limitations, some of the researchers have successfully identified the SNPs through the Next Generation Sequencing (NGS)-enabled method, genotyping-by-sequencing (GBS): 20K SNPs (164 wheat DH lines), 76K SNPs (14 sugarcane accessions), and 84K SNPs (151 sugarcane clones) (Poland et al., 2012; Yang et al., 2017). The Universal Network-Enabled Analysis Kit (UNEAK pipeline) was developed as a reference method and supplemented with the TASSEL-GBS SNP calling software useful for complex genomes (Lu et al., 2013). Several studies have identified large numbers of SNP markers in sugarcane through GBS (Balsalobre et al., 2017; Yang et al., 2017) or hybridization-based target enrichment method combined with NGS (Song et al., 2016) or the generation of an Axiom SNP array (Aitken et al., 2016). Assuming the pseudo-diploid model, the heterozygous alleles of single-dose markers were scored as one group and used in the genomic prediction studies in sugarcane (Deomano et al., 2020; Hayes et al., 2021; Voss-Fels et al., 2021; Yadav et al., 2021). Though this has resolved multi-dosage effects of markers, it does not reflect the true genomic complexity of sugarcane (Voss-Fels et al., 2021) and yet the considerable improvement in genomic prediction accuracies for cane and sugar yields under the additive and non-additive genetic models is found worthy in sugarcane. Furthermore, the use of these multi-dose markers in breeding remains a challenge and there exists a further need to address the computational challenges for the utilization of multi-dose markers in complex polyploid crops like sugarcane.

Population Structure, Genetic Diversity, and Relatedness of Training and Testing Populations

The design of the training and testing populations has to consider several parameters such as trait heritability, population structure, genetic architecture, size of the training and validation populations, and genome-wide distribution of markers. Genomic prediction accuracies are influenced by the population structure and genetic relatedness of training and testing populations (Windhausen et al., 2012; Sallam et al., 2015). The population structure arises either due to linkages between the similar alleles or genomic regions among closely related individuals or due to

the linkage disequilibrium of similar genomic regions among the distantly related individuals, which does not decay over the fewer generations of selfing or crossing. Both linkage and linkage disequilibrium contribute to the population structure and requires due considerations in genomic prediction (Daetwyler et al., 2012). The minimal population structure, maximum genetic diversity among the training population and proximal genetic relatedness between training and testing populations amplified the genomic prediction accuracies (Clark et al., 2012; Daetwyler et al., 2012; Guo et al., 2014; Sallam et al., 2015). To achieve the same, several statistical approaches were used in different studies such as k -vertex connected graph with the maximum number of edges and the K -mean clustering strategy (Maenhout et al., 2010), clusters of relationship coefficients matrices (Saatchi et al., 2011), CD mean and generalized CD (Rincent et al., 2012, 2017), stratified sampling based on Euclidian distance and Ward's hierarchical clustering, CDmean, PEVmean, and stratified CDmean (Isidro et al., 2015), and the weighted additive relationship matrix with a stratified sampling (Zhang et al., 2010; Wang et al., 2018a; Berro et al., 2019).

Several GWAS studies revealed that population structure is evident in sugarcane. The population structure analysis in the panel comprising of 134 accessions of cultivars and popular parental lines displayed a very strong linkage disequilibrium with four different groups (Barreto et al., 2019). The population structure analysis in a 96 genotypes panel comprising of wild species by using the target region amplified polymorphism (TRAP) markers presented the two subgroups for sucrose metabolic pathways and three subgroups for lignin metabolic pathways. Most of the traditional cultivars, accessions of *S. spontaneum* and *Erianthus* spp. formed one subpopulation, whereas the modern cultivars were assigned into the second subpopulation (Junior et al., 2020). The population structure analysis in 97 elite and historic sugarcane varieties by using 6,534 InDel and SNP markers revealed the strong structural differentiation into two major subgroups and a stronger marker-trait correlation with sucrose traits (Fickett et al., 2019). Sugarcane crop is a photosensitive crop which requires specific environmental regimes for plant developmental activities (Baez-Gonzalez et al., 2017). The crop growth and developments differ with agro-climatic zones such as tropical and subtropical regions. The spatial variation is well-known in sugarcane and the association mapping based on a single location resulted in the identification of markers, which does not possess any significance for utilization in sugarcane breeding programmes (Wei et al., 2010). Therefore, the formation of training and testing populations has to provide due consideration to the multi-environmental evaluation and environmental covariates to enhance the genomic prediction accuracies (Pandey et al., 2020) in sugarcane.

Trait Heritability and Genetic Architecture in Training and Testing Populations

The genetic architecture of quantitative traits describes the characteristic features of broad-sense heritability or proportion of heritable total phenotypic variation, referring to the number of genes, genomic regions, the magnitude of gene effects, and their relative contributions to the additive, non-additive and epistatic

gene actions (Holland, 2007). Broad-sense heritability accounts for the total heritable variation or genotypic variation due to additive and non-additive genes. The narrow-sense heritability explains the proportion of genetic variation governed by additive genes. Broad-sense heritability is beneficial for the prediction of the genetic gain due to the selection of superior plant types in ground nursery, and narrow-sense heritability used to predict the better parental cross combination in hybridization programmes (Jackson, 2005). However, sugarcane has a long history of low narrow-sense heritability for cane yield (Wei and Jackson, 2017), a large proportion of genetic variability governed by non-additive genes (Yadav et al., 2021). Hence, parental selection based on meager breeding value does not help to enhance the genetic gain in sugarcane (Wei and Jackson, 2017). Very high broad-sense heritability was reported in sugarcane for stalk number, stalk diameter, brix, bagasse, fiber, and lignin content (Gouy et al., 2013). The make use of genome-wide markers and genomic selection aids in assessing the non-additive genetic variance in order to increase the frequency of favorable alleles in the breeding populations and selection of heterotic clones in sugarcane (Voss-Fels et al., 2021; Yadav et al., 2021).

GENOMIC SELECTION SCHEMES FOR SUGARCANE BREEDING PROGRAMMES

The genomic selection schemes applied at various stages of sugarcane breeding such as germplasm characterization and core germplasm formation, pre-breeding, and genetic base-broadening programmes, selection of parents for hybridization and prediction of progeny performance, varietal development and deployment are summarized in **Table 1**. The feasibility and suitability of these schemes in sugarcane breeding programmes are discussed.

Genomic Prediction and Core Collection in Field Gene Banks

The large number of germplasm collections maintained at gene banks have issues of germplasm duplication, the constraint of rejuvenation of a large number of accessions, unsatisfactory phenotyping, and utilization in breeding programmes (Diez et al., 2018). The core set formation is highly significant in resolving the problems of duplications and the ideal core collection is composed of 10% of the entire germplasm collection, representing 70% of the alleles of the entire germplasm (Brown, 1989). The various strategies such as stratified random sampling (Brown, 1989) and sampling based on multivariate clustering (Franco et al., 2005) were used in the core collection of germplasm. The molecular markers were also used in core germplasm formation in various crops (Zaharieva et al., 2001; Hao et al., 2006; Krichen et al., 2012; Dutta et al., 2015; Liu et al., 2015). The advent of next-generation sequencing technologies (NGS) and reduction in the cost of sequencing have given a new avenue for genome-wide SNP discovery and their utilization in genomic prediction in germplasm and core germplasm. Genome-wide marker-based germplasm characterization and core collection are validated in crops like wheat (Crossa et al.,

2016b; Kehel et al., 2020), sorghum (Yu et al., 2016), soybean (De Azevedo Peixoto et al., 2017), and cauliflower (Thorwarth et al., 2018).

The global collections of sugarcane germplasm are maintained at World Collections of Sugarcane and Related Grasses (WCSR), Canal Point, Florida, and the Indian Council of Agriculture Research (ICAR)-Sugarcane Breeding Institute, Coimbatore, India. About 1,002 accessions and 3,345 accessions of *Saccharum* germplasm are maintained at Canal point and Coimbatore, respectively (Amalraj and Balasundaram, 2006; Nayak et al., 2014). Besides, wild relatives are also maintained in Fiji, Brazil, Australia, China, and many other countries and are actively being used in the breeding programme (Ramdoyal and Badaloo, 2002; Wang et al., 2008; da Silva, 2017; Bhuiyan et al., 2019; Cursi et al., 2021). Several studies of core collection in *Saccharum* germplasm are described based on morphological features and molecular markers (Balakrishnan et al., 2000; Tai and Miller, 2001; Balakrishnan and Nair, 2003; Amalraj et al., 2006; Nayak et al., 2014; Shadmehr et al., 2017; Tena Gashaw et al., 2018; Fickett et al., 2019). The phenotypic characterization of *Saccharum* germplasm has many hurdles. First, tall and long duration crops are more prone to lodging and sucker development, which directly influences biomass and cane yield and juice quality parameters (Berding et al., 2005). Second, sugarcane is a photosensitive crop and spatial variation or morphological expressions are influenced by specific environmental regimes (Waldron et al., 1967; Gosnell, 1973; Pereira, 1983; Shanmugavadivu and Rao, 2009; Wei et al., 2010). Third many countries prohibited the field evaluation of *S. spontaneum* germplasm which are classified as noxious weed due to their fast-growing, spreading type with rhizomatous roots (Todd et al., 2017). Fourth, morphological characterization of germplasm based on fewer morphological descriptors and environmentally sensitive quantitative data does not reflect the complete genetic variability of *Saccharum* germplasm and unlike the utilization of genome-wide markers that helps to capture the complete genetic variability of genetic resources (Nybom and Lácis, 2021). Therefore, germplasm characterization and the formulation of core germplasm by using genome-wide markers could be the best option for crops like sugarcane which additionally captures the complete genetic variability.

In-depth pedigree analysis of the sugarcane germplasm has shown that only a limited number of basic *Saccharum* species clones were used in the development of breeding lines in sugarcane (Jackson, 2005; Kumar et al., 2012). These breeding lines were further advanced through recurrent selection and utilized in the varietal developmental programmes (Jackson, 2005). The pedigree analysis of Indian “Co” varieties (**Figure 1**) showed that only two *S. spontaneum* accessions viz., *S. spontaneum* CBE and *S. spontaneum* Java, were used in the breeding programmes (Kumar et al., 2012). Although there were efforts made to utilize the *Saccharum* germplasm in sugarcane hybridization programme for pre-breeding and base-broadening programmes (Wang et al., 2008; Mohanraj and Nair, 2014; da Silva, 2017; Nair et al., 2017; Cursi et al., 2021), still it was not a complete utilization of *Saccharum* germplasm or all the favorable alleles contributing to the cane yield and sucrose.

TABLE 1 | The application of genomic prediction and genomic selection in sugarcane breeding programmes.

Sl. no	Breeding programmes	Objectives	Similar studies in other crops
1	Characterization of <i>Saccharum</i> field gene bank by using genome-wide markers	To characterize the <i>Saccharum</i> field gene bank for genomic diversity, to establish identity of accessions and to verify the duplications in gene banks.	Maize (Lu et al., 2009) Rye (Bolibok-Bragoszewska et al., 2014) Apple (Buiteveld et al., 2021)
2	Core collections of <i>Saccharum</i> field gene bank	Core sampling of <i>Saccharum</i> and allied germplasm representing the total genomic diversity present in the gene bank.	Soybean (Jeong et al., 2019) Wheat (Pascual et al., 2020) Rice (Kumar et al., 2020a)
3.	Genomic prediction in <i>Saccharum</i> gene bank	To expedite the characterization of <i>Saccharum</i> germplasm for cane yield, commercial cane sugar (CCS) and tolerances to biotic and abiotic stresses.	Sorghum (Yu et al., 2016) Wheat landraces (Cossa et al., 2016b; Kehel et al., 2020) Soybean (Peixoto et al., 2016) Cauliflower (Thorwarth et al., 2018)
4.	Genome-wide association studies and genomic prediction	To identify the genomic regions associated with agronomic traits (cane yield and CCS) and, biotic and abiotic stress tolerances.	Eucalyptus (Müller et al., 2017) Winter wheat (Kristensen et al., 2019; Odilbekov et al., 2019) Wheat and Barley (Tsai et al., 2020) Maize (Liu et al., 2021)
5.	Pre-breeding or wide hybridization or nobilization	To introgress new alleles and genomic regions from <i>Saccharum</i> germplasm identified through germplasm characterization, GWAS and genomic prediction studies into the working germplasm.	Maize (Gorjanc et al., 2018) Eucalyptus (Tan et al., 2017) Pear (Kumar et al., 2019)
6.	Genetic base-broadening	To improve the breeding value of the genetic stocks identified from pre-breeding by backcrossing with parental lines or recurrent selection	Eucalyptus (Tan et al., 2017) Wheat (Singh et al., 2018) Pear (Kumar et al., 2019) Maize (Mayer et al., 2020)
7.	Recurrent selection breeding programmes (Gouy et al., 2013)	To augment the favorable alleles in the population/parental pool for cane yield, CCS, and tolerances through recurrent selection cycles	Rice (Grenier et al., 2015) <i>Brassica napus</i> (Zhao et al., 2016) Maize (Müller et al., 2017) Blueberry (Ferrão et al., 2021)
8.	Genomic prediction of parental cross combination and hybridization	Prediction of parental cross/progenies combination through estimated breeding values, general and specific combining ability of parental lines	Apple (Kumar et al., 2020b) Maize (Kadam et al., 2016; Jarquin et al., 2021; Liu et al., 2021) Wheat (Lado et al., 2017) Cassava (Wolfe et al., 2017)
9.	Progeny assessment and clonal selection (Deomano et al., 2020; Hayes et al., 2021; Yadav et al., 2021)	Prediction of superior plant types/progenies based on the broad-sense heritability or additive and non-additive genetic variance	Cassava (Wolfe et al., 2017) Eucalyptus (Resende et al., 2017) Sugarcane (Deomano et al., 2020; Hayes et al., 2021; Yadav et al., 2021) Oil palm (Nyouma et al., 2020) Macadamia nut (O'Connor et al., 2021)
10	Multi-environment trial or deployment of cultivars to target environments	To predict the genotype × environment interactions in multi-environment trials and to identify the stable varieties suitable for the target environment To predict the marker × environment interaction in multi-environment trials and identify the environment sensitive genomic regions Incorporation of environmental covariates into the genomic models and to predict the impact of environmental covariates on genotype performance and deployment of varieties.	Barley (Oakey et al., 2016) Chick pea (Roorkiwal et al., 2018) Maize (Schulz-Streeck et al., 2013) Wheat (Lopez-Cruz et al., 2015; Cossa et al., 2016a) Rice (Monteverde et al., 2019) Wheat (Heslot et al., 2014) Barley (Malosetti et al., 2016) Rice (Monteverde et al., 2019)

Therefore, characterization of the *Saccharum* germplasm with genome-wide markers and genome-wide association studies certainly identifies the alleles contributing to the cane yield, sucrose content, and other agronomic traits. The genome-wide markers capture the complete genetic variability present in germplasm and are also useful in the formulation of core germplasm, improving the precision of selection in gene-specific introgression in pre-breeding and enriching the favorable alleles

and genes in genetic base-broadening and recurrent breeding programmes. The two major concerns of a sugarcane breeder, as described by Jackson (2005) *viz.*, the concern of narrow genetic base and potential opportunities to broaden the genetic base by utilizing a large number of basic *Saccharum* germplasm, were easily addressed by adopting the genomic prediction in *Saccharum* gene bank and genomic selection schemes in pre-breeding and genetic base-broadening programmes. As described

in sorghum (Yu et al., 2016), *Saccharum* germplasm can be used for genomic prediction by considering the part of the germplasm as the training population and the remaining germplasm as the validation population.

Genomic Selection for Pre-breeding and Genetic Base-Broadening Programme

Pre-breeding and genetic base-broadening is the most important component of the sugarcane breeding programme followed in many countries (Ramdoyal and Badaloo, 2002; Wang et al., 2008; Matsuoka et al., 2014; da Silva, 2017; Cursi et al., 2021). The importance of *S. spontaneum* and other *Saccharum* spp. in the development of cultivars for bioenergy and commercial sugar production has been previously reviewed (Wang et al., 2008; Matsuoka et al., 2014; da Silva, 2017; Cursi et al., 2021) and novel genetic resources for disease resistance were identified (Bhuiyan et al., 2019). The utilization of *S. spontaneum* accession “Mandalay” in varietal development programmes resulted in many varieties in Australia (Reffay et al., 2005; Piperidis et al., 2021) and the high yield sugarcane cultivar LCP 85-384 has genetic lineage from *S. spontaneum* and *S. barberi* (Milligan et al., 1989). The studies involving genotyping of 400 markers in 232 biparental populations, which were derived from crosses involving Mandalay as a grandparent, exhibited that 25% of genomic regions originated from the “Mandalay” (Reffay et al., 2005). Consistent efforts were made to utilize wild genetic resources to broaden the genetic base of Indian working germplasm involving many *Saccharum* spp. such as *S. officinarum*, *S. robustum*, *S. spontaneum* as well as *E.arundinaceus*, and *Erianthus bengalense* (Moharaj and Nair, 2014; Ravinder et al., 2015; Nair et al., 2017). Nevertheless, phenotypic selection might not have captured the complete favorable alleles. Furthermore, the utilization of genome-wide markers, genome-wide association studies, and genomic selection helped in enhancing the frequency of favourable alleles and minimising the linkage drag, which are commonly associated with wide-hybridization and genetic base-broadening programmes in sugarcane (Roach, 1989).

Pre-breeding, a bridge between crop improvement and plant genetic resources, provides an opportunity for introgression of desirable genes with minimal linkage drag and enhances the adaptability of the cultivars (Sharma et al., 2013). It is combined with genomic-assisted selection which is helpful in the identification of climate-smart alleles/haplotypes in gene banks and in the development of climate-resilient varieties (Varshney et al., 2018). Moreover, it was also demonstrated in other clonally propagated crops (Tan et al., 2017; Kumar et al., 2019). In similar ways, the feasibility of genomic prediction in sugarcane pre-breeding has to be explored and the methodology is described in **Figure 2**. The pre-breeding in *Saccharum* spp. requires 3–4 generations of backcrossing with recurrent parents. Pre-breeding in sugarcane involves the following steps: (i) genome-wide association studies in germplasm and identification of elite accessions with more number of desirable genes/trait-specific accessions; (ii) hybridization between elite accessions with noble *S. officinarum* clones or commercial sugarcane varieties;

(iii) identification of true interspecific hybrids through genomic/cytoplasmic/5S rRNA/Inter Transcriber Spacer specific markers; (iv) optimization of the training and testing populations at each generation or back cross programmes which generally requires 3–4 generations of backcrossing/crossing with commercial sugarcane varieties; (v) genomic prediction model building and retraining of the genomic prediction models to optimize and predict the genotype with the highest Genomic Estimated Breeding Value (GEBV) for utilization in the next generation of backcrossing or crossing. The genomic prediction model could also be able to help in the swift development of genetic stocks in pre-breeding and genetic base-broadening in sugarcane.

Genomic Selection for Superior Parental Cross Combination and Clonal Selection at an Early Stage

The varietal development programme in sugarcane breeding is a kind of recurrent selection. The parental lines for hybridizations are selected based on breeding values estimated from the performance of families or progenies of proven crosses (Chang and Milligan, 1992; Jackson, 2005; Burnquist et al., 2010; Lingle et al., 2010; Stringer et al., 2010; Mendes de Paula et al., 2020). Sugarcane is a clonally propagated crop, amenable for the exploitation of both additive and non-additive gene action. The additive genetic variance determines the proportion of genetic variability transmitted to progenies or breeding value. This is helpful in the selection of an elite parental pool and prediction of parental cross combinations. The proportion of non-additive genetic variance is suitable for the selection of elite plant types from the ground nursery (Jackson, 2005; Mendes de Paula et al., 2020). A very high significant Specific Combining Ability (SCA) variance is observed for cane yield in an unselected population of 35 families indicating the predominance of non-additive genetic variance (Zhou, 2020). The evaluation of 79 families derived from 38 parental lines revealed the predominance of non-additive genetic variance for cane yield and additive genetic variance for brix and sucrose (Mendes de Paula et al., 2020). Sugarcane has a long history of low narrow-sense heritability and a higher proportion of non-additive genetic variance for cane yield as compared with CCS (Wei and Jackson, 2017; Mendes de Paula et al., 2020; Zhou, 2020; Yadav et al., 2021). Hence, the selection of parental lines based on breeding value does result in a low genetic gain and require due consideration for non-additive genetic variance in the choice of a parental pool for hybridization programmes.

The breeding programme in sugarcane is aimed at enhancing the genetic gain for CCS and cane yield. There is no strong correlation between cane yield and CCS. The absence of a large correlation between the cane yield and CCS indicates that either gene governing these traits are independent or acting pleiotropically (Jackson and McRae, 2001; Zaharieva et al., 2001). Both traits are polygenic traits, governed by many genes with minor effects (Hoarau et al., 2002; Ming et al., 2002; Aitken et al., 2008), and the genetic gain through phenotypic selection is low, as the probability of progenies with combinations of

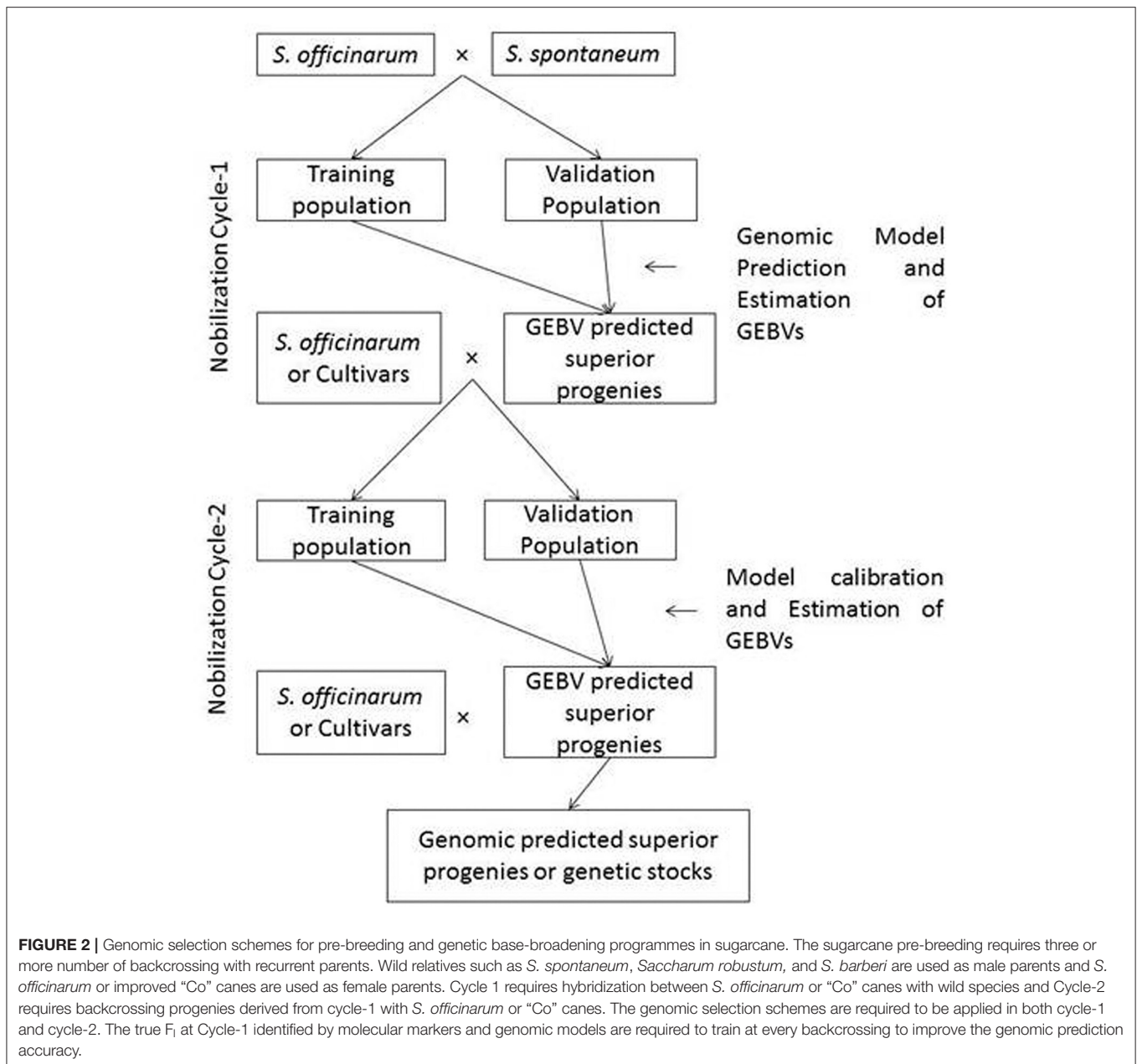


FIGURE 2 | Genomic selection schemes for pre-breeding and genetic base-broadening programmes in sugarcane. The sugarcane pre-breeding requires three or more number of backcrossing with recurrent parents. Wild relatives such as *S. spontaneum*, *Saccharum robustum*, and *S. barberi* are used as male parents and *S. officinarum* or improved “Co” canes are used as female parents. Cycle 1 requires hybridization between *S. officinarum* or “Co” canes with wild species and Cycle-2 requires backcrossing progenies derived from cycle-1 with *S. officinarum* or “Co” canes. The genomic selection schemes are required to be applied in both cycle-1 and cycle-2. The true F_1 at Cycle-1 identified by molecular markers and genomic models are required to train at every backcrossing to improve the genomic prediction accuracy.

superior alleles for sucrose and cane yield is very low (Jackson, 2005). The recurrent selection combined with family evaluation increases the probability of the accumulation of favorable alleles in the population and the chances of selecting the superior plant types combined with both cane yield and CCS is high (Chang and Milligan, 1992; Lingle et al., 2010). The genomic prediction schemes are extensively used in other crops for estimation of non-additive genetic variance and prediction of the best of parental cross combination based on their general and specific combining ability variances and effects (Kadam et al., 2016; Lado et al., 2017; Wolfe et al., 2017; Jarquin et al., 2020; Wang et al., 2020). The encouraging results of genomic prediction in recurrent selection schemes are also

reported in many crops (Zhao et al., 2016; Ferrão et al., 2021). Recently, the genomic selection for recurrent selection schemes suitable to sugarcane was reviewed (Yadav et al., 2020) and encouraging results for estimating the breeding values based on additive genetic variance with high prediction accuracies are demonstrated (Gouy et al., 2013; Deomano et al., 2020; Hayes et al., 2021). The simulation and empirical research have demonstrated the significant contribution of non-additive genetic variance and genome-wide heterozygosity for cane yield and CCS in sugarcane (Voss-Fels et al., 2021; Yadav et al., 2021).

At present, the parental cross combination in sugarcane is selected based on breeding values and past proven crosses. A large number of crosses are made and 30–50,000 of progenies

are evaluated for identification of superior plant types combined with high cane yield, CCS, and resistance to many diseases like smut, red rot, and yellow leaf disease and tolerances to abiotic stresses (Park et al., 2007; Nair, 2011). The clonal trials are evaluated in single row experiments and advanced trials in multiple row experiments. The evaluation and selection in the ground nursery and clonal trials have many technical and logistics problems. First, a large portion of genetic variability for CCS is governed by additive genes and there is a high correlation between early and advanced clonal selection cycles. For cane yield, the non-additive genetic variance contributes significantly to the genetic variance and is highly influenced by environmental factors. The correlation between the smaller and larger plot experiments is very low for the cane yield as compared with CCS due to field competitive effects. Hence, the selection of clones in the ground nursery and early clonal trials are less reliable (Jackson and McRae, 2001). Second, the lodging and suckering propensity in the ground nursery and clonal trials are major hurdles for breeders in choosing the superior plant types. Third, trait measurement for sucrose content varies with a sound cane, i.e., millable cane developed from first formed tillers, and non-sound canes, i.e., millable canes developed from late emerged tillers. For example, CCS is highly variable with sound and non-sound canes (Berding et al., 2005; Jackson, 2005). Fourth, the crop expression varies with agro-climatic regions, such as tropical and subtropical regions and the location-specific ground nursery essentially requires for identification of location-specific varieties (Park et al., 2007; Nair, 2011). Considering all these factors, genomic selection could assist the breeders in the selection of potential genotypes that combine with the high cane yield, CCS, and tolerance to biotic and abiotic stresses. With the accurate multi-environment phenotyping of the parental pool and populations derived from recurrent selection schemes, the genomic prediction could help in identifying the plant types suitable to different agro-ecological regions and even a common phenotype suitable across zones. Additionally, it helps in reducing the population size to be evaluated in the ground nursery and subsequent clonal trials. Genomic selection is an evolving science, and genomic prediction models fitted with additive and non-additive components and genome-wide heterozygosity has demonstrated its efficiency in sugarcane (Gouy et al., 2013; Deomano et al., 2020; Hayes et al., 2021; Voss-Fels et al., 2021; Yadav et al., 2021). Further developments in data science, training, and retraining of genomic statistical models will impart benefits to the sugarcane breeders in the selection of the superior cross combinations and varieties in sugarcane.

Genomic Predictions for Environment Covariates and Varietal Deployment

The genotypes developed from the breeding programmes are adapted to the specific environments and the differential performance of genotypes in different environments is due to genotype \times environment ($G \times E$) interactions. The repeatable $G \times E$ factors in multi-environment trials are estimated through single dimension based statistical models such as ANOVA and

regression approaches. The ANOVA based statistical model partitioning the $G \times E$ interactions into the main effect of genotypes, environments and their interactions as fixed effects/variables and residuals/error components as random variables (Elias et al., 2016). The linear regression-based $G \times E$ interaction (Finlay and Wilkinson, 1963) introduce the slopes for environmental means and identifies the genotypes based on the slope and trait mean value. The $G \times E$ interactions caused by several factors (Gauch, 2006) and statistical models *viz.*, additive main effects and multiplicative interaction effects model (AMMI), site regression model (SREG), shifted multiplicative model (SHMM), genotype regression model (GREG), and completely multiplicative model (COMM), are used for partitioning of the $G \times E$ interactions into more than one factor (Elias et al., 2016). The linear and multiplicative models that treat the genotypes and environments as fixed effects do not consider the heterogeneity of variance. They are suitable only for the balanced dataset as they do not consider the variable replication numbers, which can be accommodated under mixed effect models.

The mixed model framework permits the utilization of correlated heterogeneous variance under the variance-covariance structure of random variables and estimates the association between known environmental and genetic parameters. Factor regression is used for the analysis of multi-environment trials and statistically helps to test the sensitivity of genotypes with environmental covariates (Gauch et al., 2008). The partial least square regression analysis on $G \times E$ interaction allows the use of environmental covariates as independent variables to find the most influential environmental variables contributing to the interactions (Vargas et al., 1998). The factorial regression analysis offer aid in integrating the environmental covariates into the model and determine the most influential environmental parameters affecting the trait/QTL expression (Malosetti et al., 2004).

The genomic prediction models can be applied to multi-environment trials, which require characterization of $G \times E$ interactions over locations, environments and years. The covariance structures of markers, pedigree, environmental covariates, marker \times environmental covariates are incorporated into the genomic models to predict the $G \times E$ interactions, environmental sensitivity of the genotypes, and identify the environment-sensitive genomic regions/QTL contributing to the $G \times E$ interaction and assessing the prediction accuracy in untested environments (Cossa et al., 1999; Burgueño et al., 2012; Schulz-Streeck et al., 2013; Heslot et al., 2014; Montesinos-López et al., 2016; Oakey et al., 2016; Monteverde et al., 2019). Multiple environment-genomic prediction models demonstrated the prediction of the genome-wide environment-specific genomic regions and the model envisaged the environmental covariates associated with phenology and crop growth models (Malosetti et al., 2016).

The differential response of genotypes to different environments is highly evident in sugarcane such as: (i) sandy soils have high discriminative powers in differentiating the environments and genotypes (Glaz and Kang, 2008); (ii) multi-environment trials involving seven locations revealed

the significance of genotype \times location interaction, which is higher than genotype \times crop year interaction and inferred that testing genotypes across locations are more important than repeating the trials in the same locations (Guilly et al., 2017); (iii) time of ratooning/planting and harvesting season contributes significantly to the G \times E interactions for cane yield and sugar yield (Gilbert et al., 2006), partly due to the non-repeatable interactions (Ramburan, 2014); (iv) spatial variation also significantly contributes to the G \times E interactions in sugarcane (Wei et al., 2010). Therefore, multi-environment trials play a crucial role in the selection of stable region-specific better-performing varieties suitable for cultivation across locations. In the Australian sugarcane varietal development programme, the final stage of clonal selection is chosen through evaluation in the final assessment trials conducted in four agro-climatic zones *viz.*, Northern, Burdekin, Central, and Southern regions (Park et al., 2007). The Louisiana varietal identification programme involves replicating the Outfield Variety Tests in many locations (Breaux, 1984). The Brazilian breeding programme under Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESEA, Brazil) network requires assessment in multiple locations and harvests in the final phase of recommendation of clones (Barbosa et al., 2012). In India, sugarcane growing regions are broadly categorized into five different agro-climatic regions or zones, namely Peninsular, North-West, North-East, North-Central, and East-Coast regions, wherein each zone has its multi-environment trials for final varietal identification (Nair, 2011). The implementation of genomic schemes by integrating with environment covariates for the prediction of clonal performance in each agro-climatic zones certainly helps in the identification of the environment-specific genomic regions and superior clones combined with cane yield, CCS, and tolerances to biotic and abiotic stresses. The genomic prediction for clonal performance in different locations is validated for cane yield, CCS, and fiber content with high prediction accuracy (Yadav et al., 2021) and further research is required to address the challenges of genomic modeling to accommodate more parameters related to environmental parameters.

GENETIC GAIN PER UNIT COST AND UNIT TIME, AND RESOURCE ALLOCATION

The genomic selection reduces the number of selection cycles and reduce the duration of the breeding cycles. The genetic gain in breeding has to be assessed in terms of gain per unit time and cost rather than the gain per cycle (Bernardo, 2008). The superiority of genetic gain due to genomic selection over the phenotypic selection and marker-assisted selection with the reduced cost and duration of selection was demonstrated in many crops (Bernardo and Yu, 2007; Wong and Bernardo, 2008; Yabe et al., 2018). The genomic selection minimizes the phenotyping and maximizes the genotyping and is worth considering when the phenotyping cost is much higher than genotyping (Bernardo and Yu, 2007). In Indian conditions, the cost of genotyping through 50K SNP chips and genotyping by sequencing is around

Rs. 10,500.00 per sample. The cost of genotyping in a ground nursery or progeny assessment trials, which accommodates 12,000 seedlings per ha costs about Rs. 12.60 crores, the First Clonal Trial (4,800 clones/ha) costs around Rs. 5.04 crores, Second Clonal Trial (1,200 clones/ha) costs around 1.26 crores, Pre-Zonal Varietal Trial (150 entries with two replications per ha) costs around Rs.15.75 lakh, and AICRP Trials (50 entries with three replications per ha) costs around Rs. 5.25 lakhs. Therefore, the genotyping cost in India is still higher and highly expensive as compared with phenotypic selections and similar reports of the higher genotyping cost of the SNP array (\$95/sample) was also reported in Australia (Voss-Fels et al., 2021). Yet, the superiority of genetic gain through genomic selection as compared with phenotypic selection is significantly high in terms of net return and scientifically embracing to implementation of the genomic selection at the progeny assessment stage (Beyene et al., 2019; Voss-Fels et al., 2021). The genomic-wide selection for the estimation of GEBVs for germplasm, lines derived from pre-breeding, parental lines, progeny performance in progeny assessment trials and multi-environment trials is essentially required to enhance the genetic gain in sugarcane.

Optimization and efficient resource allocations in the genomic selection are necessary for improving genetic gain and prediction efficiency. Strategies for resource allocation depends on the role of genomic selection in breeding population, such as (i) size, composition and design of training and validation populations, (ii) plot size for breeding experiments varies with stages of breeding programmes, such as parental selection, early and advanced testing, and (iii) the number of replications within and across the environments (Lorenz and Nice, 2017). The composition of the training population is determined based on the objective of genomic selection schemes. The training and validation populations could be diverse germplasm, elite parental lines of diverse origin, segregating breeding lines from the same cross or diverse biparental population, early testing of breeding lines and multi-environment trials (Windhausen et al., 2012; Yu et al., 2016; Roorkiwal et al., 2018; Ozimati et al., 2019). Moreover, each study has its significance in plant breeding and demands different strategies for resource allocation.

The genomic selection in the early stages of segregating the population is a trade-off between the cost of genomic selection and trait heritability and family size. The phenotypic selection in a large segregating population with highly heritable traits is highly efficient, whereas low heritable traits, such as yield in a large segregating population, require the molecular markers to assist in the selection of desirable plant types. Further, yield measured on the single plant basis in segregating population is inaccurate in most of the crops (Heslot et al., 2015). Even after the selection of elite plant type based on GEBV, phenotypic evaluation is needed to remove the inferior genotype if any developed due to the large allelic mutation or rare alleles becoming homozygous (Hayes et al., 2009). Considering the crop duration, crop architecture, and other field related problems, such as lodging and suckering propensity, variation in trait value due to sampling errors in sound and non-sound canes, field competitive effects (Jackson and McRae, 2001), the simulation studies demonstrated the economic benefits of implementing

the genomic selection in sugarcane (Voss-Fels et al., 2021), and genomic prediction accuracies are also experimentally validated (Yadav et al., 2021).

The recurrent selection procedure is effective in augmenting the frequency of favorable alleles in the breeding population. The marker- or genomic-assisted selection helps to reduce the number of breeding cycles and the number of individuals evaluated in each selection cycle. The marker-assisted recurrent selection scheme aims at gene pyramiding after identifying the QTL (Servin et al., 2004), whereas the genomic assisted recurrent selection scheme aimed at gene pyramiding without identification of QTL (Bernardo and Yu, 2007; Bernardo, 2009). The genomic-assisted recurrent selection could reduce the prediction accuracies after several cycles (Muleta et al., 2019) and requires genomic remodeling at each cycle of selection, which may add to cost (Heslot et al., 2015). The marker-assisted selection schemes operate with fewer markers on the biparental population while the other operates with a large number of markers, further adding to the cost (Heslot et al., 2015). Even after the selection of elite plant type by using GEBVs, phenotypic evaluation is needed to remove the inferior genotype if any developed due to the large allelic mutation or rare alleles becoming homozygous (Hayes et al., 2009). The genomic assisted recurrent selection accelerates the genetic gain over a shorter period in both small and larger populations. The cost per unit gain is lower for oligogenic traits in smaller populations and polygenic traits in large populations as compared with phenotypic recurrent selection (Muleta et al., 2019). The recurrent selection cycles in sugarcane operate in two ways. First, the selection of ideal plant type from the ground nursery, vigorous testing for the identification of promising clones and further utilization in hybridization programme forms a kind of random recurrent selection cycle. Second, the 2–3 recurrent selection cycles maximize the genetic gain or trait values (Shanthi et al., 2008; Yadav et al., 2020), and the simulation study has validated the cost-benefit ratio and resource allocation to maximize the genetic gain (Voss-Fels et al., 2021).

The multi-environment trials are the final phase of cultivar identification and need robust statistical analysis with optimum resource allocation. The resource allocation depends on the number of replications, locations, number of test entries, etc. An adequate number of replications are required for each location to control the microenvironment variations and experimental design such as Alpha design, which is very efficient in accounting for the micro-environment variation (González-Barríos et al., 2019). The optimal number of replications and test locations require optimization based on historical data and the number of mega environments estimated from genotype + genotype × environment (GGE) biplot analysis (Yan, 2015; Yan et al., 2015). The four replications were found optimum based on GGE biplot analysis of multi-environmental trials (Baxevanos et al., 2017). The very high significant genotype × location interactions and very low interactions for genotype × crop and genotype × year components were observed. Based on the ranking of genotypes from all replicates and two replicates in multi-environment trials, inferred that two replications did not affect the precision of selections and save 33–50% of the

experimental area (Yan, 2021). The extremely low variance for genotype × replication interactions was observed and a reduction in the number of replications from eight to four did not reduce the precision of the experiments in sugarcane (Brown and Glaz, 2001). The historic mega-environments are useful in differentiating the genotypes and helps in resource allocation. The moderate mega-environment are highly useful to achieve the optimum genetic gain (González-Barríos et al., 2019).

The unbalanced design or sparse testing refers to the multi-environment trials where all the genotypes are not tested in all environments and sparse testing saves the resources. Unbalanced sparse testing with an appropriate genomic prediction model enhances the prediction accuracies and saves the resources (Endelman et al., 2014; González-Barríos et al., 2019; Jarquin et al., 2020). The resource allocation depends on many factors such as the genetic relationship of the training population, genomic prediction model calibration, target environments, number of replications, and many other factors (Lorenz and Nice, 2017). Several factors such as complex polyploid and its impact on genotyping, crop agronomy, crop architecture, crop duration, biomass, sucrose accumulation pattern, and other parameters necessitate consideration for genomic selection. The empirical studies are required to develop a comprehensive theoretical framework on resource allocations in genomic selection in sugarcane.

CONCLUSIONS

Sugarcane is a C4 crop with a great potential for high biomass production and a major source of raw material for sugar production and bioenergy. Its crop improvement has many bottlenecks such as the long cycle of breeding duration, complex polyploidy, high degree of heterozygosity, narrow genetic base, and fewer basic germplasm utilized in pre-breeding and linkage drag during wide hybridization and limited financial and manpower resources. The genomic selection schemes are highly helpful in reducing the duration of breeding cycles, population size, and selection of desirable parents and development of varieties. The characterization of *Saccharum* germplasm with genome-wide markers captures the total genomic diversity of gene bank and assists in the formation of core germplasm and genomic prediction to identify the genes associated with cane yield, CCS, and tolerances to biotic and abiotic stresses. The genomic selection in pre-breeding and genetic base-broadening programmes helps in the precise introgression of genes into the parental clones and genomic-assisted recurrent selection is useful in augmenting the favorable alleles in the population. The genomic prediction for characterization of parental clones for their general and specific combining ability leverage the breeders in selecting the elite parents and progenies combined with favorable alleles for cane yield, CCS, and tolerances to biotic and abiotic stresses. The integration of environmental covariates into genomic models predicts the better performing varieties for target environments and deployment of varieties. Genomic selection is an evolving science; appropriate genomic models

and breeding strategies strengthen the prediction accuracies and enhance the genetic gain in sugarcane.

AUTHOR CONTRIBUTIONS

CM conceptualized, wrote, and reviewed the draft. KA wrote, revised, provided inputs on molecular markers, SNPs, genomic selections, and resource allocations. CA, GS, PV, HM, RV,

GH, GA, and BR provided inputs on molecular markers, genomic schemes, and resources allocations. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

We are grateful to the ICAR-Sugarcane Breeding Institute, Coimbatore, India for supporting this study.

REFERENCES

- Aitken, K. S., Farmer, A., Berkman, P., Muller, C., Wei, X., Demano, E., et al. (2016). Generation of a 345 K sugarcane SNP chip. *Int. Soc. Sugar Cane Technol. Proc. Congr.* 29, 1165–1172. Available online at: <https://issct.org/wp-content/uploads/proceedings/2016/Molecular-Biology-papers/document-4-%20Molecular%20Biology%20Papers%202016.pdf>
- Aitken, K. S., Hermann, S., Karno, K., Bonnett, G. D., McIntyre, L. C., and Jackson, P. A. (2008). Genetic control of yield related stalk traits in sugarcane. *Theoret. Appl. Genet.* 117, 1191–1203. doi: 10.1007/s00122-008-0856-6
- Amalraj, V. A., Balakrishnan, R., Jebadhas, A. W., and Balasundaram, N. (2006). Constituting a core collection of *Saccharum spontaneum* L. and comparison of three stratified random sampling procedures. *Genet. Resour. Crop Evol.* 53, 1563–1572. doi: 10.1007/s10722-005-8510-5
- Amalraj, V. A., and Balasundaram, N. (2006). Status of sugar-cane genetic resources in India. *Plant Genet. Resources Newsletter* 148, 26–31. Available online at: https://www.bioversityinternational.org/fileadmin/PGR/article-issue_148-art_4-lang_en.html
- Baez-Gonzalez, A. D., Kiniry, J. R., Meki, M. N., Williams, J., Alvarez-Cilva, M., Ramos-Gonzalez, J. L., et al. (2017). Crop parameters for modeling sugarcane under rainfed conditions in Mexico. *Sustainability* 9:1337. doi: 10.3390/su9081337
- Balakrishnan, R., Nair, N. V., and Sreenivasan, T. V. (2000). A method for establishing a core collection of *Saccharum officinarum* L. Germplasm based on quantitative-morphological data. *Genet. Resources Crop Evolut.* 47, 1–9. doi: 10.1023/A:1008780526154
- Balakrishnan, R., and Nair, N. V. (2003). Strategies for developing core collections of sugarcane (*Saccharum officinarum* L.) germplasm- comparison of sampling from diversity groups constituted by three different methods. *PGR Newsletter FAO Biodiversity* 134, 33–41. Available online at: https://www.bioversityinternational.org/fileadmin/PGR/article-issue_134-art_108-lang_es.html
- Balsalobre, T. W. A., da Silva Pereira, G., Margarido, G. R. A., Gazaffi, R., Barreto, F. Z., Anoni, C. O., et al. (2017). GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18, 1–19. doi: 10.1186/s12864-016-3383-x
- Barbosa, M. H. P., Resende, M. D. V., Bressiani, J. A., Silveira, L. C. I., and Peternelli, L. A. (2005). Selection of sugarcane families and parents by Reml/Blup. *Cropp Breeding Appl. Biotechnol.* 5, 443–450. doi: 10.12702/1984-7033.v05n04a10
- Barbosa, M. H. P., Resende, M. D. V., Dias, L. A., Barbosa, G. V., Oliveira, R. A., Peternelli, L. A., et al. (2012). Genetic improvement of sugar cane for bioenergy: the brazilian experience in network research with RIDESA. *Crop Breed. Appl. Biotechnol.* 12, 87–98. doi: 10.1590/S1984-70332012000500010
- Barreto, F. Z., Bacheга Feijó Rosa, J. R., Almeida Balsalobre, T. W., 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:e219843. doi: 10.1371/journal.pone.0219843
- Baxevas, D., Tsialtas, J. T., Vlachostergios, D., and Goulas, C. (2017). Optimum replications and locations for cotton cultivar trials under Mediterranean conditions. *J. Agricult. Sci.* 155, 1553–1564. doi: 10.1017/S0021859617000648
- Berding, N., Hurney, A. P., Salter, B., and Bonnett, G. D. (2005). Agronomic impact of sucker development in sugarcane under different environmental conditions. *Field Crops Res.* 92, 203–217. doi: 10.1016/j.fcr.2005.01.013
- Berkman, P. J., Bundock, P. C., Casu, R. E., Henry, R. J., Rae, A. L., and Aitken, K. S. (2014). A survey sequence comparison of saccharum genotypes reveals allelic diversity differences. *Trop. Plant Biol.* 7, 71–83. doi: 10.1007/s12042-014-9139-3
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Sci.* 48, 1649–1664. doi: 10.2135/cropsci2008.03.0131
- Bernardo, R. (2009). Genomewide selection for rapid introgression of exotic germplasm in maize. *Crop Sci.* 49, 419–425. doi: 10.2135/cropsci2008.08.0452
- Bernardo, R., and Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47, 1082–1090. doi: 10.2135/cropsci2006.11.0690
- Berro, I., Lado, B., Nalin, R. S., Quincke, M., and Gutiérrez, L. (2019). Training population optimization for genomic selection. *Plant Genome* 12:190028. doi: 10.3835/plantgenome2019.04.0028
- Beyene, Y., Gowda, M., Olsen, M., Robbins, K. R., Pérez-Rodríguez, P., Alvarado, G., et al. (2019). Empirical comparison of tropical maize hybrids selected through genomic and phenotypic selections. *Front. Plant Sci.* 10:150. doi: 10.3389/fpls.2019.01502
- Bhat, S. R., and Gill, S. S. (1985). The implications of 2n egg gametes in nobilization and breeding of sugarcane. *Euphytica* 34, 377–384. doi: 10.1007/BF00022932
- Bhuiyan, S. A., Garlick, K., and Piperidis, G. (2019). *Saccharum spontaneum*, a novel source of resistance to root-lesion and root-knot nematodes in sugarcane. *Plant Dis.* 103, 2288–2294. doi: 10.1094/PDIS-02-19-0385-RE
- Bolibok-Bragoszewska, H., Targońska, M., Bolibok, L., Kilian, A., and Rakoczy-Trojanowska, M. (2014). Genome-wide characterization of genetic diversity and population structure in Secale. *BMC Plant Biol.* 14, 1–15. doi: 10.1186/1471-2229-14-184
- Breaux, R. D. (1984). Breeding to enhance sucrose content of sugarcane varieties in Louisiana. *Field Crops Res.* 9, 59–67. doi: 10.1016/0378-4290(84)90006-6
- Brown, A. H. D. (1989). Core collections: A practical approach to genetic resources management. *Genome* 31, 818–824. doi: 10.1139/g89-144
- Brown, A. H. D., Daniels, J., and Latter, B. D. H. (1968). Quantitative genetics of sugarcane. *Theor. Appl. Genet.* 38, 361–369. doi: 10.1007/BF00934169
- Brown, J. S., and Glaz, B. (2001). Analysis of resource allocation in final stage sugarcane clonal selection. *Crop Sci.* 41, 57–62. doi: 10.2135/cropsci2001.41157x
- Buiteveld, J., Koehorst-van Putten, H. J., Kodde, L., Laros, I., Tumino, G., Howard, N. P., et al. (2021). Advanced genebank management of genetic resources of European wild apple, *Malus sylvestris*, using genome-wide SNP array data. *Tree Genet. Genomes* 17:32. doi: 10.1007/s11295-021-01513-y
- Burgueño, J., de los Campos, G., Weigel, K., and Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52, 707–719. doi: 10.2135/cropsci2011.06.0299
- Burnquist, W. L., Redshaw, K., and Gilmour, R. F. (2010). Evaluating sugarcane R&D performance: evaluation of three breeding programs. *Proc. Int. Soc. Sugar Cane Technol.* 27, 1–15. Available online at: <https://issct.org/wp-content/uploads/proceedings/2010/2010%20Papers.pdf>
- Chandel, A. K., da Silva, S. S., Carvalho, W., and Singh, O. V. (2012). Sugarcane bagasse and leaves: Foreseeable biomass of biofuel and bio-products. *J. Chem. Technol. Biotechnol.* 87, 11–20. doi: 10.1002/jctb.2742
- Chang, Y. S., and Milligan, S. B. (1992). Estimating the potential of sugarcane families to produce elite genotypes using bivariate prediction methods. *Theor. Appl. Genet.* 84, 633–639. doi: 10.1007/BF00224162

- Clark, S. A., Hickey, J. M., Daetwyler, H. D., and van der Werf, J. H. J. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Select. Evolut.* 44:4. doi: 10.1186/1297-9686-44-4
- Clevenger, J., Chavarro, C., Pearl, S. A., Ozias-Akins, P., and Jackson, S. A. (2015). Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations. *Mol. Plant* 8, 831–846. doi: 10.1016/j.molp.2015.02.002
- Crossa, J., De Los Campos, G., Maccaferri, M., Tuberosa, R., Burgueño, J., and Pérez-Rodríguez, P. (2016a). Extending the marker \times Environment interaction model for genomic-enabled prediction and genome-wide association analysis in durum wheat. *Crop Sci.* 56, 2193–2209. doi: 10.2135/cropsci2015.04.0260
- Crossa, J., Jarquín, D., Franco, J., Pérez-Rodríguez, P., Burgueño, J., Saint-Pierre, C., et al. (2016b). Genomic prediction of gene bank wheat landraces. *G3: Genes Genomes Genet.* 6, 1819–1834. doi: 10.1534/g3.116.029637
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011
- Crossa, J., Vargas, M., Van Eeuwijk, F. A., Jiang, C., Edmeades, G. O., and Hoisington, D. (1999). Interpreting genotype \times environment interaction in tropical maize using linked molecular markers and environmental covariables. *Theor. Appl. Genet.* 99, 611–625. doi: 10.1007/s001220051276
- Cursi, D. E., Hoffmann, H. P., Barbosa, G. V. S., Bressiani, J. A., Gazaffi, R., Chapola, R. G., et al. (2021). History and current status of sugarcane breeding, germplasm development and molecular genetics in Brazil. *Sugar Tech.* doi: 10.1007/s12355-021-00951-1
- da Silva, J. A. (2017). The Importance of the Wild Cane *Saccharum spontaneum* for Bioenergy Genetic Breeding. *Sugar Tech* 19, 229–240. doi: 10.1007/s12355-017-0510-1
- Daetwyler, H. D., Kemper, K. E., van der Werf, J. H. J., and Hayes, B. J. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *J. Anim. Sci.* 90, 3375–3384. doi: 10.2527/jas.2011-4557
- De Azevedo Peixoto, L., Moellers, T. C., Zhang, J., Lorenz, A. J., Bhering, L. L., Beavis, W. D., et al. (2017). Leveraging genomic prediction to scan germplasm collection for crop improvement. *PLoS ONE* 12:179191. doi: 10.1371/journal.pone.0179191
- 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
- Diez, M. J., De la Rosa, L., Martín, I., Guasch, L., Cartea, M. E., Mallor, C., et al. (2018). Plant genebanks: Present situation and proposals for their improvement. The case of the Spanish network. *Front. Plant Sci.* 871, 1–13. doi: 10.3389/fpls.2018.01794
- Dutta, M., Phogat, B. S., Kumar, S., Kumar, N., Kumari, J., Pandey, A. C., et al. (2015). “Advances in Wheat genetics: from genome to field,” in: *Advances in Wheat Genetics: From Genome to Field. Proceedings of the 12th International Wheat Genetics Symposium*, eds Y. Ogihara, S. Takumi, and H. Handa (Springer Japan) 33–45.
- Edmé, S. J., Miller, J. D., Glaz, B., Tai, P. Y. P., and Comstock, J. C. (2005). Genetic contribution to yield gains in the Florida sugarcane industry across 33 years. *Crop Sci.* 45, 92–97. doi: 10.2135/cropsci2005.0092
- Elias, A. A., Robbins, K. R., Doerge, R. W., and Tuinstra, M. R. (2016). Half a century of studying genotype \times Environment interactions in plant breeding experiments. *Crop Sci.* 56, 2090–2105. doi: 10.2135/cropsci2015.01.0061
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:19379. doi: 10.1371/journal.pone.0019379
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Endelman, J. B., Atlin, G. N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M. E., et al. (2014). Optimal design of preliminary yield trials with genome-wide markers. *Crop Sci.* 54, 48–59. doi: 10.2135/cropsci2013.03.0154
- FAOSTAT (2019). *FAOSTAT Crops and Livestock Products*. Food and Agriculture Organization of the United Nations. <http://www.fao.org/faostat/en/#data/QLCL> (accessed September 7, 2021).
- Ferrão, L. F. V., Amadeu, R. R., Benevenuto, J., de Bem Oliveira, I., and Munoz, P. R. (2021). Genomic selection in an outcrossing autotetraploid fruit crop: lessons from blueberry breeding. *Front. Plant Sci.* 12:676326. doi: 10.3389/fpls.2021.676326
- Fickett, N., Gutierrez, A., Verma, M., Pontif, M., Hale, A., Kimbeng, C., et al. (2019). Genome-wide association mapping identifies markers associated with cane yield components and sucrose traits in the Louisiana sugarcane core collection. *Genomics* 111, 1794–1801. doi: 10.1016/j.ygeno.2018.12.002
- Finlay, K. W., and Wilkinson, G. N. (1963). The analysis of adaptation in plant breeding programme. *Aust. J. Agric. Res.* 14, 742–754. doi: 10.1071/AR9630742
- Franco, J., Crossa, J., Taba, S., and Shands, H. (2005). A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Sci.* 45, 1035–1044. doi: 10.2135/cropsci2004.0292
- Gauch, H. G. (2006). Statistical analysis of yield trials by AMMI and GGE. *Crop Sci.* 46, 1488–1500. doi: 10.2135/cropsci2005.07-0193
- Gauch, H. G., Piepho, H. P., and Annicchiarico, P. (2008). Statistical analysis of yield trials by AMMI and GGE: Further considerations. *Crop Sci.* 48, 866–889. doi: 10.2135/cropsci2007.09.0513
- Gilbert, R. A., Shine, J. M., Miller, J. D., Rice, R. W., and Rainbolt, C. R. (2006). The effect of genotype, environment and time of harvest on sugarcane yields in Florida, USA. *Field Crops Res.* 95, 156–170. doi: 10.1016/j.fcr.2005.02.006
- Glaz, B., and Kang, M. S. (2008). Location contributions determined via GGE biplot analysis of multi-environment sugarcane genotype-performance trials. *Crop Sci.* 48, 941–950. doi: 10.2135/cropsci2007.06.0315
- González-Barrios, P., Díaz-García, L., and Gutiérrez, L. (2019). Mega-environmental design: Using genotype \times environment interaction to optimize resources for cultivar testing. *Crop Sci.* 59, 1899–1915. doi: 10.2135/cropsci2018.11.0692
- Gorjanc, G., Gaynor, R. C., and Hickey, J. M. (2018). Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor. Appl. Genet.* 131, 1953–1966. doi: 10.1007/s00122-018-3125-3
- Gosnell, J. M. (1973). “Some factors affecting flowering in sugarcane,” in *Proceedings of The South African Sugar Technologists’ Association* (Mount Edgecombe), 144–147. Available online at: https://sasta.co.za/download/10/1970-1979/3930/1973_gosnell_some-factors-affecting-flowering-2.pdf
- 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, 2575–2586. doi: 10.1007/s00122-013-2156-z
- Grenier, C., Cao, T. V., Ospina, Y., Quintero, C., Châtel, M. H., Tohme, J., et al. (2015). Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. *PLoS ONE* 10:136594. doi: 10.1371/journal.pone.0136594
- Grivet, L., and Arruda, P. (2002). Sugarcane genomics: Depicting the complex genome of an important tropical crop. *Curr. Opin. Plant Biol.* 5, 122–127. doi: 10.1016/S1369-5266(02)00234-0
- Guilly, S., Dumont, T., Thong-Chane, A., Barau, L., and Hoarau, J. Y. (2017). Analysis of multi-environment trials (MET) in the sugarcane breeding program of Réunion Island. *Euphytica* 213, 1–20. doi: 10.1007/s10681-017-1994-1
- Guo, Z., Tucker, D. M., Basten, C. J., Gandhi, H., Ersoz, E., Guo, B., et al. (2014). The impact of population structure on genomic prediction in stratified populations. *Theor. Appl. Genet.* 127, 749–762. doi: 10.1007/s00122-013-2255-x
- Hao, C. Y., Zhang, X. Y., Wang, L. F., Dong, Y. S., Shang, X. W., and Jia, J. Z. (2006). Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. *Mol. Breed.* 17, 69–77. doi: 10.1007/s11032-005-2453-6
- Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K., and Goddard, M. E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet. Select. Evolut.* 41, 1–9. doi: 10.1186/1297-9686-41-1
- 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, 1455–1462. doi: 10.1007/s00122-021-03782-6
- Heller-Uszynska, K., Uszynski, G., Huttner, E., Evers, M., Carlig, J., Caig, V., et al. (2011). Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Mol. Breed.* 28, 37–55. doi: 10.1007/s11032-010-9460-y
- Heslot, N., Akdemir, D., Sorrells, M. E., and Jannink, J. L. (2014). Integrating environmental covariates and crop modeling into the genomic selection

- framework to predict genotype by environment interactions. *Theor. Appl. Genet.* 127, 463–480. doi: 10.1007/s00122-013-2231-5
- Heslot, N., Jannink, J. L., and Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Sci.* 55, 1–12. doi: 10.2135/cropsci2014.03.0249
- Hoang, N. V., Furtado, A., Botha, F. C., Simmons, B. A., and Henry, R. J. (2015). Potential for genetic improvement of sugarcane as a source of biomass for biofuels. *Front. Bioeng. Biotechnol.* 3:182. doi: 10.3389/fbioe.2015.00182
- Hoarau, J. Y., Grivet, L., Offmann, B., Raboin, L. M., Diorflar, J. P., Payet, J., et al. (2002). Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTLs for yield components. *Theor. Appl. Genet.* 105, 1027–1037. doi: 10.1007/s00122-002-1047-5
- Hogarth, D. M. (1971). Quantitative inheritance studies in sugar-cane. I. Estimation of variance components. *Aust. J. Agric. Res.* 22, 93–102. doi: 10.1071/AR9710093
- Hogarth, D. M., Wu, K. K., and Heinz, D. J. (1981). Estimating genetic variance in sugarcane using a factorial cross design. *Crop Sci.* 21:21. doi: 10.2135/cropsci1981.0011183X002100010006x
- Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* 10, 156–161. doi: 10.1016/j.pbi.2007.01.003
- Isidro, J., Jannink, J. L., Akdemir, D., Poland, J., Heslot, N., and Sorrells, M. E. (2015). Training set optimization under population structure in genomic selection. *Theor. Appl. Genet.* 128, 145–158. doi: 10.1007/s00122-014-2418-4
- Jackson, P., and McRae, T. A. (2001). Selection of sugarcane clones in small plots: effects of plot size and selection criteria. *Crop Sci.* 41, 315–322. doi: 10.2135/cropsci2001.412315x
- Jackson, P. A. (2005). Breeding for improved sugar content in sugarcane. *Field Crops Res.* 92, 277–290. doi: 10.1016/j.fcr.2005.01.024
- Jarquín, D., de Leon, N., Romay, C., Bohn, M., Buckler, E. S., Ciampitti, I., et al. (2021). Utility of climatic information via combining ability models to improve genomic prediction for yield within the genomes to fields maize project. *Front. Genet.* 11:e592769. doi: 10.3389/fgene.2020.592769
- Jarquín, D., Howard, R., Crossa, J., Beyene, Y., Gowda, M., Martini, J. W. R., et al. (2020). Genomic prediction enhanced sparse testing for multi-environment trials. *G3: Genes Genomes Genet.* 10, 2725–2739. doi: 10.1534/g3.120.401349
- Jeong, N., Kim, K. S., Jeong, S., Kim, J. Y., Park, S. K., Lee, J. S., et al. (2019). Korean soybean core collection: Genotypic and phenotypic diversity population structure and genome-wide association study. *PLoS ONE* 14:224074. doi: 10.1371/journal.pone.0224074
- Jiang, Y., and Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics* 201, 759–768. doi: 10.1534/genetics.115.177907
- Junior, C. A. D. K., Manechini, J. R. V., Corrêa, R. X., Pinto, A. C. R., da Costa, J. B., Favero, T. M., et al. (2020). Genetic Structure analysis in sugarcane (*Saccharum* spp.) using target region amplification polymorphism (trap) markers based on sugar- and lignin-related genes and potential application in core collection development. *Sugar Tech.* 22, 641–654. doi: 10.1007/s12355-019-00791-0
- Kadam, D. C., Potts, S. M., Bohn, M. O., Lipka, A. E., and Lorenz, A. J. (2016). Genomic prediction of single crosses in the early stages of a maize hybrid breeding pipeline. *G3 Genes Genomes Genet.* 6, 3443–3453. doi: 10.1534/g3.116.031286
- Kehel, Z., Sanchez-Garcia, M., El Baouchi, A., Aberkane, H., Tsvilikas, A., Charles, C., et al. (2020). Predictive characterization for seed morphometric traits for gene bank accessions using genomic selection. *Front. Ecol. Evolut.* 8, 1–11. doi: 10.3389/fevo.2020.00032
- Krichen, L., Audergon, J. M., and Trifi-Farah, N. (2012). Relative efficiency of morphological characters and molecular markers in the establishment of an apricot core collection. *Hereditas* 149, 163–172. doi: 10.1111/j.1601-5223.2012.02245.x
- Kristensen, P. S., Jensen, J., Andersen, J. R., Guzmán, C., Orabi, J., and Jahoor, A. (2019). Genomic prediction and genome-wide association studies of flour yield and alveograph quality traits using advanced winter wheat breeding material. *Genes* 10, 1–19. doi: 10.3390/genes10090669
- Kumar, A., Kumar, S., Singh, K. B. M., Prasad, M., and Thakur, J. K. (2020a). Designing a mini-core collection effectively representing 3004 diverse rice accessions. *Plant Commun.* 1:100049. doi: 10.1016/j.xplc.2020.100049
- Kumar, R., Mohanraj, K., Anna Durai, A., and Premachandran, M. N. (2012). Pedigree analysis of sugarcane preant breeding pool used in evolving “Co” varieties in India. *Ind J. Genet. Plant Breed.* 72, 61–71. Available online at: <http://www.isgpb.org/article/pedigree-analysis-of-sugarcane-parental-breeding-pool-used-in-evolving-co-varieties-in-india>
- Kumar, S., Hilario, E., Deng, C. H., and Molloy, C. (2020b). Turbocharging introgression breeding of perennial fruit crops: a case study on apple. *Horticulture Res.* 7:270. doi: 10.1038/s41438-020-0270-z
- Kumar, S., Kirk, C., Deng, C. H., Shirliff, A., Wiedow, C., Qin, M., et al. (2019). Marker-trait associations and genomic predictions of interspecific pear (*Pyrus*) fruit characteristics. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-45618-w
- Lado, B., Battenfield, S., Guzmán, C., Quincke, M., Singh, R. P., Dreisigacker, S., et al. (2017). Strategies for selecting crosses using genomic prediction in two wheat breeding programs. *Plant Genome* 10:128. doi: 10.3835/plantgenome2016.12.0128
- Lekshmi, M., Pazhany, A. S., Sobhakumari, V. P., and Premachandran, M. N. (2017). Nuclear and cytoplasmic contributions from *Erianthus arundinaceus* (Retz.) Jeswiet in a sugarcane hybrid clone confirmed through genomic in situ hybridization and cytoplasmic DNA polymorphism. *Genet. Resour. Crop Evol.* 64, 1553–1560. doi: 10.1007/s10722-016-0453-5
- Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* 30, 2843–2851. doi: 10.1093/bioinformatics/btu356
- Li, X., Singh, J., Qin, M., Li, S., Zhang, X., Zhang, M., et al. (2019). Development of an integrated 200K SNP genotyping array and application for genetic mapping, genome assembly improvement and genome wide association studies in pear (*Pyrus*). *Plant Biotechnol. J.* 17, 1582–1594. doi: 10.1111/pbi.13085
- Lingle, S. E., Johnson, R. M., Tew, T. L., and Viator, R. P. (2010). Changes in juice quality and sugarcane yield with recurrent selection for sucrose. *Field Crops Res.* 118, 152–157. doi: 10.1016/j.fcr.2010.05.002
- Liu, W., Shahid, M. Q., Bai, L., Lu, Z., Chen, Y., Jiang, L., et al. (2015). Evaluation of genetic diversity and development of a core collection of wild rice (*Oryza rufipogon* Griff.) populations in China. *PLoS ONE* 10:e145990. doi: 10.1371/journal.pone.0145990
- Liu, Y., Hu, G., Zhang, A., Loladze, A., Hu, Y., Wang, H., et al. (2021). Genome-wide association study and genomic prediction of *Fusarium* ear rot resistance in tropical maize germplasm. *Crop J.* 9, 325–341. doi: 10.1016/j.cj.2020.08.008
- Lopez-Cruz, M., Crossa, J., Bonnett, D., Dreisigacker, S., Poland, J., Jannink, J. L., et al. (2015). Increased prediction accuracy in wheat breeding trials using a marker \times environment interaction genomic selection model. *G3 Genes Genomes Genet.* 5, 569–582. doi: 10.1534/g3.114.016097
- Lorenz, A., and Nice, L. (2017). “Chapter 2 training population design and resource allocation for genomic selection in plant breeding,” in *Genomic Selection for Crop Improvement*, eds R. K. Varshney, M. Roorkiwal, and M. E. Sorrells (Cham: Springer International Publishing AG 2017), 7–22. doi: 10.1007/978-3-319-63170-7_2
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., et al. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics* 9:3215. doi: 10.1371/journal.pgen.1003215
- Lu, Y., Yan, J., Guimarães, C. T., Taba, S., Hao, Z., Gao, S., et al. (2009). Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theor. Appl. Genet.* 120, 93–115. doi: 10.1007/s00122-009-1162-7
- Maenhout, S., De Baets, B., and Haesaert, G. (2010). Graph-based data selection for the construction of genomic prediction models. *Genetics* 185, 1463–1475. doi: 10.1534/genetics.110.116426
- Malosetti, M., Bustos-Korts, D., Boer, M. P., and Van Eeuwijk, F. A. (2016). Predicting responses in multiple environments: Issues in relation to genotype \times Environment interactions. *Crop Sci.* 56, 2210–2222. doi: 10.2135/cropsci2015.05.0311
- Malosetti, M., Voltas, J., Romagosa, I., Ullrich, S. E., and Eeuwijk, F. A. V. (2004). Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137, 139–145. doi: 10.1023/B:EUPH.0000040511.46388.ef
- Manimekalai, R., Suresh, G., Govinda Kurup, H., Athiappan, S., and Kandalam, M. (2020). Role of NGS and SNP genotyping methods in sugarcane improvement programs. *Crit. Rev. Biotechnol.* 40, 865–880. doi: 10.1080/07388551.2020.1765730

- Matsuoka, S., Kennedy, A. J., Santos, E. G. D., dos, Tomazela, A. L., and Rubio, L. C. S. (2014). Energy cane: its concept, development, characteristics, and prospects. *Adv. Botany* 2014, 1–13. doi: 10.1155/2014/597275
- Mayer, M., Höllker, A. C., González-Segovia, E., Bauer, E., Presterl, T., Ouzunova, M., et al. (2020). Discovery of beneficial haplotypes for complex traits in maize landraces. *Nat. Commun.* 11, 1–10. doi: 10.1038/s41467-020-18683-3
- Mendes de Paula, T. O., Brasileiro, B. P., Cursi, D. E., de Freitas, E. G., dos Santos, J. M., de Resende, M. D. V., et al. (2020). Establishment of gene pools for systematic heterosis exploitation in sugarcane breeding. *Agron. J.* 112, 3847–3858. doi: 10.1002/agj2.20299
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819
- Milligan, S. B., Martin, F. A., Bischoff, K. P., Quereadeaux, J. P., Dufrene, E. O., Quebedeaux, K. L., et al. (1989). Registration of ‘LCP 85-384’ sugarcane. *Crop Sci.* 34, 819–820. doi: 10.2135/cropsci1994.0011183X003400030042x
- Ming, R., Wang, Y. W., Draye, X., Moore, P. H., Irvine, J. E., and Paterson, A. H. (2002). Molecular dissection of complex traits in autopolyploids: Mapping QTLs affecting sugar yield and related traits in sugarcane. *Theor. Appl. Genet.* 105, 332–345. doi: 10.1007/s00122-001-0861-5
- Mohanraj, K., and Nair, N. V. (2014). Biomass potential of novel interspecific hybrids involving improved clones of *Saccharum*. *Ind. Crops Prod.* 53, 128–132. doi: 10.1016/j.indcrop.2013.12.004
- Momen, M., Mehrgardi, A. A., Sheikhi, A., Kranis, A., Tusell, L., Morota, G., et al. (2018). Predictive ability of genome-assisted statistical models under various forms of gene action. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-30089-2
- Momen, M., and Morota, G. (2018). Quantifying genomic connectedness and prediction accuracy from additive and non-additive gene actions 06 Biological Sciences 0604 Genetics. *Genet. Select. Evolut.* 50, 1–10. doi: 10.1186/s12711-018-0415-9
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Pérez-Hernández, O., Eskridge, K. M., et al. (2016). A genomic bayesian multi-trait and multi-environment model. *G3 Genes Genomes Genet.* 6, 2725–2774. doi: 10.1534/g3.116.032359
- Monteverde, E., Gutierrez, L., Blanco, P., Pérez de Vida, F., Rosas, J. E., Bonnecarrère, V., et al. (2019). Integrating molecular markers and environmental covariates to interpret genotype by environment interaction in rice (*Oryza sativa* L.) grown in subtropical areas. *G3 Genes Genomes Genet.* 9, 1519–1531. doi: 10.1534/g3.119.400064
- Muleta, K. T., Pressoir, G., and Morris, G. P. (2019). Optimizing genomic selection for a sorghum breeding program in Haiti: A simulation study. *G3 Genes Genomes Genet.* 9, 391–401. doi: 10.1534/g3.118.200932
- Müller, D., Schopp, P., and Melchinger, A. E. (2017). Persistency of prediction accuracy and genetic gain in synthetic populations under recurrent genomic selection. *G3 Genes Genomes Genet.* 7, 801–811. doi: 10.1534/g3.116.036582
- Nair, N. V. (2011). Sugarcane varietal development programmes in india: an overview. *Sugar Tech* 13, 275–280. doi: 10.1007/s12355-011-0099-8
- Nair, N. V., Mohanraj, K., Sunadaravelandian, K., Suganya, A., Selvi, A., and Appunu, C. (2017). Characterization of an intergeneric hybrid of *Erianthus procerus* × *Saccharum officinarum* and its backcross progenies. *Euphytica* 213:267. doi: 10.1007/s10681-017-2053-7
- Nayak, S. N., Song, J., Villa, A., Pathak, B., Ayala-Silva, T., Yang, X., et al. (2014). Promoting utilization of *Saccharum* spp. genetic resources through genetic diversity analysis and core collection construction. *PLoS ONE* 9:e110856. doi: 10.1371/journal.pone.0110856
- Nybom, H., and Låcis, G. (2021). Recent large-scale genotyping and phenotyping of plant genetic resources of vegetatively propagated crops. *Plants* 10, 1–30. doi: 10.3390/plants10020415
- Nyouma, A., Bell, J. M., Jacob, F., Riou, V., Manez, A., Pomiès, V., et al. (2020). Genomic predictions improve clonal selection in oil palm (*Elaeis guineensis* Jacq.) hybrids. *Plant Sci.* 299:110547. doi: 10.1016/j.plantsci.2020.110547
- Oakey, H., Cullis, B., Thompson, R., Comadran, J., Halpin, C., and Waugh, R. (2016). Genomic selection in multi-environment crop trials. *G3 Genes Genomes Genet.* 6, 1313–1326. doi: 10.1534/g3.116.027524
- O’Connor, K. M., Hayes, B. J., Hardner, C. M., Alam, M., Henry, R. J., and Topp, B. L. (2021). Genomic selection and genetic gain for nut yield in an Australian macadamia breeding population. *BMC Genomics* 22, 1–12. doi: 10.1186/s12864-021-07694-z
- Odilbekov, F., Armoniené, R., Koc, A., Svensson, J., and Chawade, A. (2019). GWAS-assisted genomic prediction to predict resistance to septoria tritici blotch in nordic winter wheat at seedling stage. *Front. Genet.* 10:1224. doi: 10.3389/fgene.2019.01224
- Ozimati, A., Kawuki, R., Esuma, W., Kayondo, S. I., Pariyo, A., Wolfe, M., et al. (2019). Genetic variation and trait correlations in an East African Cassava breeding population for genomic selection. *Crop Sci.* 59, 460–473. doi: 10.2135/cropsci2018.01.0060
- Pandey, M. K., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., et al. (2017). Development and evaluation of a high density genotyping ‘‘Axiom-Arachis’’ array with 58 K SNPs for Accelerating genetics and breeding in Groundnut. *Sci. Rep.* 7, 1–10. doi: 10.1038/srep40577
- Pandey, M. K., Chaudhari, S., Jarquin, D., Janila, P., Crossa, J., Patil, S. C., et al. (2020). Genome-based trait prediction in multi-environment breeding trials in groundnut. *Theor. Appl. Genet.* 133, 3101–3117. doi: 10.1007/s00122-020-03658-1
- Park, S., Jackson, P., and Berding, N. (2007). Conventional breeding practices within the Australian sugarcane breeding program. *Proc. Aust. Soc. Sugar Cane Technol.* 29, 113–121. Available online at: <https://publications.csiro.au/rpr/pub?list=BRO&pid=procite:d43d4ddd-b273-4ddd-a5e5-8ef1d6ea7267>
- Pascual, L., Fernández, M., Aparicio, N., López-Fernández, M., Fité, R., Giraldo, P., et al. (2020). Development of a multipurpose core collection of bread wheat based on high-throughput genotyping data. *Agronomy* 10, 1–16. doi: 10.3390/agronomy10040534
- Peixoto, L. S., Nunes, J. A. R., and Furtado, D. F. (2016). Factor analysis applied to the G+GE matrix via REML/BLUP for multi-environment data. *Crop Breed. Appl. Biotechnol.* 16, 1–6. doi: 10.1590/1984-70332016v16n1a1
- Pereira, A. R. (1983). Climatic conditioning of flowering induction in sugarcane. *Agric. Meteorol.* 29, 103–110. doi: 10.1016/0002-1571(83)90043-2
- Piperidis, N., Piperidis, G., and Hont, A. D. (2021). Genomic organization of sugarcane cultivars revealed by chromosome-specific oligonucleotide probes. *Proc. Austr. Soc. Sugar Cane Technol.* 42, 573–584. Available online at: https://elibrary.sugarresearch.com.au/bitstream/handle/11079/18205/Genomic%20organization%20of%20sugarcane%20cultivars%20revealed%20by%20chromosome_specific%20oligonucleotide%20probes.pdf?sequence=1&isAllowed=y
- Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:32253. doi: 10.1371/journal.pone.0032253
- Ram, B., and Hemaprabha, G. (2020). The sugarcane variety Co 0238-A reward to farmers and elixir to India’s sugar sector. *Curr. Sci.* 118, 1643–1646. Available online at: <https://www.currentscience.ac.in/>
- Ramburan, S. (2014). A multivariate illustration and interpretation of non-repeatable genotype × environment interactions in sugarcane. *Field Crops Res.* 157, 57–64. doi: 10.1016/j.fcr.2013.12.009
- Ramdoyal, K., and Badaloo, G. H. (2002). ‘‘Prebreeding in sugarcane with an emphasis on the programme of the mauritius sugar industry research institute,’’ in *Managing Plant Genetic Diversity*, eds J. M. M. Engels, V. Ramanatha Rao, A. H. D. Brown, and M. T. Jackson (New York, NY: CABI Publishing), 307–322.
- Ravinder, K., Appunu, C., Anna Durai, A., Premachandran, M. N., Raffee Viola, V., Bakshi, Ram, et al. (2015). Genetic confirmation and field performance comparison for yield and quality among advanced generations of erianthus arundinaceus, *E. bengalense* and *saccharum spontaneum* cyto-nuclear genome introgressed sugarcane intergeneric hybrids. *Sugar Tech* 17, 379–385. doi: 10.1007/s12355-014-0333-2
- Reffay, N., Jackson, P. A., Aitken, K. S., Hoarau, J. Y., D’Hont, A., Besse, P., et al. (2005). Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. *Mol. Breed.* 15, 367–381. doi: 10.1007/s11032-004-7981-y
- Resende, R. T., Resende, M. D. V., Silva, F. F., Azevedo, C. F., Takahashi, E. K., Silva-Junior, O. B., et al. (2017). Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. *Heredity* 119, 245–255. doi: 10.1038/hdy.2017.37
- Rincent, R., Charcosset, A., and Moreau, L. (2017). Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. *Theor. Appl. Genet.* 130, 2231–2247. doi: 10.1007/s00122-017-2956-7

- Rincen, R., Laloë, D., Nicolas, S., Altmann, T., Brunel, D., Revilla, P., et al. (2012). Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: Comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192, 715–728. doi: 10.1534/genetics.112.141473
- Roach, B. T. (1989). “Origin and improvement of the genetic base of sugarcane,” in *Proceedings of the Australian Society of Sugar Cane Technologists- Annual Conference* (Tweed Heads, NSW), 34–47.
- Roorikwal, M., Jarquin, D., Singh, M. K., Gaur, P. M., Bharadwaj, C., Rathore, A., et al. (2018). Genomic-enabled prediction models using multi-environment trials to estimate the effect of genotype × environment interaction on prediction accuracy in chickpea. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-30027-2
- Saatchi, M., McClure, M. C., McKay, S. D., Rolf, M. M., Kim, J., Decker, J. E., et al. (2011). Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. *Genet. Select. Evolut.* 43, 1–16. doi: 10.1186/1297-9686-43-40
- Sallam, A. H., Endelman, J. B., Jannink, J. -L., and Smith, K. P. (2015). Assessing genomic selection prediction accuracy in a dynamic Barley breeding population. *Plant Genome* 8:20. doi: 10.3835/plantgenome2014.05.0020
- Saxena, R. K., Patel, K., Sameer Kumar, C. V., Tyagi, K., Saxena, K. B., and Varshney, R. K. (2018). Molecular mapping and inheritance of restoration of fertility (Rf) in A4 hybrid system in pigeonpea (*Cajanus cajan* (L.) Millsp.). *Theor. Appl. Genet.* 131, 1605–1614. doi: 10.1007/s00122-018-3101-y
- Schmitz, A., Kennedy, P. L., and Zhang, F. (2020). Sugarcane and sugar yields in Louisiana (1911–2018): Varietal development and mechanization. *Crop Sci.* 60, 1303–1312. doi: 10.1002/csc2.20045
- Schulz-Streeck, T., Ogutu, J. O., Gordillo, A., Karaman, Z., Knaak, C., and Piepho, H. P. (2013). Genomic selection allowing for marker-by-environment interaction. *Plant Breed.* 132, 532–538. doi: 10.1111/pbr.12105
- Servin, B., Martin, O. C., Mézard, M., and Hospital, F. (2004). Toward a theory of marker-assisted gene pyramiding. *Genetics* 168, 513–523. doi: 10.1534/genetics.103.023358
- Shadmehr, A., Ramshini, H., Zeinalabedini, M., Almani, M. P., Ghaffari, M. R., Darbandi, A. I., et al. (2017). Phenotypic variability assessment of sugarcane germplasm (*Saccharum officinarum* L.) and extraction of an applied mini-core collection. *Agriculture* 7, 1–15. doi: 10.3390/agriculture7070055
- Shanmugavadivu, R., and Rao, P. N. G. (2009). A comparison of flowering behaviour of sugarcane clones in two different locations. *Sugar Tech* 11, 401–404. doi: 10.1007/s12355-009-0070-0
- Shanthi, R. M., Bhagyalakshmi, K., Alarmelu, S., Hemaprabha, G., and Nagarajan, R. (2008). Relative performance of the sugarcane families in early selection stages. *Ind. J. Genet. Plant Breed.* 10, 114–118. doi: 10.1007/s12355-008-0019-8
- Sharma, S., Upadhyaya, H. D., Varshney, R. K., and Gowda, C. L. L. (2013). Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. *Front. Plant Sci.* 4:309. doi: 10.3389/fpls.2013.00309
- Singh, S., Vikram, P., Sehgal, D., Burguenio, J., Sharma, A., Singh, S. K., et al. (2018). Harnessing genetic potential of wheat germplasm banks through impact-oriented-prebreeding for future food and nutritional security. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-30667-4
- Song, J., Yang, X., Resende, M. F. R., Neves, L. G., Todd, J., Zhang, J., et al. (2016). Natural allelic variations in highly polyploidy *Saccharum* complex. *Front. Plant Sci.* 7:804. doi: 10.3389/fpls.2016.00804
- Sreenivasan, T., and Ahloowalia, B. (1987). “Cytogenetics,” in *Sugarcane Improvement Through Breeding*, ed D. Heinz (Amsterdam: Elsevier) 211–253. doi: 10.1016/B978-0-444-42769-4.50010-2
- Stringer, J. K., Cox, M. C., Atkin, F. C., Wei, X., and Hogarth, D. M. (2010). Family selection improves the efficiency and effectiveness of selecting original seedlings and parents. *Proc. Int. Soc. Sugar Cane Technol* 27, 1–8. Available online at: <https://issct.org/wp-content/uploads/proceedings/2010/2010%20Papers.pdf>
- Tai, P. Y. P., and Miller, J. D. (2001). A core collection for *Saccharum spontaneum* L. from the World collection of sugarcane. *Crop Sci.* 41, 879–885. doi: 10.2135/cropsci2001.413879x
- Tan, B., Grattapaglia, D., Martins, G. S., Ferreira, K. Z., Sundberg, B., and Ingvarsson, P. K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two *Eucalyptus* species and their F1 hybrids. *BMC Plant Biol.* 17, 1–15. doi: 10.1186/s12870-017-1059-6
- Tena Gashaw, E., Mekbib, F., and Ayana, A. (2018). Sugarcane landraces of Ethiopia: germplasm collection and analysis of regional diversity and distribution. *Adv. Agricult.* 2018, 1–18. doi: 10.1155/2018/7920724
- Thorwarth, P., Yousef, E. A. A., and Schmid, K. J. (2018). Genomic prediction and association mapping of curd-related traits in gene bank accessions of cauliflower. *G3 Genes Genomes Genet.* 8, 707–718. doi: 10.1534/g3.117.300199
- Todd, J., Sandhu, H., Hale, A., Glaz, B., and Wang, J. (2017). Phenotypic evaluation of a diversity panel selected from the world collection of sugarcane (*Saccharum* spp) and related grasses. *Maydica* 62:6525. doi: 10.1590/1678-4499.2016525
- Tsai, H. Y., Janss, L. L., Andersen, J. R., Orabi, J., Jensen, J. D., Jahoor, A., et al. (2020). Genomic prediction and GWAS of yield, quality and disease-related traits in spring barley and winter wheat. *Sci. Rep.* 10, 1–15. doi: 10.1038/s41598-020-63862-3
- Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., et al. (2014). A powerful tool for genome analysis in maize: Development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* 15, 1–15. doi: 10.1186/1471-2164-15-823
- Vargas, M., Crossa, J., Sayre, K., Reynolds, M., Ramirez, M. E., and Talbot, M. (1998). Interpreting genotype × environment interaction in wheat by partial least squares regression. *Crop Sci.* 38, 679–689. doi: 10.2135/cropsci1998.001183X003800030010x
- Varona, L., Legarra, A., Toro, M. A., and Vitezica, Z. G. (2018). Non-additive effects in genomic selection. *Front. Genet.* 9:78. doi: 10.3389/fgenet.2018.00078
- Varshney, R. K., Singh, V. K., Kumar, A., Powell, W., and Sorrells, M. E. (2018). Can genomics deliver climate-change ready crops? *Curr. Opin. Plant Biol.* 45, 205–211. doi: 10.1016/j.pbi.2018.03.007
- 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, 1493–1511. doi: 10.1007/s00122-021-03785-3
- Waldron, C., Glasziou, K. T., and Bull, T. A. (1967). The physiology of sugar-cane. IX Factor affecting the photosynthesis and sugar storage. *Austr. J. Biol. Sci.* 20, 1043–1052. doi: 10.1071/BI9671043
- Wang, J., Zhou, Z., Zhang, Z., Li, H., Liu, D., Zhang, Q., et al. (2018a). Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. *Heredity* 121, 648–662. doi: 10.1038/s41437-018-0075-0
- Wang, L. P., Jackson, P. A., Lu, X., Fan, Y. H., Foreman, J. W., Chen, X. K., et al. (2008). Evaluation of sugarcane × *Saccharum spontaneum* progeny for biomass composition and yield components. *Crop Sci.* 48, 951–961. doi: 10.2135/cropsci2007.10.0555
- Wang, X., Xu, Y., Hu, Z., and Xu, C. (2018b). Genomic selection methods for crop improvement: Current status and prospects. *Crop J.* 6, 330–340. doi: 10.1016/j.cj.2018.03.001
- Wang, X., Zhang, Z., Xu, Y., Li, P., Zhang, X., and Xu, C. (2020). Using genomic data to improve the estimation of general combining ability based on sparse partial diallel cross designs in maize. *Crop J.* 8, 819–829. doi: 10.1016/j.cj.2020.04.012
- Wei, X., and Jackson, P. (2017). Addressing slow rates of long-term genetic gain in sugarcane. *Int. Sugar J.* 119, 1923–1930. Available online at: <https://internationalsugarjournal.com/paper/addressing-slow-rates-of-long-term-genetic-gain-in-sugarcane/>
- Wei, X., Jackson, P. A., Hermann, S., Kilian, A., Heller-Uszynska, K., and Deomano, E. (2010). Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker-trait associations in sugarcane. *Genome* 53, 973–981. doi: 10.1139/G10-050
- Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J. L., Sorrells, M. E., et al. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3 Genes Genomes Genet.* 2, 1427–1436. doi: 10.1534/g3.112.003699
- Wolfe, M. D., Del Carpio, D. P., Alabi, O., Ezenwaka, L. C., Ikeogu, U. N., Kayondo, I. S., et al. (2017). Prospects for genomic selection in cassava breeding. *Plant Genome* 10:15. doi: 10.3835/plantgenome2017.03.0015
- Wong, C. K., and Bernardo, R. (2008). Genomewide selection in oil palm: Increasing selection gain per unit time and cost with small populations. *Theor. Appl. Genet.* 116, 815–824. doi: 10.1007/s00122-008-0715-5

- Yabe, S., Hara, T., Ueno, M., Enoki, H., Kimura, T., Nishimura, S., et al. (2018). Potential of genomic selection in mass selection breeding of an allogamous crop: An empirical study to increase yield of common buckwheat. *Front. Plant Sci.* 9:276. doi: 10.3389/fpls.2018.00276
- Yadav, S., Jackson, P., Wei, X., Ross, E. M., Aitken, K., Deomano, E., et al. (2020). Accelerating genetic gain in sugarcane breeding using genomic selection. *Agronomy* 10, 1–21. doi: 10.3390/agronomy10040585
- 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, 2235–2252. doi: 10.1007/s00122-021-03822-1
- Yan, W. (2015). Mega-environment analysis and test location evaluation based on unbalanced multiyear data. *Crop Sci.* 55, 113–122. doi: 10.2135/cropsci2014.03.0203
- Yan, W. (2021). Estimation of the optimal number of replicates in crop variety trials. *Front. Plant Sci.* 11:590762. doi: 10.3389/fpls.2020.590762
- Yan, W., Frégeau-Reid, J., Martin, R., Pageau, D., and Mitchell-Fetch, J. (2015). How many test locations and replications are needed in crop variety trials for a target region? *Euphytica* 202, 361–372. doi: 10.1007/s10681-014-1253-7
- Yang, X., Song, J., You, Q., Paudel, D. R., Zhang, J., and Wang, J. (2017). Mining sequence variations in representative polyploid sugarcane germplasm accessions. *BMC Genomics* 18, 1–16. doi: 10.1186/s12864-017-3980-3
- You, Q., Yang, X., Peng, Z., Xu, L., and Wang, J. (2018). Development and applications of a high throughput genotyping tool for polyploid crops: Single nucleotide polymorphism (SNP) array. *Front. Plant Sci.* 9:104. doi: 10.3389/fpls.2018.00104
- Yu, H., Xie, W., Li, J., Zhou, F., and Zhang, Q. (2014). A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J.* 12, 28–37. doi: 10.1111/pbi.12113
- Yu, X., Li, X., Guo, T., Zhu, C., Wu, Y., Mitchell, S. E., et al. (2016). Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nat. Plants* 2, 1–7. doi: 10.1038/nplants.2016.150
- Zaharieva, M., Santoni, S., and David, J. (2001). Use of RFLP markers to study genetic diversity and to build a core-collection of the wild wheat relative *Ae. geniculata* Roth (= *Ae. ovata* L.). *Genet. Select. Evolut.* 33:884. doi: 10.1186/BF03500884
- Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., et al. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Front. Plant Sci.* 8:1916. doi: 10.3389/fpls.2017.01.916
- Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D. J., and Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS ONE* 5:12648. doi: 10.1371/journal.pone.0012648
- Zhao, X., Li, B., Zhang, K., Hu, K., Yi, B., Wen, J., et al. (2016). Breeding signature of combining ability improvement revealed by a genomic variation map from recurrent selection population in *Brassica napus*. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep29553
- Zhou, M. (2020). General and specific combining ability effects for sugarcane yield among South African breeding populations. *South Afr. J. Plant Soil* 37, 300–307. doi: 10.1080/02571862.2020.1749318

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.

Copyright © 2021 Mahadevaiah, Appunu, Aitken, Suresha, Vignesh, Mahadeva Swamy, Valarmathi, Hemaprabha, Alagarasan and Ram. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.