Sugar Tech (Jan-Feb 2022) 24(1):320–340
https://doi.org/10.1007/s12355-021-01012-3

S.I. : HISTORY OF SUGARCANE BREEDING AND MOLECULAR GENETICS

Applications of Quantitative Genetics and Statistical Analyses in Sugarcane Breeding

Jean-Yves Hoarau[1,3,6] · Thomas Dumont[1] · Xianming Wei[2] · Philip Jackson[4] · Angélique D'Hont[5,6]

Received: 28 November 2020 / Accepted: 17 May 2021 / Published online: 26 July 2021

Abstract Sugarcane breeding programs aim to deliver new high-yielding varieties, resistant to diseases and pests, which contribute to profitability and sustainability of cane industries. These programs generally mobilize significant experimental, technological and human resources on long-term basis. Their efficiency in terms of genetic gains per unit of cost and time and their ability to release new varieties rely on the development of many breeding applications based on quantitative genetics theory and on statistical analyses of numerous experimental data from selection schemes including DNA marker data developed for some genomic breeding applications. New methodological approaches and new technologies that might better guide and support breeding research in cultivars development programs are continually sought. This paper presents an overview of the main applications developed in statistical methodology in support of the efficiency of sugarcane breeding programs. For each type of application, its conceptual and methodological framework is presented. Implementation issues are reviewed as well as the main scientific and practical achievements so far obtained.

Keywords Sugarcane breeding · Linear mixed model · Quantitative genetics · Genetic gain · Genomics applications

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12355-021-01012-3.

✉ Jean-Yves Hoarau
jean-yves.hoarau@cirad.fr

1 eRcane, 29, rue d'Emmerez de Charmoy, 97490 Saint-Clotilde, La Réunion, France

2 Sugar Research Australia, Mackay, QLD, Australia

3 UMR AGAP, CIRAD, 97490 Saint-Clotilde, La Réunion, France

4 CSIRO Agriculture, PO Aitkenval, Townsville, QLD 4814, Australia

5 UMR AGAP, CIRAD, 34398 Montpellier, France

6 AGAP, Univ. Montpellier, CIRAD, INRAE, L'Institut Agro, 34000 Montpellier, France

Introduction

Profitability and sustainability of sugarcane industries rely on several cornerstones including development of best agronomic practices and efficient breeding programs that regularly deliver improved cultivars. Efficient sugarcane breeding programs require availability of diverse germplasm and are facilitated by formulation of appropriate selection strategies and procedures, optimized resource allocation and supportive genomics approaches. Moreover, variety development programs constantly require statistical reasoning for designing breeding schemes as efficiently as possible and for properly interpreting collected data to make correct inferences and conclusions about the investigated question(s).

Statistical reasoning is a primary concern of sugarcane breeders and geneticists. The management of a breeding program is a daily massively decision-based exercise. Decisions are made regularly on planning and designing of multiple comparative experiments using biological material (germplasm, pathogen inoculum, etc.). Much statistical reasoning combines ideas about data and probability, which leads to making inferences and interpreting statistical results (Garfield 2002). Underlying this reasoning is a conceptual understanding of important ideas such as random sampling, normal distribution of quantitative traits and

hypothesis testing and a hands-on mastering of the principles of experimental design and statistical modeling for data analysis.

For instance, when a breeder aims to objectively compare several sugarcane varieties for cane yield or sugar content in a field trial, statistical reasoning is required: (1) to plan an efficient experimental design (grouping of homogeneous experimental units, random assignment of varieties to units within a group, independence between the replicated units of a single variety); (2) to control the uniformity of management applied to the trial and the protocols used to collect data to prevent any bias in final variety comparisons; and (3) to design a model of *analysis of variance* of the data affected by different sources of variation under the breeder control to detect the differences between variety means that are statistically greater than the estimated experimental error. Any glaring lack of statistical thinking in such comparative experiments can lead to unfounded decisions to promote some varieties and therefore reduce genetic progress, waste time and money.

Broadly speaking, the general organization of large sugarcane breeding programs follows some form of recurrent selection scheme (Falconer and Mackay 1996) as pictured in Fig. 1. Despite many differences in details between selection programs of different countries such as population sizes and duration (Milligan 1994; Cox et al. 2000; Scortecci et al. 2012; Zhou 2013; Dumont et al. 2019, 2021; Santchurn et al. 2021; Cursi et al. 2021), a sequence of selection trials typically follows the four typical successive stages of Fig. 1. This trial sequence is characterized by a gradually decreasing number of candidate lines tested in progressively more accurate trials (larger plot size more replicates and sites). This sequence starts with many seedlings (several tens of thousands or more) which are selected and then clonally propagated and selected in a scheme which normally spans at least ten years of experiments to end with a few superior varieties tested in pre-commercial trials. In each program, the technical choices (trading-off decisions on resource allocation and selection intensities across stages, trial designs and analyze models) and thinking mobilized in quantitative genetics are intended to provide the best genetic gain per unit of cost and time. Some breeding programs also benefit from creative genetics research using DNA markers of the genome which aim to develop molecular breeding approaches adapted to the complex polyploid genome of sugarcane (Hoarau et al. 2007).

In the management of breeding programs, many decisions have to be made (choice of the parents, candidates to be advanced into next selection stages, resource allocation efficiency, etc.). Very often, statistical analyses based on biometrics or quantitative genetics theory and using appropriate experimental data may provide genetic or statistical benchmarks to help support good decision making. The present review aims to present some useful applications of quantitative genetics and statistical analyses that can support decisions affecting the efficiency of sugarcane breeding programs. As a prelude, Table 1 provides a list of statistical and genetic terms that are widely used and essential to full comprehension of this review, based on our own experiences and on several references, all of which are cited here. These and some other basic concepts in statistical analysis of data in sugarcane breeding programs were presented in this Special Issue in the paper of Jackson (2021). Our review revises and advances these concepts further, aiming particularly to discuss some important applications in sugarcane breeding programs. It comprises four different sections: (1) The first one is an introductory

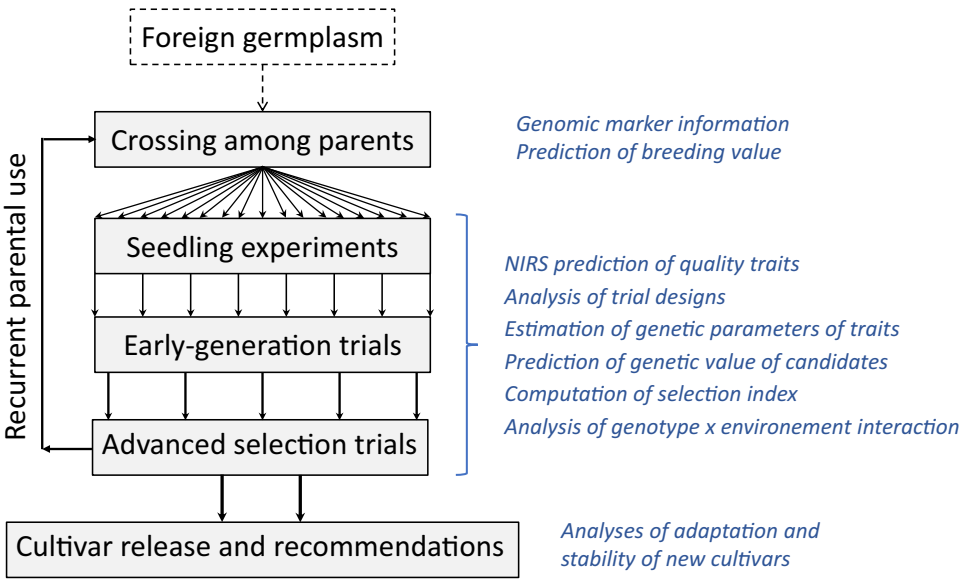

Fig. 1 Schematic flowchart of a typical sugarcane breeding program (left) and genetic and statistical analyses to support breeding and selection decisions (right) (adapted from Wei and Jackson 2016)

Table 1 Working definitions of statistical and genetic terms used throughout this review

Term	Definition
Genotype × environment (GE) interaction	Differential response for a trait between varieties when tested in multiple environments
Linear mixed model (LMM)	Statistical model explaining phenotypic data of an experiment as a linear combination of explanatory variables, some of which are fixed effects and other random effects
Variance–covariance (VCOV) matrix	A square matrix giving variances of a set of random variables on its diagonal and pairwise covariances off-diagonal
G matrix	VCOV matrix between all random effects of a LMM, except residual effects
R matrix	VCOV matrix between residual effects of a LMM
Best Linear Unbiased Prediction (BLUP)	An acronym for the statistical estimator of random effects in a LMM that has the lowest variance (*Best*) and zero bias (*Unbiased*)
BLUP Methodology	Methodology of prediction of random effects in a LMM with an internal algorithm (REML) providing BLUP values
Broad-sense heritability (h^2_{bs})	Percentage of the phenotypic variability of a trait attributable to genetic effects (as opposed to environmental effects), used to estimate accuracy of breeding experiments
Narrow-sense heritability (h^2_{ns})	Percentage of the phenotypic variability of a quantitative trait which is heritable, used to predict trait response to selection between generations
Selection index	An index used to rank varieties for their genetic merit (based on a linear combination of their genetic values for all traits of interest weighted by economic coefficients)
Quantitative trait allele (QTA)	Allele at a locus that is involved in the genetic architecture of a trait and exerts an (favorable or unfavorable) quantitative effect on the trait value
Genome-wide association studies (GWAS)	Search for statistical associations between a trait variation in a population of individuals and variations in their genome-wide DNA marker profiles to locate QTAs
Genomic selection (GS)	Methodologies of prediction of the genetic or breeding values of candidate genotypes for a trait in statistical models using their genome-wide DNA marker profiles

reminder of issues related to the management of breeding programs and the efficient analysis of breeding experiments based on mixed models; (2) the second one deals with statistical analysis of selection trials; (3) the third section is devoted to quantitative genetics approaches to optimize selection schemes; and (4) finally, the last one deals with analysis of DNA marker related to molecular breeding approaches.

Breeding Program Management and Data Modeling Approaches

Developing improved sugarcane varieties in a breeding program is a challenging task. It involves designing breeding strategies and making numerous decisions based on principles of quantitative genetics, statistical reasoning and data modeling approaches. Breeding strategies define a detailed plan of actions to operate a variety development program expected to provide genetic gains for target traits. Efficiency of breeding strategies can be assessed using the breeder's equation benchmark that predicts the rate of genetic gain (ΔG) expected from a selection pressure exerted on a population (Lush 1937). Understanding the principles of this equation is of fundamental importance to understanding how plant breeding works. The first paragraph of this section briefly presents this equation, its key ingredients, their meaning and how it can be used. Breeding decisions encompass all choices made at different stages of a program: choosing the parents to be crossed, the varieties to be advanced into the successive selection stages all the way to the release of new improved commercial varieties. Efficiency of breeding decisions depends on the parameters depicted in the breeder's equation. Appropriate design and statistical analysis of breeding field experiments can be used to estimate these parameters, and appropriate mathematical models can predict optimal configurations of breeding program design which maximize ΔG for any limited amount of resources and funding. Each breeding experiment, whatever its purpose, is characterized by specific features and frequent experimental constraints. These latter can be flexibly and efficiently handled with linear mixed models (LMM) approaches. LMMs constitute a powerful general framework of analysis of experimental data (Balzarini 2002; Piepho et al. 2008). It has the advantage of providing unbiased statistical inferences likely supporting objective conclusions in the breeding issue surveyed. The second paragraph of this section presents the concept of LMM and the benefits that can be

drawn from it when analyzing sugarcane breeding experiments.

The Breeder's Equation Benchmark

The equation of genetic gain (ΔG) known as the breeder's equation can be understood as the improvement in the mean genetic value of a quantitative trait of interest for a population over a defined time period, e.g., one breeding or selection cycle. This equation which expresses the predicted rate of genetic gain for a trait that can be achieved in a given breeding or selection cycle is as follows:

$$\Delta G = \frac{ih^2\sigma_P}{L} \quad (1)$$

where ΔG is the rate of genetic gain, i represents the selection intensity, h^2 represents a heritability estimate of the trait, σ_P^2 is the observed phenotypic variation and L is the interval in time units to complete the desired breeding cycle. If intergenerational gains (i.e., gain in progeny performance each breeding cycle from crossing selecting parents) are being considered, the h^2 parameter is the narrow-sense heritability of the trait (h_{ns}^2). The narrow-sense heritability represents the proportion of the phenotypic variance that is passed to progeny and σ_P^2 is the phenotypic variation among the genitors (being selected from the previous progeny population). If clonal selection is being considered, the h^2 parameter is the broad-sense heritability of the trait (h_{bs}^2), and this is the proportion of the phenotypic variance attributable to total genetic effects (as opposed to environmental effects) and σ_P^2 is the phenotypic variation among tested individuals. The quantitative parametric framework of the breeder's equation is a tool that can help identify bottlenecks that limit the rate of genetic gain in breeding programs and therefore also help identify steps to improve current breeding and selection methods to give faster gains. It can also help predict efficiencies of different plans of actions, for instance, strategies based on use of new technologies (Yadav et al. 2020).

Analysis of Breeding Field Trials with Linear Mixed Model (LMM)

Linear models conceived to analyze data from experiments in sugarcane breeding programs often require (in addition to the residual term), one or several terms representing *sources of variation* which can be considered to arise due to *random* effects. The definition of *random* and *fixed* effects in statistical analysis and some advantages in considering some important sources of variation (e.g., clones or families) in breeding programs as *random* effects were discussed by Jackson (2021) in this issue. Linear mixed models (LMM) can be conveniently used to analyze datasets containing both fixed-effect and random-effect terms.

One of the major advantages of LMM (unlike ANOVA composed of fixed-effect terms) is that they can very easily accommodate unbalanced experimental designs. In sugarcane breeding programs and research generally, datasets frequently have one or several unbalanced aspects, not only in a joint analysis of several experiments (e.g., having different replicate numbers or variety sets across trials), but also in single experiment, due to unbalanced numbers of replicates per variety (e.g., resulting from missing values or different numbers of plots of clones due to limited planting material of some clones) or unbalanced incomplete blocks frequently designed to deal with potential spatial heterogeneity in large trials.

A second major advantage of LMMs is these represent a flexible general framework to analyze datasets that may exhibit particular covariance structures arising in data due to joint variability (covariance) between some random-effect terms in data modeling (Henderson 1984). Many sugarcane breeding experiments have a range of important covariance structures, either arising (1) from relationships between experimental units of spatial nature (e.g., plots close to each other in field trials are often more similar to each other on average due to similar soil or other factors), or of a temporal nature (error effects across different years in the same plot are likely to be correlated with some extent), (2) from kinship relationships between some genetic entries (family or genotypes) or (3) because of a need of a joint analysis of several traits. As emphasized by Margarido et al. (2015), '*breeding programs typically leverage data collected for many traits, in multiple locations and along several years. Consequently, genetic and residual (co)variances are expected to be different across traits and environments, which in turn makes this type of data particularly suited for mixed model analysis*.'

Table 2 presents an overview of some emblematic genetic and breeding applications related either to breeding or to selection issues that are performed in support of sugarcane programs and which necessarily may benefit from the implementation of LMMs.

The general equation of a LMM usually expressed in matrix algebra notation is (Piepho et al. 2008):

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Zu} + \boldsymbol{e} \quad (2)$$

where $\boldsymbol{Y}_{n\times 1}$ is the response vector of n observations; $\boldsymbol{\beta}_{p\times 1}$ the vector of p fixed effects corresponding to one or several variable(s) chosen by the investigator to be fixed in nature; $\boldsymbol{u}_{q\times 1}$ the vector of q random effects of one or several random variable(s) and $\boldsymbol{e}_{n\times 1}$ the error terms, both being unobservable and unknown vectors of random values; and $\boldsymbol{X}_{n\times p}$ and $\boldsymbol{Z}_{n\times q}$ the incidence matrices. Assumptions regarding the structure of the $\boldsymbol{G}$ matrix of variance–

Table 2 Examples of applications in sugarcane breeding and genetics using linear mixed models (LMM)

	Objective	Needed information
Breeding applications	Estimation of trait heritability at narrow-sense	Pedigree among lines, phenotypic data, experiment designs
	Prediction of breeding value (BV) of parents	Pedigree among lines, phenotypic data, experiment designs
	Computation of selection index	BV estimates and partial regression coefficients
	Prospection of haplotype of agronomic interest through QTL[a] or GWAS[b]	Phenotypic data, experiment designs, marker data (genetic maps)
Selection applications	NIRS prediction of quality traits	Calibration data base, NIRS spectra data of lines to be predicted
	Analysis of trial	Phenotypic data (experiment designs)
	Spatial analysis of a trial	Phenotypic data, spatial position of individual plots (experiment designs)
	Prediction of genetic value (GV) of candidate lines	Phenotypic data, experiment designs
	Estimation of genetic parameters (variance components, broad-sense heritability, etc.)	Phenotypic data, experiment designs
	Analysis of genotype × environment interactions	Multi-environment trial database, trial designs
	Computation of selection index	GV estimates and partial regression coefficients
	Genomic selection (GS) and GWAS applications	Marker data, kinship information, experiment designs
	Analysis of performance and stability of varieties across environments	Multi-environment trial database, trial designs

[a]Quantitative trait loci

[b]Genome-wide association studies

covariance (VCOV) of random effects in $\boldsymbol{u}$ and of the $\boldsymbol{R}$ matrix of VCOV of random terms in $\boldsymbol{e}$ will define a particular mixed model (Balzarini 2002). The estimation of random and fixed effects, respectively, known as *best linear unbiased predictions* (BLUP) or *estimations* (BLUE) and that of the variance components are usually obtained with the *restricted maximum likelihood* (REML) algorithm (Henderson 1984). This algorithm solves the mixed-model equations of Eq. 2 by searching the best estimates of these BLUE and BLUP parameters and of the variance components using a recursive numerical method until reaching convergence. Some statistical software which fits LMMs via the REML method includes ASReml (Gilmour et al. 2009), SAS (Littell et al. 2006), R (R Development Core Team 2010) and GenStat (Payne et al. 2009). Because this estimation procedure is based not on any analytical computations but on recursive approximations, LMM can handle unbalanced or incomplete datasets.

BLUP concept and methodology of LMMs imply a fundamental shrinkage property inherent to the assessment of the effects of random variables. Figure 2 illustrates this shrinkage property in a concrete example (Online resource 1) of a mixed-model analysis of an unbalanced experiment of sugarcane varieties tested in a randomized complete trial, in which variety effect is considered as a random variable. The above-average individual means are shrunk downward toward the overall population mean, whereas below-average individual means are shrunk upward toward population mean. The degree of shrinkage of the variety effects toward the population mean depends on variance components and becomes more pronounced as error variance increases in magnitude relative to variance due to the random effects being predicted. The shrinkage adjustment (toward the population mean) is generally on a proportional basis, so that it is largest in absolute value for performances of genotypes which are either extremely high or extremely low. This property implies an expectation that the highest yielding (or lowest yielding) varieties in an experiment will perform less (more) well in a future trial. In this regard, BLUP methodology allows inferences for variety performance broader than the current experiment and therefore leads to more objective and cautious selection decision making. Generally speaking, selection approaches exploiting BLUP methodology are advisable, particularly in unbalanced dataset contexts. One of the desirable properties of BLUPs is that these values maximize the correlation of predicted genotypic values and true genotypic values and provide optimized predictions of variety performance (Piepho et al. 2008).

When implementing LMM, terms implying unbalanced data, terms representing nuisance effects (e.g., block effects) and terms whose levels represent a random sample

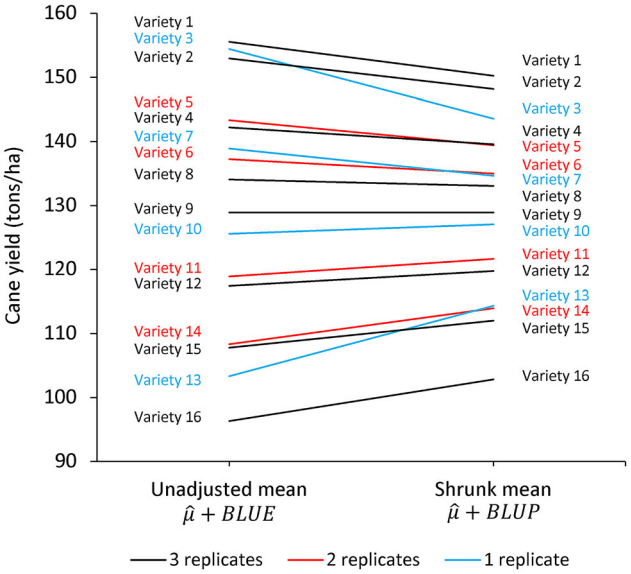

Fig. 2 Illustration of the shrinkage property of the BLUP methodology for random-effect assessment: comparison between unadjusted means ($\hat{\mu} + BLUE$) and shrunk means ($\hat{\mu} + BLUP$) of 16 sugarcane varieties for cane yield (tons/ha) experimented in a complete randomized trial implying an unbalanced number of replicates per variety (see Electronic Supplementary Material 1 for data and models). Unadjusted variety means inferred from a model-based analysis considering variety as a *fixed-effect* term are the addition of the general mean ($\hat{\mu}$) and the *best linear unbiased estimation* (BLUE) and represent the arithmetic means of the varieties. Shrunk variety means represent the addition of general mean ($\hat{\mu}$) and the *best linear unbiased prediction* (BLUP) of variety considered as a *random-effect* term in a mixed model solved using the *residual maximum likelihood* (REML) algorithm. The BLUP methodology is characterized by shrinkage properties, i.e., above-average individual means are shrunk downward toward the overall mean, whereas below-average individual means are shrunk upward toward the overall mean. Ranking of varieties tested on the basis of the same number of replicates remain unchanged. The lower the number of replicates, the greater the shrinkage effect. The shrinkage effect adjusts performance all the more as it is either extremely high or extremely low

from an infinite population of possible levels (e.g., clones or families) should usually be considered as random effects (Galwey 2014). Each random variable generates a variance component which constitutes a fraction of the total phenotypic variance. However, terms with a relatively modest number of treatments (e.g., if only a few cultivars are compared in a trial) are frequently considered as fixed variables. This choice may allow for valid statistical comparisons between treatments but generally provide statistical findings that should be regarded meaningful only in the context of the studied trial (local inference).

LMMs may allow powerful analysis of experimental data, particularly when dealing with complex designs. For example, many flexible options are currently available to test and compare different structures of VCOV (Littell et al. 2006) of random (***G*** matrix) and/or residual (***R*** matrix) effects (variances homogeneous or not; zero, constant or heterogeneous covariances with possible links of dependence or not). Options of VCOV structures chosen to be tested depend on investigator's motivations and their ability to explain data can easily be compared with likelihood-based criteria (AIC, BIC, likelihood ratio tests). Specifications of LMMs possibly using particular VCOV structures and other functions can easily accommodate presence of natural field variations at local and/or global scales for more realistic analyses (Gilmour et al. 1997). Moreover, LMMs allow gathering information of several experiments that share common genetic entries (e.g., family or genotypes) in a single analysis with specific within-trial error variance (Smith et al. 2005; Atkin et al. 2009; Wei et al. 2010; Jackson et al. 2007) with possible combinations of factors such as between harvests and locations (Smith et al. 2007; Pastina et al. 2012; Balsalobre et al. 2016). Such strategies of meta-analyses of trials may allow much broader and accurate results and conclusions than analyses of individual experiments.

Statistical Analysis of Selection Trials

Because selection experiments are expensive, researchers wish to reduce the probability of failure in interpreting their data and making selections when evaluating their genetic entries (families or genotypes). The four basic pillars of experimental designs are: (1) randomization of the assignment of genetic entries to the different experimental units; (2) blocking by grouping experimental units (usually plots) with similar features (e.g., soil type) together; (3) replication of genetic entries; and (4) choosing optimal size of experiment units (Cochran and Cox 1957; Casler 2015). Broad-sense heritability (h^2_{bs}) of the investigated traits represents a key statistical indicator of the efficiency of a selection trial and the accuracy of measurement protocol. The higher a heritability value, the better the degree of confidence that the experimenter can give to the phenotypic values of the varieties to estimate their genetic values. Broad-sense heritability can be defined, at the trial mean level, as the following ratio of the genetic variance (σ^2_g) over the phenotypic variance of variety means (σ^2_P):

$$h^2_{bs} = \sigma^2_g / \sigma^2_p = \sigma^2_g / \left(\sigma^2_g + \sigma^2_e / n \right) \quad (3)$$

where n represents the number of replicates per variety and σ^2_e the within-trial environmental variance, commonly referred to as error variance or residual variance. The higher the within-trial environmental variance, the less accurate the estimate of the mean value of the tested genetic entries. To maximize genetic progress (ΔG) in selection decisions resulting from a breeding experiment (Eq. 1), it is important to optimize the heritability

parameter by controlling as much as possible any sources of errors likely increasing environmental variance.

Seedling Assessment Trials

Seedling experiments raised from seed germination, which is the first selection stage, usually consist of replicated family plots. Programs worldwide may differ in how these experiments are handled depending on resources available (human, land, equipment), nature of data recorded and scope of information expected to be gained. The best strategy in terms of genetic gains and costs is a combined-family and within-family selection (Kimbeng and Cox 2003). Family appraisal is commonly based on several replicates (Shanthi et al. 2008; Pedrozo et al. 2011; Barbosa et al. 2012; Zhou 2013). A few dozen of seedlings (Wu et al. 1978; Leite et al. 2009) per family plots with a minimum of two rows in any replicated trial design were considered by Leite et al. (2006) as being appropriate to evaluate family means, which is a reasonably good indicator of the proportion of elite clones (Chang and Milligan 1992).

Data records of family plots for cane yield and sucrose content analyzed with mixed models (Chang and Milligan 1992; Atkin et al. 2009; Stringer et al. 2011; Neto et al. 2013) not only allow ranking objectively family performance but also can provide a prediction of the breeding value of parents (see "Selection of Parents"). Intensity of seedling selection within families may vary with increased intensity in families with lower rankings. In case of logistical constraints, combined-family and within-family selection can rely on visual rating of yield on a semi-quantitative scale (Daniels 1972). Performance of individual seedlings may be strongly affected by competition and environmental effects (Skinner et al. 1987). However, despite these potential problems, Jackson (2018a) showed significant correlations between visual ratings and subsequent performances in next clonal stage, even with sugar content (more vigorous seedlings tended to have higher sugar content in the next stage). Using a stochastic simulation modeling, he also showed that overall genetic gains in the first two stages of selection were not very sensitive to selection intensity in seedlings.

Clonal Assessment Trials

Clonal assessment trials in sugarcane selection schemes following the seedling stage consist of early and advanced selection trials (Fig. 1) in a multi-stage process that frequently lasts about 10 years. Whatever the experiment design to be analyzed, LMM and its BLUP methodology is the tool of choice to estimate genotype values and genotype differences (especially when dealing with unbalanced or incomplete data set) and to incorporate effects of natural field variations to reduce residual error variance and therefore increase the broad-sense heritability (Eq. 3) and statistical power of analyses (see "Spatial Analysis of Field Trials").

Early-Generation Trials The first clonal stage of selection usually consists of single-row plots with many (thousands) un-replicated candidates tested in large trials. Sugarcane breeders may use various trial layouts (augmented, or row–column (RC) or incomplete blocks) or p-rep design (Cullis et al. 2006) enabling adjustment for environmental heterogeneity and repeated plots of check cultivars to estimate error. Greater weighting is usually given to traits having a reasonably good repeatability such as sugar content, disease susceptibility and flowering propensity (Skinner et al. 1987). The question of how much relative weighting to give to different traits to optimize overall gain from selection can be difficult to resolve precisely. However, selection index theory (see "Use of Selection Index") can be used to objectively help address this complex question. Selection for cane yield (CY) based on direct measurement if enough resources (human and/or equipment) or indirectly through a properly calibrated logistic regression using agro-morphological components of CY (Zhou 2018), should be considered cautiously due to competition effects between plots of small size (Jackson and McRae 2001). This phenomenon can inflate or reduce the yield of strong or of weak competitors, respectively, while competitive ability is of no value in commercial fields of pure stands.

Advanced Selection Trials The more advanced stages of selection usually consist of replicated trials of multiple-row plots (e.g., latinized RC or alpha designs or randomized complete blocks). The reduced effect of competition between varieties allows selection for yield with more confidence. Some large programs routinely assess sugarcane quality (fiber, brix, pol, purity and derived sucrose content) with automated NIRS evaluation systems using shredded cane (Berding and Marston 2010; Taira et al. 2010; Roussel et al. 2015; Koonjah et al. 2019) which require minimal sample preparation and are high speed and cost-effective. Decisions on selection can be done with an index of several traits (see "Use of Selection Index") and culling levels for some pest/disease traits. The choice of the selection stage from which experimental genotypes would deserve to be tested in several locations may be guided by some studies of genotype × environment (GE) interaction (see "Analysis of genotype × environment interactions"). Resistance to important local diseases not efficiently screened for and which requires costly artificial inoculation

tests is usually considered at the end of the scheme when the number of elites still in the race is reasonably modest.

Analysis of Genotype × Environment Interactions

Investigating genotype × environment (GE) interactions is useful when striving to improve efficiency of breeding programs in terms of best management of resource allocation. GE interaction can be analyzed using several trials with common genetic entries (e.g., family or genotypes), such as multi-environment trials (MET) in final selection stages. LMM and its BLUP methodology is the tool of choice for analyzing GE interactions. It allows unbiased estimates of the variance components from frequently imbalanced datasets. Imbalance can arise, for example, in different replicates among clones within a trial or differences in clones tested across trials.

GE studies in sugarcane often partition total variation into genotype (G), location (L), crop-year (C) and GL, GC and GLC interactions to assess the significance of each interaction components. In this terminology, 'crop-year' refers to the joint (confounded) effect of different crops (plant, first ratoon, second ratoon, etc.) and years. For example, a MET consisting of a randomized complete block design repeated at multiple locations could be analyzed with the following model:

$$Y_{jklm} = \mu + L_j + C_{k(j)} + R_{l(j)} + \underline{G}_m + \underline{GL}_{mj} + \underline{GC}_{mk} + \underline{GLC}_{mjk} + \underline{e}_{jklm} \quad (4)$$

where Y_{jklm} designates the observation of genotype m in replicate l in crop year k at location j, μ is the grand mean, L_j is the location main effect at j, $C_{k(j)}$ is the crop year effect for k at location j, $R_{l(j)}$ is the replication effect for l at location j, $\underline{G}_m$ is the main effect for genotype m, $\underline{GL}_{mj}$ is the effect of the genotype × location interaction, $\underline{GC}_{mk}$ is the effect of the genotype × crop-year interaction, $\underline{GLC}_{mjk}$ is the effect of the genotype × crop year × location interaction and e_{jklm} is the error term. In this example of model, the underlined terms relating to genotypic (G) and interaction (GC, GL and GLC) effects in Eq. 4 represent random effects from independent Gaussian variables and are estimated on the basis of BLUPs, while the other terms not underlined represent fixed-effect variables. In this model, the $\boldsymbol{G}$ matrix of VCOV of genotype effects has a simple identity structure ($\boldsymbol{G} = \sigma^2_G \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix whose dimension equals the number of genotypes). Only these four random variables and the residual ones impact the variability of the phenotypic data. Partition of phenotypic variance (σ^2_P) into elementary variance components ($\sigma^2_P = \sigma^2_G + \sigma^2_{GL} + \sigma^2_{GC} + \sigma^2_{GLC} + \sigma^2_e$) in MET studies allow comparing magnitude of genotypic variance (σ^2_G) to that of genotype × environment (GE) interactions ($\sigma^2_{GL} + \sigma^2_{GC} + \sigma^2_{GLC}$) and these latter between each other. A GE interaction variance that would be significant (e.g., 30% or greater) in comparison with the G variance, would justify a selection program developed across multiple representative locations and/or data collecting during several crop years. Moreover, in cane industries composed of relatively similar agro-climatic cultivation zones, the size of the genotype × crop year (GC) interaction variance might be much larger than that of the genotype × location (GL) , and in this case, it would be appropriate to evaluate clones across several crops than on many locations. Conversely, if GL interaction variance was greater than that of GC interaction, evaluation across different locations would be more important than evaluation across different crop-years (Guilly et al. 2017). Under a scenario of a high GE interaction, it may be more efficient to conduct multiple breeding or selection programs and METs within each program.

GE studies based on MET data can also be conducted using graphical tools such as *Genotype plus Genotype × Environment* (GGE) biplot (Yan and Tinker 2005, 2006). GGE biplots allow investigation in more detail of the relationships between environments and the response of genotypes across environments. It is an effective method based on principal component analysis (PCA) to fully explore GE interaction pattern in MET data. A two-way table of adjusted genotype means × locations ($\overline{Y}_{mj}$) is first centered to the mean trait value (μ_j) of each location j and divided by its standard deviation (s_j) to obtain a 'standardized GGE matrix.' This standardized GGE matrix is then subjected to singular value partitioning between genotype and environment eigenvectors (Yan 2002). Biplots based on the first couples of principal components (PCs) graphically approximate variation of MET data. This allows visualization at a glance of the level of proximity between environment as well as genotypes' yield potential and their stability across the environments. Figure 3 shows an example of a GGE biplot for a trait along with its visual interpretation.

Many GGE studies have been carried out in sugarcane breeding programs for various objectives. These include determining the appropriateness of specific test sites for evaluating performance of families and clones for an industry (Ramburan et al. 2012) and checking dissimilarity levels among sites to reduce redundancy in collected data (Glaz and Kang 2008; Guilly et al. 2017) and to assess and make recommendations on potential new cultivars depending on their yield potential and stability (Luo et al. 2015). GGE studies may be very useful in guiding decisions to optimize resource allocation among selection stages (Milligan 1994; Brown and Glaz 2001) and to assess tradeoffs between site number and crop (ratoon) number within stages of selection (Ramburan and Zhou 2011), or to

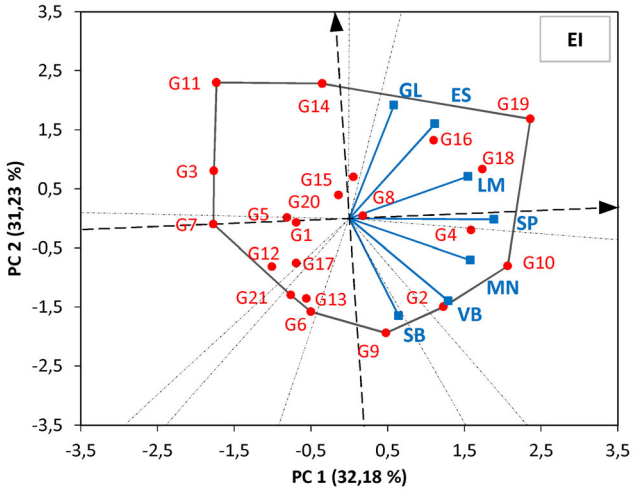

Fig. 3 An example of genotype plus genotype × environment (GGE) biplot for the economic index (EI) data of 21 genotypes (G1 to G21) tested in seven environments (GL, ES, LM, SP, MN, VB and SB) reproduced here with courtesy of Guilly et al. (2017). The biplot allows direct visualization of the magnitude of genotypic and environmental variation in the same units for both principal components (PCs). Both PCs explained globally a large part of the GGE data (63.41%) for the studied trait. Environment vectors have positive abscissas and the variation among them is first discriminated by PC2, while variation among genotypes is discriminated by both PC1 and PC2. The vector view provides a succinct summary of the interrelationships among environments. The smaller an angle between two vectors, the more similar the two environments are in terms of genotypic response for the studied trait. The orthogonal projection of varieties onto each environment vector approximates the ranking of their performance in each environment. The polygon formed by connecting the genotypes that are further away from the biplot origin contains all genotypes. The orthogonal reference frame defined by the two dotted arrows defines average environment vectors (AE) for PC1 and PC2 scores of all environments. Genotypes with the highest (or lowest) abscissas in this frame pinpoint the genotypes exhibiting the highest (or lowest) mean performances across all environments. The highest ordinates in absolute value (either positive or negative) in this frame pinpoint the genotypes exhibiting the highest instability of their performance across environments

provide benchmarks for designing best strategies of meta-analysis of trials for faster genetic gains (Jackson et al. 2007).

The analysis of data of multi-local trials obtained from multiple harvests can also be conceived by omitting the terms of GE interactions (GL, GC, GLC) in the conventional model of analysis (Eq. 4) and by considering non-null co-variances between genetic effects across harvest-by-location environments. The main idea of this alternative strategy of analysis is to test for and exploit genetic correlations between trials and between harvests arising from repeated measures on the same genotypes taken across locations and harvests. This view of the existence of genetic correlations between trials and between harvests in multiple-harvest-location (M-H-L) experiments of perennial crops (such as sugarcane) is intuitive and appealing. It could provide more efficient analysis (and therefore more accurate results) than the conventional analysis since it may capture more complex relationships between genotype performance in different environments that can exist between many M-H-L data. Implementing such an analysis is relatively complex although the mathematical expression of ad hoc models consists in a simplification of the conventional model (Eq. 4 in which all interaction terms are removed). Briefly, the greater complexity of the analysis arises from the fact that it is necessary to consider a $\boldsymbol{G}$ matrix of VCOV of genotype effects much larger than that of the conventional model ($\boldsymbol{G} = \sigma_G^2 \boldsymbol{I}$) and more elaborate. The increase in the dimension of the $\boldsymbol{G}$ matrix is proportional to the combining number between locations and harvests and the $\boldsymbol{G}$ matrices to be tested which need to have nonzero covariances between many genetic effects (off-diagonal values), are necessarily more elaborate than simple basic matrices (identity or diagonal ones). Moreover, to find the best fitting model a selection process of models is required between different variance models for genetic effects -$\boldsymbol{G}$- and different variance models for residuals effects -$\boldsymbol{R}$- (Smith et al. 2007; Balsalobre et al. 2016). Therefore, advanced skills in mixed modeling are required. The paper of Smith et al. 2007 provides a comprehensive strategy of analysis and presents appropriate models and informative case studies.

Despite their relative complexity, analyses of M-H-L data that test for genetic correlations across trials and across harvests and provide information on their pattern have great advantages as illustrated by several studies. When applied to MET data in the final assessment stage of a selection program, this strategy of combined analysis of all M-H-L data can predict genetic value of candidates (BLUP) for individual harvest by location combinations more accurately than in conventional separate analyses (Smith et al. 2007). Thence, recommendations of variety release for local adaptation can be improved as well as recommendation for broad adaptation based on the best mean performance across environments (using a selection index calculated with variety BLUP of each environment). When applied to early selection stage from multi-local experiments, this combined analysis can capture in a more realistic way both the heterogeneity variance and potentially complex covariance structures existing between locations and crop-years at genetic level (Balsalobre et al. 2016). Therefore, higher broad-sense heritability values of important agronomic traits can be obtained and subsequently increased genetic gains in selection decisions. When applied to sugarcane QTL studies relative to quantitative traits, the combined analysis paying attention to model dependencies (correlations) between harvests and locations allows to identify stable QTLs that can be distinguished from environment-sensitive QTLs, which

contributes to better understanding of genetic architecture of complex traits (Pastina et al. 2012).

Spatial Analysis of Field Trials

Spatial variations within field trials usually arise to some degree because of variation across trial in soil type or any other factor affecting sugarcane growth. This reduces accuracy of genotype comparisons and can be particularly problematic in early-generation trials using limited replications. Spatial variation inflates environmental variance within trials (σ_e^2) and therefore reduces broad-sense heritability (h_{bs}^2) of traits and consequently genetic gain (ΔG) predictable from selection (Eq. 1). Investigation of small- or large-scale variations in a field can be done following the methodology proposed by Gilmour et al. (1997). This consists of a multi-step approach aiming at selecting the best spatial mixed model among different mathematical options using diagnostic tools (data fit statistics, variogram of residuals, trellis plots). In early-generation sugarcane trials, corrections with this multi-step approach can result in more objective estimates of variety effects for cane yield and sugar content (Matassa et al. 1998; Stringer and Cullis 2002; Edme et al. 2007). Even in advanced multi-local trials where several replicates facilitate more precise estimates of genotype effects, correction for spatial variations based on anisotropic autoregressive models within trials (Smith et al. 2007; Ostengo et al. 2015) can contribute to gains of precision (when considering among-trial heteroscedasticity). A new spatial model, using bi-dimensional penalized spline functions (Velazco et al. 2017; Rodríguez-Álvarez et al. 2018) formulated in the framework of LMM (Lee et al. 2013), can also provide smooth curve-fitting maps of any type of spatial variation in large sugarcane trials for any quantitative trait (Hoarau et al. 2019). Though attractive for routine analysis of a large number of trials (no procedure of model selection), the current implementation of this model (Rodríguez-Álvarez et al. 2016) does not yet allow flexible covariance structure between some random effects (in $\boldsymbol{G}$ or in $\boldsymbol{R}$ matrices). In addition, modern technology such as electromagnetic induction mapping (based on the electrical conductivity of the soil) used to characterize environment could be utilized to account for soil variability in a trial (Wei et al. 2015).

Analysis of Disease/Pest Resistance Trials

Evaluation of elite candidates for resistance to diseases and pests constitutes a major component of sugarcane selection programs. Analysis of variance (ANOVA) which is the most common method of analysis for designed experiments is not suitable for many types of resistance traits. This analysis is usually based on an assumption that experimental errors are independently and normally distributed with a homogeneous variance (Steel and Torrie 1980). This assumption is often not met for data collected as percentage (e.g., % infected plants), counts (e.g., number of symptoms), binomial data (e.g., disease incidence) or data based on a short ordinal measurement (e.g., 1–4 scale). A common approach is to transform the data in the hope that it will better meet this necessary assumption. In case of failure, nonparametric statistics using ranks (such as Friedman or Kruskal–Wallis tests) could be used but exploit less information than available in the data and have a lower power. The best solution is to use generalized linear mixed models (GLMMs) with appropriate link functions (log, logit or probit). GLMMs allow properties of many resistance traits to be handled more easily (Engel and Keen 1994), as illustrated by several resistance studies to some sugarcane bioagressors (Ahmad et al. 2007; Gouy et al. 2013; Fartek et al. 2012). GLMM offers the great advantage of easily handling joint analysis of trials sharing varieties in common. Sugarcane selection programs usually conduct routine resistance screening trials for important diseases including a common set of susceptible, tolerant and resistant standards. Stringer et al. (2012, 2013) proposed an interesting multiple step procedure of analyses using GLMM to exploit all historical data of standards used in current trials to best interpret these. This procedure may reduce the risk of discarding high-yielding varieties incorrectly rated as susceptible or releasing a variety incorrectly rated as resistant that could suffer losses in case of epidemics.

Quantitative Genetics to Optimize Selection

Estimation of Genetic Parameters

Narrow-sense heritability (h_{sn}^2) of a trait is the key parameter to predict gains from selection of parents for a new cycle of crossing. This parameter can be estimated from experiments of unselected progenies through biometrical approaches (parent–offspring regressions) or quantitative genetic studies using mixed models. Generally speaking, narrow-sense heritability estimates for any trait are specific to the particular populations of genotypes studied and may vary greatly from one population to another. In commercial sugarcane breeding programs, estimates of h_{ns}^2 for cane yield (CY) are generally low (i.e., < 0.20) (Gravois et al. 1991; Neto et al. 2013; Pisaroglo de Carvalho et al. 2014; Jackson 2018b). This is reflective of the importance or predominance of non-additive gene action relative to additive genetic effects (Hogarth et al. 1981; Neto et al. 2013; Pisaroglo de Carvalho et al. 2014; Zhou 2019; Mendes de Paula et al. 2020). Hence, performance of clones for cane yield tends to have a low

predictive value of their general breeding value (BV). Estimates of h^2_{ns} for sucrose content are variable from one population to another one but are frequently moderate (Gravois et al. 1991; Hogarth et al. 1981; Liu et al. 2007) to high (Ramdoyal and Badaloo 1998), and therefore, BVs of clones are predictable on the basis of their clonal performance. For disease resistance, moderate h^2_{ns} estimates have been reported for smut (Wu et al. 1988; Chao et al. 1990), rust (Hogarth et al. 1993; Ramdoyal et al. 2000), SCYLV (Hoarau et al. 2018) and leaf-scald (Bressiani et al. 2007) and high estimates for resistance to a borer species (White et al. 2001) in different experimental populations. Therefore, for these bioagressors, selection of parents based on their resistance levels may be effective. Considering the breeder's equation (Eq. 1) in the varietal creation phase, the breeder would gain by estimating periodically the narrow-sense heritability of target traits in his current populations of genitors to check whether he could expect or not significant values. Knowledge of narrow-sense heritability in breeding populations enables a breeder to adjust criteria for selection of parents. If h^2_{ns} is large for a trait, then greater weighting can be placed on direct phenotypic performance of candidate parents (e.g., with resistance to diseases), while if h^2_{ns} is low, then assessment of progeny performance is usually required to obtain an accurate assessment of breeding value of the trait (e.g., with cane yield).

Breeding for increased sucrose yield may be constrained in some breeding program populations by slightly negative genetic correlations between cane yield and sucrose content (Kang et al. 1983; Gravois et al. 1991; Milligan et al. 1996; Jackson 2005; Dumont et al. 2019) and by limited genetic variability for sucrose content (Jackson 2005; Hogarth et al. 1997). Rapid improvement in short breeding cycles is feasible for sucrose content (Kennedy 2005) due to its relatively good heritability but intense selection pressure for sucrose content may adversely impact on cane yield because of a slightly negative correlation between these two components. This can be explainable by either loci located in or associated with pleiotropic genes having an opposite effect for cane yield and sucrose content (Pinto et al. 2010; Magarido et al. 2015). These breeding constraints justify the use of selection indices integrating the relative economic importance of all target traits, their heritability and genetic correlations. Their use can make the work of breeder more predictable, practical and rational, and improve genetic gains for overall economic value which is affected by multiple traits (see "Use of Selection Index").

Selection of Parents

The success of any sugarcane breeding program largely depends on the choice of the parent clones and the parental combinations. However, because there are numerous traits of commercial importance that must be considered in breeding, the criteria and strategies for choosing parents are complex, as the economic and genetic parameters of many traits need to be taken into account simultaneously. Parents may come from advanced trials due interesting clonal values (sucrose yield, disease resistance, ratooning ability, etc.), from introduced foreign germplasm, or from specific improvement programs (e.g., high sucrose or base-broadening germplasm). Given the high cost of maintenance, the management of a parent collection is a dynamic ongoing process with regular discard of relatively poorer performing parents to enable addition of better new ones. Breeding strategies usually exploit general combining ability (GCA) and specific combining ability (SCA) of parents and parental combinations, respectively. GCA is mostly affected by additive genetic effects[1] and SCA mostly affected by non-additive effects[2] (Falconer and Mackay 1996). Unlike GCA, SCA is not predictable since non-additive effects are transmitted only weakly. Usually, a large number of crosses are made to identify the small number of elite crosses that may produce a higher proportion of superior progenies and increase their number of seedlings. In order to maximize the rate of genetic gain over time (ΔG) in a recurrent selection approach (aiming at ongoing improvement of parent population), the best strategy to rationalize resource use is to select and cross the parents showing highest GCA. The GCA of a parent for a trait, i.e., its general breeding value (BV), is predictable with the mean performance of a set of derived bi-parental progenies and of closely related progenies of any generation level (Piepho et al. 2008). To improve a parent population over a long term for several traits at the same time (which is the strategy of most sugarcane breeding programs), the use of a selection index combining BV of all traits is the most rational option for choosing parents to favor in crosses and parents to discard (see "Use of Selection Index" section). BVs for a trait can be estimated with progeny assessment data (see "Seedling Assessment Trials") through a LMM and its BLUP methodology using an additive relationship matrix between all parents inferred from their pedigree (Piepho 2008; Barbosa et al. 2005; Neto et al. 2013). Atkin et al. (2009) showed in the Australian sugarcane

[1] An effect of a gene that is related to the number of copies of each allele in a linear or additive fashion and whose effects are independent to variation in other genes.

[2] Interaction effects between alleles within or between loci, of nature potentially complex.

program that a minimum of three generations of pedigree and five years of historical data allowed good estimates of parent BVs (in reasonable computing times). Highly significant correlations between mid-parent predicted BVs and family performances were obtained by Wei et al. (2013) for cane yield and sucrose content and indicated that a strategy of parent selection driven by BVs allowed production of higher-performing families with greater chance for selecting superior offspring.

Use of Selection Index

For sugar or sugar/bio-energy production, an ideal variety should be superb in a number of characteristics, which generally include high yielding, high and suitable sugar content and quality, resistance to disease and tolerance to abiotic stress. Unfortunately, it is extremely difficult to find a genotype possessing all the desirable characteristics and almost all genotypes in a selection program have both positive and negative characteristics. The requirement of improving multiple traits and the rare combination of all desirable traits make it critical to evaluate each clone by compromising objectively the negatives with positives. For example, to compare two clones, one with high yield but lower sugar content and another with opposite characteristics, we need to know how high in yield (or sugar content) can compensate the lower sugar content (or yield). Selection index theory was developed to specifically address these kinds of issues and has been applied successfully in many genetic improvement programs since 1930s (Smith 1936; Hazel 1943).

There are two key concepts or steps to take in developing a selection index: defining the breeding objective and deriving the selection index. The breeding objective is simply the purpose of a breeding program. In most sugarcane breeding programs, for example, the breeding objective may be reasonably defined as maximizing the economic profits per tonnes of sugar produced, for the whole industry, as in the example reported by Wei et al. (2006b) for the Australian sugarcane industry. For sugar production, this objective was expressed as:

$$H = v_{TCH} * GV_{TCH} + v_{CCS} * GV_{CCS} + \ldots \quad (5)$$

where v_{TCH} is the economic weight of cane yield in tonnes of cane per hectare (TCH), which is defined as the additional economic profit per tonnes of sugar by improving cane yield by one unit (1 tonnes/hectare) and the GV_{TCH} is the genetic value of the variety under investigation. v_{CCS}, GV_{CCS} and terms for other traits are defined by the same way. All the traits (such as commercial cane sugar, CCS) that could be important in affecting the profits of sugar production and which may be selected for in the breeding program should be included in the objective equation. The determination of economic weights entirely depends on the processes within a production system (e.g., growing, harvesting, transport, milling and marketing) and how a variety can impact on the cost of those processes per unit of product (sucrose) produced (Wei et al. 2006b). Therefore, developing a selection index in any breeding program starts from defining the targeted production system, collecting all the economic data from establishing a crop in the paddock to the final market where the product is sold. This is followed by examining how sugarcane traits will impact on the revenue, costs and (by difference) the profits per tonnes of sugar produced for the targeted industry. The impact on profit expressed in terms of economic value per unit change of the trait is the economic weight for that trait. This approach can be illustrated by an example with the harvesting operation and cane yield/sugar content. It is assumed for this example that maximizing economic profit for the whole industry from sugar production is the breeding objective and also that costs of harvesting one tonnes of cane are \$7 and average sugar content is 14%. An improvement in cane yield will have no impact on the costs of harvesting per tonnes of sugar because the costs of harvesting increase in proportion to increased sugar production and revenue. However, improving CCS by 1% would have an impact as the costs in terms of producing one tonnes of sugar from the perspective of harvesting has changed from \$7/(1 TCH * 14%) = \$50/tonnes of sugar to \$7/(1 TCH * 15%) = \$46.67/tonnes of sugar. That is, by improving sugar content by one unit (from 14 to 15%), the increased profit from producing one tonnes of sugar attributed specifically to decreased harvesting costs will be (\$50 \$46.67) \$3.33/tonnes sugar.

Selection index theory is a methodology to help breeders to maximize the rate of progress in relation to their breeding objective. It is not obvious to many breeders that all the traits identified in the breeding objective refer to the measurements under commercial production environments. For example, some traits (e.g., physiological traits or yield components) may not have direct economic value but may still be useful in a selection index if they have a high correlation with another trait which does have direct economic value. For example, cane yield measured in early selection stage trials in single-row plots which may be strongly affected by competition effects has some degree of correlation with yield measured in a pure stand. Therefore, development of an optimal selection index needs to reflect such kind of issues to have greatest utility. Conventionally, a selection index is normally expressed as:

$$I = w_{TCH} * P_{TCH} + w_{CCS} * P_{CCS} + \ldots \quad (6)$$

where w_{TCH} is a coefficient for cane yield and P_{TCH} is the phenotypic measurements for cane yield based on experimental trials. Unlike the economic weights in

breeding objective, index coefficients are derived from variance components and economic weights as follows (White and Hodge 1989):

$$\mathbf{w} = \mathbf{P}_{\mathrm{II}}^{-1}\mathbf{G}_{\mathrm{IH}}\mathrm{v} \quad (7)$$

where $\mathbf{w}$ is a vector of index coefficients for all traits collected from trials and included in the selection index to predict the genetic values of traits included in breeding objective; $\mathbf{P}_{\mathrm{II}}^{-1}$ is inverse of the phenotypic variance–covariance (VCOV) matrix among all the traits in selection index (Eq. 6); and $\mathbf{G}_{\mathrm{IH}}$ is genetic VCOV matrix between traits in selection index (Eq. 6) and traits in breeding objective (Eq. 5).

Nowadays, data from trials are generally analyzed by methods described above and BLUPs for each of selection criteria would be available for constructing selection index. Under this circumstance, a similar index as in Eq. 6 can be constructed with BLUPs replacing the phenotypic values and the coefficients estimated as (Schneeberger et al. 1992):

$$\mathbf{w} = \mathbf{G}_{\mathrm{II}}^{-1}\mathbf{G}_{\mathrm{IH}}\mathbf{v} \quad (8)$$

where $\mathbf{G}_{\mathrm{II}}^{-1}$ is the inverse of the genetic VCOV matrix among selection criteria included in the selection index and $\mathbf{G}_{\mathrm{IH}}$ and $\mathbf{v}^{-1}$ are same defined as in Eq. 7. Apparently if all the traits in the selection index (Eq. 6) are the same as those in breeding objectives (Eq. 5) and also their estimated genetic values are perfectly associated with the genetic values in breeding objective, Eq. 8 becomes $\mathbf{w} = \mathbf{v}$; in other words, the selection index (Eq. 6) is identical to breeding objective (Eq. 5).

It should be noted that comparing with the economic weight relatively is more important to accurately estimate the genetic values. Accuracy of estimation can be improved by experimental design, field operation and statistical models, which are covered in the above sections.

Plant breeders usually prefer to compare the testing clones with known cultivars as standards or checks in the same trials. For this purpose, selection index values for experimental clones can be scaled in relation to the mean selection index values of all standard cultivars (or checks) in the same trials. This scaled value is known as relative economic genetic value (rEGV), and an example of this is illustrated by Wei et al. (2008) for the Australian sugarcane breeding program.

Analysis of DNA Marker Data

Applications using DNA markers have been developed to help speed-up genetic gains (ΔG) in sugarcane breeding programs by reducing duration of breeding cycles in the breeder equation (Eq. 1). However, the complex genome of sugarcane makes genomic applications more difficult than for simple diploid plants. Modern sugarcane cultivars are highly polyploid and aneuploid hybrids which contain 100–130 chromosomes and 10–13 hom(e)ologous copies of most loci (D'Hont et al. 1996, 1998; Piperidis and D'Hont 2020). Genomic applications based on the study of DNA polymorphism need to be tailored to this complex genome organization. This section only deals with the principles and issues of statistical genetics related to development of genomic applications based on markers and with scientific and practical results obtained in breeding. These principles of statistical analyses apply to any type of markers, i.e., low-to-medium (RFLP, SSR, EST-derived SSRs, AFLPs, DArTs) or high (next-generation sequencing) throughput marker systems. A full presentation of DNA marker technologies developed for sugarcane is beyond the scope of this paper. For this matter, the reader can refer to Aitken (2021) in this special issue.

Quantitative Trait Allele (QTA) Mapping

A QTA study aims to dissect the genetic architecture of a trait by localizing the chromosomal regions containing causal genes (loci) and by estimating the nature (additivity, dominance and epistasis) of the effects of their alleles and allele effect size. Discovery of alleles of significant effect constitutes the initial step to develop marker-based breeding approaches in support to conventional breeding. Many QTA studies have been performed in sugarcane on selfed or bi-parental progenies by analyzing low-dose markers, i.e., single dose (SD) and double (DD) markers used to construct genetic linkage maps (Alwala and Kimbeng 2010). These latter contain large gaps due to the relative lack of SD and DD markers in genomic zones of the genome that may be redundant. Statistical methods for QTA search include: (1) single analysis of mapped or unlinked SD and DD markers and di-SD epistatic interactions through ANOVAs (Hoarau et al. 2002; Reffay et al. 2005; Aitken et al. 2006, 2008; Piperidis et al. 2008; Nibouche et al. 2012) or regression models and likelihood ratio tests (Pinto et al. 2010; Santos et al. 2015); (2) simple interval mapping (SIM) on linkage maps constructed for each parent with their specific SD markers (Ming et al. 2002); composite interval mapping (CIM) on separate parental maps (Aljanabi et al. 2007; Alwala et al. 2009; Singh et al. 2013; Yang et al. 2018) or on a joined map (Ukoskit et al. 2019) or an integrated map (Garcia et al. 2006; Margarido et al. 2007) of both parents (Balsalobre et al. 2017; Gutierrez et al. 2018) using co-factors selected outside mapping regions to increase statistical power; or (3) more sophisticated QTA search methodologies through SIM applied on an integrated map which use mixed models exploiting multiple-harvest-location (Pastina et al. 2012; Costa et al. 2016) or multiple-trait-environment (Margarido

et al. 2015) datasets with appropriate (co)variances structures for modeling heterogeneity and correlation of genetic effects to limit residual variation and increase statistical power for detecting QTA effects. In contrast to diploid plants where comparison is between two alternative alleles per locus, in modern interspecific cultivars there could be about 10 to 13 segregating factors per locus (Piperidis and D'Hont 2020). In such polyploid context with predominantly polysomic inheritance, one intuitively expects to be able to tag only the most (or the least) favorable allele, and only it, among the whole allelic series (Hoarau et al. 2002). In diploids, the detectable QTA repertoires may greatly vary from one variety to another. On the contrary, in the genetic context of sugarcane where the expression of a QTA results from its differential confrontation against a large allelic background, QTA repertoires of different varieties are expected to better overlap than in some diploid species, at least for QTA sets showing the largest effects (McIntyre et al. 2005; Piperidis et al. 2008; Ukoskit et al. 2019).

Sugar Yield-Related Traits The expression of quantitative traits related to sugar yield (sugar content, cane yield and its morphological components) is expected to result from the combined action of numerous alleles and the influence of environmental conditions. Progenies derived from modern interspecific cultivars (Hoarau et al. 2002; Reffay et al. 2005; Pinto et al. 2010; Singh et al. 2013; Costa et al. 2016; Balsalobre et al. 2016) or from crosses between cultivars and *S. officinarum* accessions (Aitken et al. 2006, 2008) usually exhibit QTA of small individual effect (less than 10% of the phenotypic variation explained in large progenies). However, the direction of their effects (positive or negative) is always consistent across years whatever their significance level. These results reflect alleles with large additive effects for sugar content and stalk morphology related to cane yield which have been readily accumulated to multiple doses if positive or eliminated from prior breeding generations or from ancestral domestication if negative (Aitken et al. 2006). Such situation should diminish the internal contrast that determines trait segregations and the magnitude of the QTA effects (Hoarau et al. 2002). Identification of relatively large QTA effects is more common in mapping populations derived from *S. spontaneum* (Ming et al. 2002; Alwala et al. 2009; Ukoskit et al. 2019). Hence, such interspecific progenies should be more suitable to try to develop marker-based breeding approaches to tag unfavorable haplotypes to be purged from cultivated germplasm or novel favorable haplotypes worth to be introgressed.

Resistance to Diseases and Pests Resistance to diseases and pests constitutes a major criterion of selection of all sugarcane breeding programs. Sources of resistance have been searched for through QTA studies for many bioagressors. Nibouche et al. (2012) and Gutierrez et al. (2018) who, respectively, studied resistance to a borer species and to leaf-scald disease failed to detect any major allele of resistance in surveyed varieties. However, Costet et al. (2012a) found a major QTA of resistance for *sugarcane yellow leaf virus* (SCYLV) in the genotype 'MQ76-53' which accounted for R^2 = 32% of the phenotypic variance of a large a progeny (n = 196). Aljanabi et al. (2007) identified a major allele of resistance (R^2 = 23.8%; n = 227) for yellow spot (*Mycovellosiella koepkei*) in the cultivar 'M134/75.' More recently, Yang et al. (2018) found a major allele of resistance (R^2 = 58%; n = 173) for orange rust (*Puccinia kuehnii*) in the cultivar 'CP95-1039' based on its high density SNP linkage map. These studies pave the way for the search of markers tightly linked with these resistance alleles to develop useful breeding applications to increase the level of resistance to these diseases (evaluate the frequency and the effect of these genes among breeding germplasm, identify other potential sources of resistance, screening of resistant clones in breeding programs that may indicate potential parents for future crosses). Such applications have already been developed for brown rust resistance with diagnostic markers associated with the major and durable *Bru1* resistance gene (Glynn et al. 2013; Racedo et al. 2013; Li et al. 2017; Parco et al. 2017) found in the cultivar 'R570' a long time ago (Daugrois et al. 1996). The identification of markers tightly associated with this resistance gene needed long steps of genetic (Asnaghi et al. 2004) and physical (Le Cunff et al. 2008) mapping using sorghum and rice genomes as well as the study of the haplotype bearing it (Costet et al. 2012b). Nowadays, the advent of the first sugarcane reference genomes (Garsmeur et al. 2018; Zhang et al. 2018) should help expedite the design of diagnostic markers of a resistance haplotype that would be detected in a QTA study. Any sequence of interest presumably linked to a target haplotype could be mapped directly onto a sugarcane reference genome to develop diagnostic markers and search for possible candidate genes.

Genome-Wide Association Studies (GWAS)

Genome-wide association studies (GWAS) have become popular in the last decade in sugarcane. This application consists of statistical associations between segregating markers scattered over the genome and the agronomic variations of traits among panels (populations) of genotypes from many different parents, but with a common ancestral basis. Therefore, GWAS enable the discovery of a much greater number of alleles of agronomic interest than in QTA studies. GWAS have been facilitated by the advent

of high-throughput marker systems and the existence of large blocks of linkage disequilibrium in modern sugarcane cultivars ($\sim$ 5 cM) (Raboin et al. 2008). The most common methodology to detect marker-trait associations (MTA) uses the linear mixed model (LMM) of Yu et al. (2006) which takes into account the genetic structure ($\boldsymbol{Q}$) of the surveyed population and a matrix of kinship ($\boldsymbol{K}$) between its individuals inferred from marker data to best control the risk of spurious associations. This LMM can be expressed as follows:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Sm} + \boldsymbol{Qv} + \boldsymbol{Zu} + \boldsymbol{e} \quad (9)$$

where $\boldsymbol{y}$ is the vector of the phenotypic values of the trait of interest; $\boldsymbol{\beta}$ is the vector of all fixed effects related to the trial design; $\boldsymbol{m}$ is the vector the fixed effects of the DNA markers; $\boldsymbol{v}$ is the vector of population effects; $\boldsymbol{u}$ is the vector of polygene background effects of each individual; $\boldsymbol{e}$ is the vector or residuals effects; $\boldsymbol{Q}$ is a matrix of coefficients relating $\boldsymbol{y}$ to $\boldsymbol{v}$; and $\boldsymbol{X}$, $\boldsymbol{S}$ and $\boldsymbol{Z}$ are incidence matrix of 1 s and 0 s relating $\boldsymbol{y}$ to $\boldsymbol{\beta}$, $\boldsymbol{m}$ and $\boldsymbol{u}$, respectively. The variances of the random effects are assumed to be $\mathrm{Var}(\boldsymbol{u}) = 2\,\boldsymbol{K}\,\sigma_g^2$ and $\mathrm{Var}(e) = \boldsymbol{R}\,\sigma_e^2$, where $\boldsymbol{K}$ is an n x n matrix of relative kinship coefficients matrix between all pairs of individuals; $\boldsymbol{R}$ is an n x n identity matrix; and σ_g^2 and σ_e^2 are the genetic and the environmental variances, respectively. When this system of mixed-model equations is solved, the study of P-values related to marker effects allows to identify MTA that are statistically significant.

Many GWAS have been conducted on quantitative agronomic traits related to cane yield and cane quality, either among panels of elite and/or historic sugarcane clones (Wei et al. 2010; Banerjee et al. 2015; Racedo et al. 2016; Barreto et al. 2019; Fickett et al. 2019; Zan et al. 2020) or among panels mainly containing representatives of ancestral species (Yang et al. 2019a, b). Results obtained are all the more directly useful for genomic breeding applications, as GWAS effectively control the risks of false-positive MTAs and that allele effects are found in large populations experimented in robust experiments. GWAS have also been conceived to prospect for resistance alleles to many diseases, such as Pachymetra root rot, leaf scald and Fiji disease (Wei et al. 2006a), smut (Wei et al. 2006a; Gouy et al. 2015); red rot (Singh et al. 2016), brown rust and gumming disease (Gouy et al. 2015); SCYLV (Débibakas et al. 2014; Gouy et al. 2015) or orange rust (McCord et al. 2019). When successful, individual MTA usually do not explain more than R^2 = 15% of the phenotypic variation of large populations due to the very large segregating factors in the studied panels. However, any interesting global values of any set of independent loci in a multiple regression ($R^2 > 50\%$) would pave the way of direct applications in marker-assisted breeding to select higher resistant varieties more rapidly than in often tedious inoculation experiments.

Genomic Selection (GS)

Genomic selection (GS) is a novel method for selecting individuals that have the ability to speed up genetic gains (ΔG) of complex traits in conventional breeding programs (Breeder equation Eq. 1). Its principle is to predict the genetic performance or the breeding value of individuals, on the basis of their genome-wide DNA marker profiles by using a predictive model previously calibrated with a representative phenotyped and genotyped 'Training Population' (Crossa et al. 2011). GS exploits the whole DNA marker information by simultaneously accounting the effect of each marker across the entire genome to predict the genetic value of individuals. Statistical models and algorithms to estimate marker effects and implement GS predictions include parametric (GBLUP, RR-BLUP), Bayesian and nonparametric (Kernels) methods (Lorenz et al. 2011). Only a few studies have investigated GS for sugarcane. Gouy et al. (2013) studied ten agronomically important traits in two small variety panels (167 individuals) from breeding program in Reunion Island and Guadeloupe with a small number of markers (1499 DArT). They found small to substantial (0.11 to 0.62) accuracy values of prediction, depending on traits studied. These results could, however, be regarded as encouraging given the limits of their study (small marker number and training population size). Deomano et al. (2020) studied in Australia three relatively large populations (hundreds to one thousand individuals) with many SNP markers (47–57,000). Range of prediction accuracies for cane yield and sucrose content was found promising (0.25–0.45) and strongly supports the potential usefulness of GS for sugarcane breeding. This latter opinion stems from the fact that, on the one hand, the levels of narrow-sense heritability of yield components in sugarcane (see "Estimation of Genetic Parameters") are low (tonnage of cane) to moderate (sugar content) and that, on the other hand, most current GS models predominantly exploit additive marker effects and very poorly capture non-additive effects. Therefore, genomic prediction based on predominately additive effects of DNA markers would better predict breeding value (BV) and would have the potential to accelerate genetic gains in breeding by shortening generation intervals (Yadav et al. 2020). Experimental work of Castellani et al. (2018) performed in Brazil gives some credit to these prospects. Using 22,000 genome-wide SNP markers, these authors evaluated the accuracy to predict the genomic estimated breeding values (GEBV) of parents using a large set of parents (1276) with progeny data. Phenotypic BVs were estimated using progeny data and

pedigree information. Correlation between observed (BV) and predicted (GEBV) values reached an average of 0.45 and 0.60 for sucrose content and cane yield, respectively. These results strongly support the feasibility of GS to predict breeding values of parents and select the best parents and superior predicted cross combinations. To improve the ability of GS to predict clonal value (i.e., total genotypic value) of candidates for assistance purposes of selection programs, extended statistical models that consider non-additive effects could be beneficial to derive more precise marker effects and, ultimately, additional gains in prediction accuracies (Yadav et al. 2020). Indeed, in a recent study, the base additive GS model (including random additive genetic effects) which was extended with two additional random non-additive effects (dominance and additive-additive epistatic interaction effects) was found to significantly improve the prediction ability (+ 17%) of clonal performance for cane yield (Yadav et al. 2021), a trait well known for its low values of narrow heritability. Finally, for both predictions of breeding value and clonal value, inclusion of allele dosage information in GS models is a matter of high interest in polyploid species. In sugarcane, until now, only pseudo-diploid models have been used to account for heterozygosity. However, polyploidy implies different types of heterozygotes in terms of allele dosage, and this factor contributes to phenotypic variation. Research focused on the influence of allele dosage could help improve the current performance of genomic selection in sugarcane.

Conclusions

This article provides an overall picture of applications of quantitative genetics and statistical analyses developed to support the efficiency of breeding programs dedicated to sugarcane industries. These concerns could be regarded as 'old science,' if major technological innovations had not appeared for two decades allowing to renew ways of thinking in conducting most applications. Major innovations have emerged with concrete achievements to sustain overall genetic gains in sugarcane breeding programs. Beyond the optimization of the designs of variety trials and of their statistical analysis mainly related to skills in biometrics, many innovations have emerged in statistical genetics. The development of modeling approaches allowing inferences to be drawn from experimental and genetic data related to quantitative genetics theory happens to provide very helpful information to drive more efficient breeding programs and therefore save time and costs. This was possible thanks to the advent of powerful software in statistics along with an unprecedented development of computing/algorithmic science. These computing facilities enable solving efficiently more complex prediction issues related to many sugarcane breeding applications using higher volume of data. Despite dreadful pitfalls in any thinking about the use of molecular markers in the highly complex polyploid context of sugarcane, creative and very helpful applications have been devised in support to breeding programs. These help understand the genetic architecture of some key traits, provide markers to target and follow up in breeding and selection processes chromosome regions of agronomic interest and provided the first genomic breeding and selection models using high-throughput genomic information. We believe that future progress in the development of improved or innovative applications in support to sugarcane breeding programs may reside in closer scientific collaborations in statistical genetics between operational breeders, molecular biologists, biometricians and computer scientists, especially in the genomics age.

Declarations

Conflict of interest The authors declare no conflict of interests regarding the publication of this paper.

References

Ahmad, Y.A., L. Costet, J.H. Daugrois, S. Nibouche, P. Letourmy, J.C. Girard, and P. Rott. 2007. Variation in infection capacity and in virulence exists between genotypes of Sugarcane yellow leaf virus. *Plant Disease* 91 (3): 253–259.

Aitken, K.S., P.A. Jackson, and C.L. McIntyre. 2006. Quantitative trait loci identified for sugar related traits in a sugarcane (Saccharum spp.) cultivar × Saccharum officinarum population. *Theoretical and Applied Genetics* 112 (7): 1306–1317.

Aitken, K.S., S. Hermann, K. Karno, G. Bonnett, L. McIntyre, and P.A. Jackson. 2008. Genetic control of yield related stalk traits in sugarcane. *Theoretical and Applied Genetics* 117 (7): 1191–1203.

Aljanabi, S.M., Y. Parmessur, H. Kross, S. Dhayan, S. Saumtally, K. Ramdoyal, L.J.C. Autrey, and A. Dookun-Saumtally. 2007. Identification of a major quantitative trait locus (QTL) for yellow spot (*Mycovellosiella koepkei*) disease resistance in sugarcane. *Molecular Breeding* 19 (1): 1–14.

Alwala, S., and C.A. Kimbeng. 2010. Molecular genetic linkage mapping in *Saccharum*: strategies, resources and achievements. In *Genetics, genomics and breeding of sugarcane*, ed. R. Henry and C. Kole, 69-96. CRC Press, Taylor & Francis Group, Science Publisher.

Alwala, S., C.A. Kimbeng, J.C. Veremis, and K.A. Gravois. 2009. Identification of molecular markers associated with sugar-related traits in a *Saccharum* interspecific cross. *Euphytica* 167 (1): 127–142.

Asnaghi, C., D. Roques, S. Ruffel, C. Kaye, J.Y. Hoarau, H. Telismart, J.C. Girard, L.M. Raboin, A.M. Risterucci, L. Grivet, and A. D'Hont. 2004. Targeted mapping of a sugarcane rust resistance gene (*Bru1*) using bulked segregant analysis and AFLP markers. *Theoretical and Applied Genetics* 108 (4): 759–764.

Atkin, F.C., M.J. Dieters, and J.K. Stringer. 2009. Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. *Theoretical and Applied Genetics* 119 (3): 555–565.

Balsalobre, T.W., M.C. Mancini, G.D.S. Pereira, C.O. Anoni, F.Z. Barreto, H.P. Hoffmann, A.P.D. Souza, A.A.F. Garcia, and M.S. Carneiro. 2016. Mixed modeling of yield components and brown rust resistance in sugarcane families. *Agronomy Journal* 108 (5): 1824–1837.

Balsalobre, T.W., G. da Silva Pereira, G.R.A. Margarido, R. Gazaffi, F.Z. Barreto, C.O. Anoni, C.B. Cardoso-Silva, E.A. Costa, M.C. Mancini, H.P. Hoffmann, A.P. de Souza, A.A.F. Garcia, and M.S. Carneiro. 2017. GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genomics* 18 (1): 72.

Balzarini, M. 2002. Applications of mixed models in plant breeding. In *Quantitative genetics, genomics and plant breeding*, ed. M.S. Kang, 353–363. New York: CABI Publ.

Banerjee, N., A. Siraree, S. Yadav, S. Kumar, J. Singh, D.K. Pandey, and R.K. Singh. 2015. Marker-trait association study for sucrose and yield contributing traits in sugarcane (*Saccharum* spp. hybrid). *Euphytica* 205 (1): 185–201.

Barbosa, M.H.P., M.D.V. de Resende, J.A. Bressiani, L.C.I. da Silveira, and L.A. Peternelli. 2005. Selection of sugarcane families and parents by Reml/Blup. *Crop Breeding and Applied Biotechnology* 5: 443–450.

Barbosa, M.H.P., M.D.V. Resende, L.A. dos Dias, G.V.D.S. Barbosa, R.A.D. Oliveira, L.A. Peternelli, and E. Daros. 2012. Genetic improvement of sugar cane for bioenergy: the Brazilian experience in network research with RIDESA. *Crop Breeding and Applied Biotechnology* 12 (SPE): 87–98.

Barreto, F.Z., J.R.B.F. Rosa, T.W.A. Balsalobre, M.M. Pastina, R.R. Silva, H.P. Hoffmann, A. Pereira de Souza, A.A.F. Garcia, and M.S. Carneiro. 2019. A genome-wide association study identified loci for yield component traits in sugarcane (*Saccharum* spp.). *PLoS ONE* 14 (7): e0219843.

Berding, N., and D.H. Marston. 2010. Operational validation of the efficacy of Spectracane^{TM}, a high-speed analytical system for sugarcane quality components. *Proceedings of the Australian Society of Sugarcane Technologists* 32: 445–459.

Bressiani, J.A., W. Burnquist, and J.A. Da Silva. 2007. Breeding sugarcane for leaf scald resistance: a genetic study. *Journal American Society Sugar Cane Technologists* 27: 15–22.

Brown, J.S., and B. Glaz. 2001. Analysis of resource allocation in final stage sugarcane clonal selection. *Crop Science* 41 (1): 57–62.

Casler, M.D. 2015. Fundamentals of experimental design: guidelines for designing successful experiments. *Agronomy Journal* 107 (2): 692–705.

Castellani, L.G., I. Brun, and M. Butterfield. 2018. Genomic breeding value estimation of sugarcane genotypes based on high-density SNP. Abstract in *Plant and Animal Genome XXVIII Conference* W954.

Chang, Y.S., and S.B. Milligan. 1992. Estimating the potential of sugarcane families to produce elite genotypes using univariate cross prediction methods. *Theoretical and Applied Genetics* 84 (5–6): 662–671.

Chao, C.P., J.W. Hoy, A.M. Saxton, and F.A. Martin. 1990. Heritability of resistance and repeatability of clone reactions to sugarcane smut in Louisiana. *Phytopathology* 80 (7): 622–626.

Cochran, W.G., and G.M. Cox. 1957. *Experimental designs*, 2nd ed. New-York: John Wiley and Sons Inc.

Costa, E.A., C.O. Anoni, M.C. Mancini, F.R.C. Santos, T.G. Marconi, R. Gazaffi, M.M. Pastina, D. Perecin, M. Mollinari, M.A. Xavier, L.R. Pinto, A.P. Souza, and A.A.F. Garcia. 2016. QTL mapping including codominant SNP markers with ploidy level information in a sugarcane progeny. *Euphytica* 211 (1): 1–16.

Costet, L., L.M. Raboin, M. Payet, A. D'Hont, and S. Nibouche. 2012a. A major quantitative trait allele for resistance to the Sugarcane yellow leaf virus (Luteoviridae). *Plant Breeding* 131 (5): 637–640.

Costet, L., L. Le Cunff, S. Royaert, L.M. Raboin, C. Hervouet, L. Toubi, H. Telismart, O. Garsmeur, Y. Rousselle, J. Pauquet, S. Nibouche, J.C. Glaszmann, J.Y. Hoarau, and A. D'Hont. 2012b. Haplotype structure around *Bru1* reveals a narrow genetic basis for brown rust resistance in modern sugarcane cultivars. *Theoretical and Applied Genetics* 125 (5): 825–836.

Cox, M., M. Hogarth, and G. Smith. 2000. Cane breeding and improvement. In *Manual of cane growing*, ed. Hogarth and Allsopp, 91-108. Bureau of Sugar Experiment Stations.

Crossa, J., P. Pérez, G. de los Campos, G. Mahuku, S. Dreisigacker, and C. Magorokosho. 2011. Genomic selection and prediction in plant breeding. *Journal of Crop Improvement* 25 (3): 239–261.

Cullis, B., A. Smith, and N. Coombes. 2006. On the design of early generation variety trials. *Journal of Agricultural, Biological and Environmental Statistics* 11 (4): 382–393.

Cursi, D.E., H.P. Hoffmann, G.V.S. Barbosa, J.A. Bressiani, R. Gazaffi, R.G. Chapola, A.R. Fernandes Junior, T.W.A. Balsalobre, C.A. Diniz, J.M. Santos, and M.S. Carneiro. 2021. History and current status of sugarcane breeding, germplasm development and molecular genetics in Brazil. *Sugar Tech* (in press).

D'Hont, A., L. Grivet, P. Feldmann, J.C. Glaszmann, S. Rao, and N. Berding. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Molecular and General Genetics* 250 (4): 405–413.

D'Hont, A., D. Ison, K. Alix, C. Roux, and J.C. Glaszmann. 1998. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. *Genome* 41 (2): 221–225.

Daniels, J. 1972. Description of sugarcane clones 1. Agricultural description. *Proceedings of the International Society of Sugar Cane Technologists* 14: 112–119.

Daugrois, J.H., L. Grivet, D. Roques, J.Y. Hoarau, H. Lombard, J.C. Glaszmann, and A. D'Hont. 1996. A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570.' *Theoretical and Applied Genetics* 92 (8): 1059–1064.

Débibakas, S., S. Rocher, O. Garsmeur, L. Toubi, D. Roques, A. D'Hont, J.Y. Hoarau, and J.H. Daugrois. 2014. Prospecting sugarcane resistance to sugarcane yellow leaf virus by genome-wide association. *Theoretical and Applied Genetics* 127 (8): 1719–1732.

Deomano, E., P. Jackson, X. Wei, K. Aitken, R. Kota, and P. Perez-Rodriguez. 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. *Molecular Breeding* 40: 38.

Dumont, T., A. Thong-Chane, L. Barau, B. Siegmund, and J.Y. Hoarau. 2019. Genetic variabilities and genetic gains for yield components in regional sugarcane breeding programmes on Réunion Island. *Sugar Tech* 21 (6): 868–878.

Dumont, T., L. Barau, A. Thong-Chane, J. Dijoux, M. Mellin, J. Daugrois and J.Y. Hoarau. 2021. Sugarcane breeding in Reunion: challenges, achievements and future prospects. *Sugar Tech* (in press).

Edme, S.J., P.Y.P. Tai, and J.D. Miller. 2007. Relative efficiency of spatial analyses for non-replicated early-stage sugarcane field experiments. *Journal American Society of Sugar Cane Technologists* 27: 89–104.

Engel, B., and A. Keen. 1994. A simple approach for the analysis of generalized mixed linear models. *Statistica Neerlandica* 48: 1–22.

Falconer, D.S., and T.F.C. Mackay. 1996. *Introduction to Quantitative Genetics*, 4th ed., 464. Essex, UK: Longman.

Fartek, B., S. Nibouche, P. Turpin, L. Costet, and B. Reynaud. 2012. Resistance to *Melanaphis sacchari* in the sugarcane cultivar R 365. *Entomologia Experimentalis Et Applicata* 144 (3): 270–278.

Fickett, N., A. Gutierrez, M. Verma, M. Pontif, A. Hale, C. Kimbeng, and N. Baisakh. 2019. Genome-wide association mapping identifies markers associated with cane yield components and sucrose traits in the Louisiana sugarcane core collection. *Genomics* 111 (6): 1794–1801.

Galwey, N.W. 2014. *Introduction to mixed modelling. Beyond regression and analysis of variance*. Second edition, John Wiley & Sons edition, 486 pages.

Garcia, A.A.F., E.A. Kido, A.N. Meza, H.M.B. Souza, L.R. Pinto, M.M. Pastina, C.S. Leite, J.A.G. da Silva, E.C. Ulian, A. Figueira, and A.P. Souza. 2006. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theoretical and Applied Genetics* 112 (2): 298–314.

Garfield, G. 2002. The challenge of developing statistical reasoning. *Journal of Statistics Education* 10: 3. https://doi.org/10.1080/10691898.2002.11910676.

Garsmeur, O., G. Droc, R. Antonise, J. Grimwood, B. Potier, K. Aitken, J. Jenkins, G. Martin, C. Charron, C. Hervouet, L. Costet, N. Yahiaoui, A. Healey, D. Sims, Y. Cherukuri, A. Sreedasyam, A. Kilian, A. Chan, M.A. Van Sluys, K. Swaminathan, C. Town, H. Bergès, B. Simmons, J.C. Glaszmann, E. Van der Vosser, R. Henry, J. Schmutz, and A. D'Hont. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature Communications* 9 (1): 1–10.

Gilmour, A.R., B.R. Cullis, and A.P. Verbyla. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological, and Environmental Statistics* 2 (3): 269–293.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, and R. Thompson. 2009. *ASReml user guide 3.0*. VSN Int. Hemel Hempstead, UK.

Glaz, B., and M.S. Kang. 2008. Location contributions determined via GGE biplot analysis of multi environment sugarcane genotype-performance trials. *Crop Science* 48: 941–950.

Glynn, N.C., C. Laborde, R.W. Davidson, M.S. Irey, B. Glaz, A. D'Hont, and J.C. Comstock. 2013. Utilization of a major brown rust resistance gene in sugarcane breeding. *Molecular Breeding* 31 (2): 323–331.

Gouy, M., Y. Rousselle, D. Bastianelli, P. Lecomte, L. Bonnal, D. Roques, J.C. Efile, S. Rocher, J.H. Daugrois, L. Toubi, S. Nabeneza, C. Hervouet, H. Telismart, M. Denis, A. Thong-Chane, J.C. Glaszmann, J.Y. Hoarau, S. Nibouche, and L. Costet. 2013. Experimental assessment of the accuracy of genomic selection in sugarcane. *Theoretical and Applied Genetics* 126 (10): 2575–2586.

Gouy, M., Y. Rousselle, A. Thong-Chane, A. Anglade, S. Royaert, S. Nibouche, and L. Costet. 2015. Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. *Euphytica* 202 (2): 269–284.

Gravois, K.A., S.B. Milligan, and F.A. Martin. 1991. Additive genetic effects for sugarcane yield components and implications for hybridization. *Tropical Agriculture (trinidad)* 68 (4): 376–380.

Guilly, S., T. Dumont, A. Thong-Chane, L. Barau, and J.Y. Hoarau. 2017. Analysis of multienvironment trials (MET) in the sugarcane breeding program of Reunion Island. *Euphytica* 213 (9): 213.

Gutierrez, A.F., J.W. Hoy, C.A. Kimbeng, and N. Baisakh. 2018. Identification of genomic regions controlling leaf scald resistance in sugarcane using a bi-parental mapping population and selective genotyping by sequencing. *Frontiers in Plant Science* 9: 877.

Hazel, D.L. 1943. The genetic basis for constructing selection indexes. *Genetics* 28: 476–490.

Henderson, C.R. 1984. *Applications of linear models in animal breeding*. ON, Canada: Univ. of Guelph.

Hoarau, J.Y., L. Grivet, B. Offmann, L.M. Raboin, J.P. Diorflar, J.P. Payet, M. Hellmann, A. D'Hont, and J.C. Glaszmann. 2002. Genetic dissection of a modern sugarcane cultivar (*Saccharum* spp.). II. Detection of QTLs for yield components. *Theoretical and Applied Genetics* 105 (6–7): 1027–1037.

Hoarau, J.Y., G. Souza, A. D'Hont, M. Menossi, L.R. Pinto, A.P. Souza, L. Grivet, M.C.F. Menck, E.C. Cesar, and M. Vincentz. 2007. Sugarcane: A tropical crop with a highly complex genome. In *Functional plant genomics*, ed. J.F. Morot-Gaudry, P. Lea, and J.F. Briat, 481–499. Enfield: Science Publishers.

Hoarau, J.Y., D. Roques, J.C. Efile, N. Adjanoh-Lubin, S. Débibakas, and J.H. Daugrois. 2018. Narrow-sense heritability estimates of resistance of sugarcane to Sugarcane yellow leaf virus. *International Sugar Journal CXX* 1435: 548–553.

Hoarau, J.Y., A. Thong-Chane, L. Barau, and T. Dumont. 2019. Comparison between randomized complete-block and alpha designs in sugarcane variety trials in the light of a spatial modelling of fertility trends. *Proceedings of the International Society of Sugar Cane Technologists* 30: 121–127.

Hogarth, D.M., K.K. Wu, and D.J. Heinz. 1981. Estimating genetic variance in sugarcane using a factorial cross design. *Crop Science* 21 (1): 21–25.

Hogarth, D.M., C.C. Ryan, and P.W.J. Taylor. 1993. Quantitative inheritance of rust resistance in sugarcane. *Field Crops Research* 34 (2): 187–193.

Hogarth, D.M., M.C. Cox, and J.K. Bull. 1997. Sugarcane improvement: past achievements and future prospects. *Crop Improvement for the 21st Century. Research Signpost*, Trivandrum, India 29–56.

Jackson, P.A. 2005. Breeding for improved sugar content in sugarcane. *Field Crops Research* 92: 277–290.

Jackson, P.A. 2018a. What is the optimal selection intensity of seedlings (stage 1) in sugarcane breeding programs? *International Sugar Journal CXX* 1435: 542–546.

Jackson, P.A. 2018b. Advances in conventional sugarcane breeding programmes. In *Achieving sustainable cultivation of sugarcane Volume 2: Breeding, pests and diseases*, ed. Rott, P., Burleigh Dodds Science Publishing, Cambridge, UK, 2018, (ISBN: 978 1 78676 148 4; www.bdspublishing.com).

Jackson, P.A. 2021. An introduction to some concepts in statistical analysis and quantitative genetics for sugarcane breeding programs. *Sugar Tech* (submitted).

Jackson, P.A., and T.A. McRae. 2001. Selection of sugarcane clones in small plots. *Crop Science* 41 (2): 315–322.

Jackson, P.A., S. Chapman, A. Rattey, X. Wei, and M. Cox. 2007. Genotype × region interactions and implications for sugarcane breeding programs. *Proceedings of the Australian Society of Sugarcane Technologists* 29: 9p.

Kang, M.S., J.D. Miller, and P.Y.T. Tai. 1983. Genetic and phenotypic path analyses and heritability in sugarcane. *Crop Science* 23 (4): 643–647.

Kennedy, A.J. 2005. Breeding improved cultivars for the Caribbean by utilization of total biomass production. *Proceedings of the International Society of Sugar Cane Technologists* 25: 491–499.

Kimbeng, C.A., and M.C. Cox. 2003. Early generation selection of sugarcane families and clones in Australia: a review. *Journal of the American Society of Sugar Cane Technologists* 23: 20–39.

Koonjah, S.S., A. Beekharry, M.G.H. Badaloo, C. Henderson, and A.D. Saumtally. 2019. Evaluation of Near Infrared Spectroscopy for the direct analysis of cane quality characters. *Universal Journal of Agricultural Research* 7 (5): 169–176.

Le Cunff, L., O. Garsmeur, L.M. Raboin, J. Pauquet, H. Telismart, A. Selvi, L. Grivet, R. Philippe, D. Begum, M. Deu, L. Costet, R. Wing, J.C. Glaszmann, and A. D'Hont. 2008. Diploid/polyploid syntenic shuttle mapping and haplotype-specific chromosome walking toward a rust resistance gene (*Bru1*) in highly polyploid sugarcane ($2n \sim 12x \sim 115$). *Genetics* 180 (1): 649–660.

Lee, D.J., M. Durbán, and P. Eilers. 2013. Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. *Computational Statistics & Data Analysis* 61: 22–37.

Leite, M.S.D.O., L.A. Peternelli, M.H.P. Barbosa, P.R. Cecon, and C.D. Cruz. 2009. Sample size for full-sib family evaluation in sugarcane. *Pesquisa Agropecuária Brasileira* 44 (12): 1562–1574.

Leite M.S.D.O., L.A. Peternelli, M.H.P. Barbosa. 2006. Effects of plot size on the estimation of genetic parameters in sugarcane families. *Crop Breeding and Applied Biotechnology* 6: 40–46.

Li, W.F., X.Y. Wang, Y.K. Huang, R.Y. Zhang, H.L. Shan, J. Yin, and Z.M. Luo. 2017. Molecular detection of *Bru1* gene and identification of brown rust resistance in Chinese sugarcane germplasm. *Sugar Tech* 19 (2): 183–190.

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2006. *SAS for mixed models*, 2nd ed. Cary: SAS; Institute Inc.

Liu, G.F., H.K. Zhou, H. Hu, Z.H. Zhu, Y. Hayat, M.H. Xu, and Y. Yang. 2007. Genetic analysis for brix weight per stool and its component traits in sugarcane (*Saccharum officinarum*). *Journal of Zhejiang University Science B* 8 (12): 860–866.

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J.L. Jannink. 2011. Genomic selection in plant breeding: knowledge and prospects. *Advances in Agronomy. Academic Press* 110: 77–123.

Luo, J., Y.B. Pan, L. Xu, M.P. Grisham, H. Zhang, and Y. Que. 2015. Rational regional distribution of sugarcane cultivars in China. *Scientific Reports* 5 (1): 1–10.

Lush, J. 1937. *Animal breeding Plans*. Ames: Iowa State College Press.

Margarido, G.R.A., A.P. Souza, and A.A.F. Garcia. 2007. OneMap: software for genetic mapping in outcrossing species. *Hereditas* 144 (3): 78–79.

Margarido, G.R.A., M.M. Pastina, A.P. Souza, and A.A.F. Garcia. 2015. Multi-trait multi-environment quantitative trait loci mapping for a sugarcane commercial cross provides insights on the inheritance of important traits. *Molecular Breeding* 35 (8): 1–15.

Matassa, V.J., K.E. Basford, J.K. Stringer, and D.M. Hogarth. 1998. The application of spatial analysis to sugarcane variety trials. *Proceedings of the Australian Society of Sugarcane Technologists* 20: 162–168.

McCord, P., N. Glynn, and J. Comstock. 2019. Identifying markers for resistance to sugarcane orange rust (*Puccinia kuehnii*) via selective genotyping and capture sequencing. *Euphytica* 215 (9): 150.

McIntyre, C.L., V.A. Whan, B. Croft, R. Magarey, and G.R. Smith. 2005. Identification and validation of molecular markers associated with Pachymetra root rot and brown rust resistance in sugarcane using map-and association-based approaches. *Molecular Breeding* 16 (2): 151–161.

Mendes de Paula, T.O., B.P. Brasileiro, D.E. Cursi, E. Gonçalves de Freitas, J.M. dos Santos, M.D.V. de Resende, C. Kimbeng, and M.H.P. Barbosa. 2020. Establishment of gene pools for systematic heterosis exploitation in sugarcane breeding. *Agronomy Journal* 112: 3847–3858.

Milligan, S.B. 1994. Test site allocation within and among stages of a sugarcane breeding program. *Crop Science* 34 (5): 1184–1190.

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

Ming, R., Y. Wang, X. Draye, P. Moore, J. Irvine, and A.H. Paterson. 2002. Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane. *Theoretical and Applied Genetics* 105 (2–3): 332–345.

Neto, H.Z., E. Daros, J.C. Bespalhok Filho, C.A. Scapim, M.C.G. Vidigal, and P.S. Vidigal Filho. 2013. Selection of families and parents of sugarcane (*Saccharum* spp.) through mixed models by joint analysis of two harvests. *Euphytica* 193 (3): 391–408.

Nibouche, S., L.M. Raboin, J.Y. Hoarau, A. D'Hont, and L. Costet. 2012. Quantitative trait loci for sugarcane resistance to the spotted stem borer *Chilo sacchariphagus*. *Molecular Breeding* 29 (1): 129–135.

Ostengo, S., M.I. Cuenya, and M. Balzarini. 2015. Modeling spatial correlation structure in sugarcane (*Saccharum* spp.) multi-environment trials. *Journal of Crop Improvement* 29: 53–64.

Parco, A.S., A.L. Hale, M.C. Avellaneda, J.W. Hoy, C.A. Kimbeng, M.J. Pontif, P.H. McCord, T. Ayala-Silva, J.R. Todd, and N. Baisakh. 2017. Distribution and frequency of *Bru1*, a major brown rust resistance gene, in the sugarcane world collection. *Plant Breeding* 136 (5): 637–651.

Pastina, M.M., M. Malosetti, R. Gazaffi, M. Mollinari, G.R.A. Margarido, K.M. Oliveira, L.R. Pinto, A.P. Souza, F.A. Van Eeuwijk, and A.A.F. Garcia. 2012. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. *Theoretical and Applied Genetics* 124 (5): 835–849.

Payne, R.W., D.A. Murray, S.A. Harding, D.B. Baird, and D.M. Soutar. 2009. *GenStat for Windows*. 12th ed. VSN Int., Hemel Hempstead, UK.

Pedrozo, C.Â., M.H.P. Barbosa, F.L. da Silva, M.D.V. de Resende, and L.A. Peternelli. 2011. Repeatability of full-sib sugarcane families across harvests and the efficiency of early selection. *Euphytica* 182 (3): 423–430.

Piepho, H.P., J. Möhring, A.E. Melchinger, and A. Büchse. 2008. BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161 (1–2): 209–228.

Pinto, L.R., A.A.F. Garcia, M.M. Pastina, L.H.M. Teixeira, J.A. Bressiani, E.C. Ulian, M.A.P. Bidoia, and A.P. Souza. 2010. Analysis of genomic and functional RFLP derived markers associated with sucrose content, fiber and yield QTLs in a sugarcane (*Saccharum* spp.) commercial cross. *Euphytica* 172 (3): 313–327.

Piperidis, N., and A. D'Hont. 2020. Sugarcane genome architecture decrypted with chromosome specific oligo probes. *The Plant Journal* 103 (6): 2039–2051.

Piperidis, N., P.A. Jackson, A. D'Hont, P. Besse, J.Y. Hoarau, B. Courtois, K.S. Aitken, and C.L. McIntyre. 2008. Comparative genetics in sugarcane enables structured map enhancement and

validation of marker-trait associations. *Molecular Breeding* 21 (2): 233–247.

Pisaroglo de Carvalho, M., S.A. Gezan, L.A. Peternelli, and M.H. Pereira Barbosa. 2014. Estimation of additive and nonadditive genetic components of sugarcane families using multitrait analysis. *Agronomy Journal* 106 (3): 800–808.

R Development Core Team. 2010. A language and environment for statistical computing. R foundation for Statistical Computing, Vienna. www.R-project.org.

Raboin, L.M., J. Pauquet, M. Butterfield, A. D'Hont, and J.C. Glaszmann. 2008. Analysis of genome-wide linkage disequilibrium in the highly polyploid sugarcane. *Theoretical and Applied Genetics* 116 (5): 701–714.

Racedo, J., M.F. Perera, R. Bertani, C. Funes, V. González, M.I. Cuenya, A. D'Hont, B. Welin, and A.P. Castagnaro. 2013. *Bru1* gene and potential alternative sources of resistance to sugarcane brown rust disease. *Euphytica* 191 (3): 429–436.

Racedo, J., L. Gutiérrez, M.P. Perera, S. Ostengo, E.M. Pardo, M.I. Cuenya, and A.P. Castagnaro. 2016. Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biology* 16 (1): 142.

Ramburan, S., and M. Zhou. 2011. Investigating sugarcane genotype × environment interactions under rainfed conditions in South Africa using variance components and biplot analysis. *Proceedings of the South African Sugarcane Technologists* 84: 245–362.

Ramburan, S., M. Zhou, and M. Labuschagne. 2012. Investigating test site similarity, trait relations and causes of genotype × environment interactions of sugarcane in the Midlands region of South Africa. *Field Crops Research* 129: 71–80.

Ramdoyal, K., and G. Badaloo. 1998. Inheritance of agronomic traits in commercial hybrid sugar cane populations in contrasting environments and in different crop cycles. *Journal of Genetics & Breeding* 52: 361–368.

Ramdoyal, K., S. Sullivan, L.C.Y. Chong, G.H. Badaloo, S. Saumtally, and R. Domaingue. 2000. The genetics of rust resistance in sugar cane seedling populations. *Theoretical and Applied Genetics* 100 (3-4): 557–563.

Reffay, N., P.A. Jackson, K.S. Aitken, J.Y. Hoarau, A. D'Hont, P. Besse, and C.L. McIntyre. 2005. Characterisation of genome regions incorporated from an important wild relative into Australian sugarcane. *Molecular Breeding* 15 (4): 367–381.

Rodríguez-Álvarez, M.X., M.P. Boer, P.H.C. Eilers, and F.A. Van Eeuwijk. 2016. SpATS: spatial analysis of field trials with splines. R package version 1.0–4. https://cran-r.project.org/package=SpATS.

Rodríguez-Álvarez, M.X., M.P. Boer, F.A. Van Eeuwijk, and P.H.C. Eilers. 2018. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics* 23: 52–71.

Roussel, C., F.R. Warembourg, F. Davrieux, J. Bernard, and C. Poser. 2015. Development of NIRS method for routine assessment of sugarcane quality in Reunion Island. In ISSCT abstract of 11th germplasm and breeding and 8th molecular biology workshops. Pushing the frontiers of sugarcane improvement. ISSCT, St-Gilles, Reunion Island 1–5 June 2015 [BO12].

Sabatier, D., L. Thuriès, D. Bastianelli, and P. Dardenne. 2012. Rapid prediction of the lignocellulosic compounds of sugarcane biomass by near infrared reflectance spectroscopy: comparing classical and independent cross-validation. *Journal of near Infrared Spectroscopy* 20 (3): 371–385.

Santchurn, D., M.G.H. Badaloo, S. Koonjah, and A. Dookun-Saumtally. Sugarcane breeding and supporting genetics research in Mauritius. *Sugar Tech* (in press).

Santos, F.R., L.R. Pinto, L.A. Carlini-Garcia, R. Gazaffi, M.C. Mancini, B.S. Gonçalves, C.N.F. Medeiros, D. Perecin, A.A.F. Garcia, A.P. Souza, and M.I. Zucchi. 2015. Marker-trait association and epistasis for brown rust resistance in sugarcane. *Euphytica* 203 (3): 533–547.

Schneeberger, M., S.A. Barwick, G.H. Crow, and K. Hammond. 1992. Economic indices using breeding values predicted by BLUP. *Journal of Animal Breeding and Genetics* 109: 180–187.

Scortecci, K.C., S. Creste, T. Calsa, M.A. Xavier, M.G. Landell, A. Figueira, and V.A. Benedito. 2012. Challenges, opportunities and recent advances in sugarcane breeding. In *Plant breeding*, ed. I.Y. Abdurakhmonov, 267–296. Rijeka, Croatia: InTech Publisher.

Shanthi, R.M., K.V. Bhagyalakshmi, G. Hemaprabha, S. Alarmelu, and R. Nagarajan. 2008. Relative performance of the sugarcane families in early selection stages. *Sugar Tech* 10 (2): 114–118.

Singh, R.K., S.P. Singh, D.K. Tiwari, S. Srivastava, S.B. Singh, M.M. Sharma, R. Singh, T. Mohapatra, and N.K. Singh. 2013. Genetic mapping and QTL analysis for sugar yield-related traits in sugarcane. *Euphytica* 191 (3): 333–353.

Singh, R.K., N. Banerjee, M.S. Khan, S. Yadav, S. Kumar, S.K. Duttamajumder, R.J. Lai, J.D. Patel, H. Guou, D. Zhan, and A.H. Paterson. 2016. Identification of putative candidate genes for red rot resistance in sugarcane (*Saccharum* species hybrid) using LD-based association mapping. *Molecular Genetics and Genomics* 291 (3): 1363–1377.

Skinner, J.C., D.M. Hogarth, and K.K. Wu. 1987. Selection methods, criteria and indices. In *Sugarcane improvement through breeding*, ed. D.J. Heinz, 409–454. Amsterdam: Elsevier.

Smith, H.F. 1936. A discriminant function for plant selection. *Annals Eugenics* 7 (3): 240–250.

Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *The Journal of Agricultural Science* 143 (6): 449–462.

Smith, A.B., J.K. Stringer, X. Wei, and B.R. Cullis. 2007. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. *Euphytica* 157 (1–2): 253–266.

Steel, R.G.D., and J.H. Torrie. 1980. *Principles and Procedures of Statistics. A biometrical approach*. Second edition, 633. New York: McGraw-Hill Book Co., Inc.

Stringer, J.K., and B.R. Cullis. 2002. Application of spatial analysis techniques to adjust for fertility trends and identify interplot competition in early stage sugarcane selection trials. *Australian Journal of Agricultural Research* 53 (8): 911–918.

Stringer, J.K., M.C. Cox, F.C. Atkin, X. Wei, and D.M. Hogarth. 2011. Family selection improves the efficiency and effectiveness of selecting original seedlings and parents. *Sugar Tech* 13 (1): 36–41.

Stringer, J., B. Croft, S. Bhuiyan, E. Deomano, R. Magarey, M. Cox, and X. Xu. 2012. A new method of statistical analysis for sugarcane disease screening trials. *Proceedings of the Australian Society of Sugarcane Technologists* 34: 10p.

Stringer, J., B. Croft, E. Deomano, and S. Bhuiyan. 2013. Analysis of sugarcane disease screening trials over years with a mixed model to improve ratings of varieties. *Proceedings of the International Society of Sugar Cane Technologists* 28, 9 p.

Taira, E., M. Ueno, and Y. Kawamitsu. 2010. Automated quality evaluation system for net and gross sugarcane samples using near infrared spectroscopy. *Journal of near Infrared Spectroscopy* 18 (3): 209–215.

Ukoskit, K., G. Posudsavang, N. Pongsiripat, P. Chatwachirawong, P. Klomsa-ard, P. Poomipant, and S. Tragoonrung. 2019. Detection and validation of EST-SSR markers associated with sugar-related traits in sugarcane using linkage and association mapping. *Genomics* 111 (1): 1–9.

Velazco, J.G., M.X. Rodríguez-Álvarez, M.P. Boer, D.R. Jordan, P.H.C. Eilers, M. Malosetti and F.A. van Eeuwijk. 2017.

Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. *Theoretical and Applied Genetics* 130: 1375–1392.

Wei, X., and P.A. Jackson. 2016. Addressing slow rates of long-term genetic gain in sugarcane. *Proceedings of the International Society of Sugar Cane Technologists* 29: 1923–1930.

Wei, X., P.A. Jackson, L. McIntyre, K.S. Aitken, and B. Croft. 2006a. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theoretical and Applied Genetics* 114 (1): 155–164.

Wei, X., P. Jackson, M. Cox, and J. Stringer. 2006b. Maximising economic benefits to the whole sugarcane industry from the BSES-CSIRO sugarcane improvement program. *Proceedings of the Australian Society of Sugarcane Technologists* 28: 181–186.

Wei, X., P.A. Jackson, J. Stringer, and M. Cox. 2008. Relative economic genetic value (rEGV)-an improved selection index to replace net merit grade (NMG) in the Australian sugarcane variety improvement program. *Proceedings of the Australian Society of Sugarcane Technologists* 30: 174–181.

Wei, X., P.A. Jackson, S. Hermann, A. Kilian, K. Heller-Uszynska, and E. Deomano. 2010. Simultaneously accounting for population structure, genotype by environment interaction, and spatial variation in marker–trait associations in sugarcane. *Genome* 53 (11): 973–981.

Wei, X., P.A. Jackson, F. Atkin, M. Cox, G. Piperidis, and R. Parfitt. 2013. How well are the predicted breeding values of parental clones associated with family performance in the BSES-CSIRO sugarcane breeding program? *International Sugar Journal* 115 (1371): 208–211.

Wei, X., J. Stringer, B. Salter, G. Piperidis, and B. Schroeder. 2015. Improving selection accuracy in clonal assessment trials by accounting for site variability. *Proceedings of the Australian Society of Sugarcane Technologists* 37: 237–243.

White, T.L., and G.R. Hodge. 1989. *Predicting breeding values with applications in forest tree improvement*. Dordrecht, Boston, London: Kluwer academic publishers.

White, W.H., J.D. Miller, S.B. Milligan, D.M. Burner, and B.L. Legendre. 2001. Inheritance of sugarcane borer resistance derived from two measures of insect damage. *Crop Science* 41: 1706–1710.

Wu, K.K., D.J. Heinz, H.K. Meyer, and S.L. Ladd. 1978. Minimum sample size for estimating progeny mean and variance. *Crop Science* 18 (1): 57–62.

Wu, K.K., D.J. Heinz, and D.M. Hogarth. 1988. Association and heritability of sugarcane smut resistance to races A and B in Hawaii. *Theoretical and Applied Genetics* 75 (5): 754–760.

Yadav, S., P.A. Jackson, X. Wei, E.M. Ross, K. Aitken, E. Deomano, F. Aitkin, B.J. Hayes, and K.P. Voss-Fels. 2020. Accelerating genetic gain in sugarcane breeding using genomic selection. *Agronomy* 10 (4): 585.

Yadav, S., X. Wei, P. Joyce, F. Atkin, E. Deomano, Y. Sun, L.T. Nguyen, E.M. Ross, T. Cavallaro, K.S. Aitken, B.J. Hayes, and K.P. Voss-Fels. 2021. Improved genomic prediction of clonal performance in sugarcane by exploiting non-additive genetic effects. *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-021-03822-1.

Yan, W. 2002. Singular-value partitioning in biplot analysis of multi-environment trial data. *Agronomy Journal* 94: 990–996.

Yan, W., and N.A. Tinker. 2005. An integrated biplot analysis system for displaying, interpreting, and exploring genotype × environment interaction. *Crop Science* 45: 1004–1106.

Yan, W., and N.A. Tinker. 2006. Biplot analysis of multi-environment trial data: principles and applications. *Canadian Journal of Plant Science* 86 (3): 623–645.

Yang, X., M.S. Islam, S. Sood, S. Maya, E.A. Hanson, J. Comstock, and J. Wang. 2018. Identifying Quantitative Trait Loci (QTLs) and developing diagnostic markers linked to orange rust resistance in sugarcane (Saccharum spp.). *Frontiers in Plant Science* 9: 350.

Yang, X., J. Todd, R. Arundale, J.B. Binder, Z. Luo, M.S. Islam, S. Sood, and J. Wang. 2019a. Identifying loci controlling fiber composition in polyploid sugarcane (*Saccharum* spp.) through genome-wide association study. *Industrial Crops and Products* 130: 598–605.

Yang, X., Z. Luo, J. Todd, S. Sood, and J. Wang. 2019b. Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (*Saccharum* spp.). *The Plant Genome*. https://doi.org/10.1002/tpg2.20006.

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38 (2): 203–208.

Zan, F., Y. Zhang, Z. Wu, J. Zhao, C. Wu, Y. Zhao, X. Chen, L. Zhao, W. Qin, L. Yao, H. Xia, P. Zhao, K. Yang, J. Liu, and X. Yang. 2020. Genetic analysis of agronomic traits in elite sugarcane (*Saccharum* spp.) germplasm. *PLoS ONE* 15 (6): e0233752.

Zhang, J., X. Zhang, H. Tang, Q. Zhang, X. Hua, X. Ma, , ... and R. Ming. 2018 Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics* 50(11): 1565 1573.

Zhou, M.M. 2013. Conventional sugarcane breeding in South Africa: progress and future prospects. *American Journal of Plant Sciences* 4: 189–196.

Zhou, M.M. 2018. Simultaneous selection for multiple traits in sugarcane breeding populations using logistic regression models. *South African Journal of Plant and Soil* 35 (3): 165–171.

Zhou, M.M. 2019. Potential of exploiting additive and non-additive genetic effects in sugarcane breeding. *Proceedings of the International Society of Sugar Cane Technologists* 30: 482–488.