MATEUS TELES VITAL GONÇALVES

PREDICTIVE METHODS USING PEDIGREES, MARKERS, AND IMAGES FOR THE GENETIC IMPROVEMENT OF SUGARCANE

Thesis submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

Adviser: Luiz Alexandre Peternelli
Co-advisers: Gota Morota, Márcio Henrique Pereira Barbosa.

VIÇOSA - MINAS GERAIS
2024

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

Gonçalves, Mateus Teles Vital, 1993-
G635p Predictive methods using pedigrees, markers, and images
2024 for the genetic improvement of sugarcane / Mateus Teles Vital
Gonçalves. – Viçosa, MG, 2024.
1 tese eletrônica (89 f.): il. (algumas color.).

Texto em inglês.
Orientador: Luiz Alexandre Peternelli.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Estatística, 2024.
Inclui bibliografia.
DOI: https://doi.org/10.47328/ufvbbt.2024.652
Modo de acesso: World Wide Web.

1. Cana-de-açúcar - Melhoramento genético. 2. BLUP. 3. Poliploidia. 4. *Saccharum*. 5. Genômica. 6. Cana-de-açúcar - Seleção. 7. Modelos lineares (Estatística). I. Peternelli, Luiz Alexandre, 1966-. II. Universidade Federal de Viçosa. Departamento de Estatística. Programa de Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 633.612

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

MATEUS TELES VITAL GONÇALVES

PREDICTIVE METHODS USING PEDIGREES, MARKERS, AND IMAGES FOR THE GENETIC IMPROVEMENT OF SUGARCANE

Thesis submitted to the Genetics and Breeding Graduate Program of the Universidade Federal de Viçosa in partial fulfillment of the requirements for the degree of *Doctor Scientiae*.

APPROVED: September 23, 2024.

Assent:

Documento assinado digitalmente
MATEUS TELES VITAL GONCALVES
Data: 09/10/2024 15:30:49-0300
Verifique em https://validar.iti.gov.br

Mateus Teles Vital Gonçalves
Author

Documento assinado digitalmente
LUIZ ALEXANDRE PETERNELLI
Data: 09/10/2024 14:39:45-0300
Verifique em https://validar.iti.gov.br

Luiz Alexandre Peternelli
Adviser

Dedico esta tese à Divina Amâncio Teles (in memoriam)

ACKNOWLEDGEMENTS

To my family, especially my father Gilberto, my mother Wilma, and my brother Gustavo. Thank you for your love, support, and for making everything possible for me to become the person I am today.

To all people from the Universidade Federal de Viçosa, and the Genetics and Breeding Graduate Program for providing the infrastructure and all the means for my training and education.

To my advisor Luiz Alexandre Peternelli, for the direction throughout the years since I was an enthusiastic undergraduate student. Thank you for your encouragement, for sharing your expertise, and for offering so many opportunities. Also, I am thankful for the horizontal and respectful relationship you had towards me.

To my co-advisor Gota Morota, for the guidance and for receiving me as a visiting graduate student in his lab at Virginia Tech. Thank you for sharing your expertise, the fruitfully meetings, and insightful discussions.

To my co-advisor Márcio Henrique Pereira Barbosa, for the guidance and his prompt assistance whenever I had to work at the research station of the Sugarcane Genetic Breeding Program of the Universidade federal de Viçosa (PMGCA). I would also like to thank all people from the PMGCA for their assistance in planting and conducting field trials.

To my committee members, Guilherme Silva Pereira, Lucas de Paula Corrêdo, Rafael Massahiro Yassue, Matheus Ferreira, and Hélcio Duarte Pereira, thank you for your mentorship and valuable suggestions.

To my girlfriend Thaliã, for her love, patience, and support.

To all friends and colleagues I had the pleasure to spend time with during the preparation of this thesis, at the Universidade Federal de Viçosa, and at the Virginia Tech Animal Science department.

Education is a privilege and I am extremely grateful for being able to fully dedicate my time to research during my doctoral training. I would not be able to do so, without the financial support of the Brazilian government funding institutions. Therefore, I would like to acknowledge the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Brasil (CAPES) – Finance Code 001.

ABSTRACT

GONÇALVES, Mateus Teles Vital, D.Sc., Universidade Federal de Viçosa, September, 2024. **Predictive methods using pedigrees, markers, and images for the genetic improvement of sugarcane**. Adviser: Luiz Alexandre Peternelli. Co-advisers: Gota Morota and Márcio Henrique Pereira Barbosa.

The overall objective of this thesis was to assess predictive methods that leverage different sources of information, to improve genetic evaluation of sugarcane genotypes. The plant material and experimental data used consisted of early-generation field trials conducted by the genetic breeding program of the Universidade Federal de Viçosa (PMGCA). In the first chapter I discussed the content, introducing research questions and the main objectives of this thesis. In the second chapter, we compared pedigree-based best linear unbiased prediction (PBLUP), genomic-based (GBLUP), and single-step (ssGBLUP) models for the genetic evaluation of a sugarcane population in which only a subset of individuals was genotyped. Models were evaluated in two cross-validation (CV) schemes: validation using genotyped individuals (CV1) and validation using nongenotyped individuals (CV2). Our results suggest that genetic evaluation using the ssGBLUP models may be an alternative approach for sugarcane. Also, results showed that models including only pedigree information gave relatively high prediction accuracies, suggesting that pedigrees are an important source of genetic information, particularly for sugarcane and other crop species with complex polyploid genomes. In the third chapter, we evaluated the integration of unoccupied aerial systems (UAS)-based red-green-blue (RGB) imaging with pedigree and genomic prediction models to improve selection accuracy for tonnes of cane per hectare (TCH). The objectives were to estimate genetic parameters and trends for TCH and RGB-image extracted traits, and to compare the performance of single-trait with multi-trait genomic and pedigree prediction models that incorporate RGB-image extracted traits. The performance of models was evaluated in terms of Pearson's correlation between adjusted and predicted phenotypes, and mean squared error (MSE) using three cross-validation schemes, which varied in the level of phenotypic information available: ST, without secondary traits; MT-

1, secondary traits in the training set; and MT-2, secondary traits in both, training and testing sets. We used data of an augmented block design trial, consisting of 385 clones. Clones were phenotyped at the second ratoon stage for TCH, and for 12 RGB-image extracted traits collected in a single flight. In general, we found low genetic correlation between TCH and RGB-image extracted traits, and moderate narrow-sense heritability estimates for RGB-image extracted traits. Overall indirect response to selection of RGB-image extracted traits was higher compared to direct response to selection for TCH. Our results suggest that accuracies of multi-trait models that incorporated RGB-image extracted traits did not improve compared to single-trait models for predicting TCH. Future research should investigate alternative sensor technologies and optimize UAS-based data collection.

Keywords: RGB; Single-step models; Multi-trait models; Polyploid; *Saccharum spp.*; Genomic prediction.

RESUMO

GONÇALVES, Mateus Teles Vital, D.Sc., Universidade Federal de Viçosa, setembro de 2024. **Métodos preditivos utilizando pedigrees, marcadores e imagens para o melhoramento genético da cana-de-açúcar**. Orientador: Luiz Alexandre Peternelli. Coorientadores: Gota Morota e Márcio Henrique Pereira Barbosa.

O objetivo geral desta tese foi avaliar métodos preditivos que aproveitam diferentes fontes de informação, para melhorar a avaliação genética de genótipos de cana-de-açúcar. O material vegetal e os dados experimentais utilizados consistiram de ensaios de campo de fases iniciais de seleção conduzidos pelo programa de melhoramento genético da Universidade Federal de Viçosa (PMGCA). No primeiro capítulo discuti o conteúdo, introduzindo as questões de pesquisa e os principais objetivos desta tese. No segundo capítulo, comparamos os modelos de melhor predição linear não enviesada baseada em pedigree (PBLUP), baseada em genômica (GBLUP) e de etapa única (ssGBLUP) para a avaliação genética de uma população de cana-de-açúcar na qual apenas um subconjunto de indivíduos foi genotipado. Os modelos foram avaliados em dois esquemas de validação cruzada (CV): validação usando indivíduos genotipados (CV1) e validação usando indivíduos não genotipados (CV2). Nossos resultados sugerem que a avaliação genética usando os modelos ssGBLUP pode ser uma abordagem alternativa para a cana-de-açúcar. Além disso, os resultados mostraram que os modelos que incluíam apenas informações de pedigree resultaram em accuracias preditivas relativamente altas, sugerindo que pedigrees são uma fonte importante de informação genética, particularmente para cana-de-açúcar e outras espécies vegetais cultivadas que possuem genomas poliploides complexos. No terceiro capítulo, avaliamos a integração de imagens vermelho-verde-azul (RGB) baseadas em sistemas aéreos não tripulados (UAS) com modelos de predição PBLUP e GBLUP para melhorar a acurácia de seleção para toneladas de cana por hectare (TCH). Os objetivos foram estimar parâmetros genéticos e tendências para TCH e características extraídas de imagens RGB, e comparar o desempenho de modelos de predição PBLUP e GBLUP univariados e multivariados que incorporam características extraídas de imagens RGB. O desempenho dos modelos foi avaliado em termos de correlação de Pearson entre fenótipos ajustados e preditos e o erro quadrático médio (MSE) usando três esquemas de validação cruzada, que variaram no nível de informação fenotípica disponível: ST, sem características secundárias; MT-1, características secundárias no conjunto de treinamento; e MT-2, características secundárias em presentes em ambos os conjuntos, treinamento e teste. Usamos dados de um ensaio de delineamento de blocos aumentados, consistindo de 385 clones. Os clones foram fenotipados na segunda soqueira para TCH e para 12 características extraídas de imagens RGB coletadas em um único voo. Em geral, encontramos baixa correlação genética entre TCH e características extraídas de imagens RGB e estimativas moderadas de herdabilidade no sentido restrito para características extraídas de imagens RGB. Em geral, a resposta indireta a seleção para características extraídas de imagens RGB foi maior em comparação à resposta direta à seleção para TCH. Nossos resultados sugerem que as acurácias preditivas de modelos multivariados que incorporaram características extraídas de imagens RGB não melhoraram em comparação com modelos de univariados para prever TCH. Pesquisas futuras devem investigar alternativas tecnologias de sensores e otimizar a coleta de dados baseada em UAS.

Palavras-chave: RGB; Modelos single-step; Modelos multicaracterística; Poliplóide; *Saccharum spp.*; Predição genômica.

SUMMARY

1. CHAPTER 1

1.1. General Introduction

Sugarcane is used as feedstock for sugar and renewable energy production, which can contribute to reduce greenhouse gas emissions and mitigate the impacts of climate change (Moreira & Pacca, 2020; Pereira et al., 2019). The aim of sugarcane breeding programs is to develop new improved varieties that meet the industry and society needs. However, breeding improved sugarcane varieties is a slow-moving process. Broadly speaking, there are four main stages of testing and selection. The first and second stages consists of large number of selection candidates tested in unreplicated field trials. Then, third and forth stages consist of clonal assessment trials, where replicated field trials are conducted across different environments (Cursi et al., 2022).

The intrinsic characteristics of sugarcane, including a complex polyploid genome and biology, such as relatively long crop (growth) and breeding cycles hinder its genetic improvement (Deomano et al., 2020; Lu et al., 2024). Therefore, current genetic evaluation methods need optimization to increase the rates of genetic gain. Predictive methods involve assessing the performance of selection candidates without phenotyping the individuals themselves (Bernardo, 2021). Predictive methods that incorporate additional information sources may offer new opportunities, potentially helping breeders to make better selection decisions (Cursi et al., 2022).

Oftentimes, data of comparative field experiments are challenging to analyze because of high imbalance and/or lack of replications. For this reason, sugarcane breeders have relied on linear mixed models using the best linear unbiased prediction (BLUP) methodology to analyze experimental data (Henderson et al., 1984). In sugarcane, primarily, random genetic effects are considered independent and their corresponding variance-covariance structure in mixed model equations becomes an identity matrix ($\mathbf{G}\sigma_g^2 = \mathbf{I}\sigma_g^2$) (Hoarau et al., 2022). Also, in the genetic evaluation using BLUP, pedigree information can be used to derive relationship matrices that describe the genetic covariance among individuals in a given population (Barbosa et al., 2005).

In genomic prediction, marker information is likewise used to compute such relationship matrices. Presumably, marker-based relationships are expected to be more accurate compared to pedigree counterparts (B. J. Hayes et al., 2009). Indeed, pedigree records might be incomplete or not available whatsoever; pedigree-based relationships ignore relatedness between founder individuals; are more likely to contain errors; and hence over- or underestimate relationships and genetic parameters (Endelman et al., 2018; Forni et al., 2011; Munoz et al., 2014). In contrast, markers offer a way of overcoming the lack of pedigree records and can track hidden relationships due to historical coancestry (B. J. Hayes et al., 2009; VanRaden, 2008). Furthermore, the use of marker information enables the efficient detection of Mendelian sampling deviations, thereby capitalizing on within-family variation (Goddard & Hayes, 2007). The benefits of genomic selection are greater at early selection stages (Heslot et al., 2015). However, not all selection candidates can be genotyped due to budget constraints. In this context, predictive methods that can efficiently integrate different sources of information, such as the single-step genomic best linear unbiased prediction (ssGBLUP) are highly desirable.

Another important bottleneck in sugarcane breeding is the evaluation of tonnes of cane per hectare (TCH). A key objective in sugarcane breeding is to increase TCH. However, evaluating TCH is a costly, time consuming, and labor-intensive task. Unoccupied aerial systems (UAS) enables breeder to measure secondary traits that are putatively genetically correlated with TCH on large populations, in a faster and more affordable manner as compared to conventional phenotyping tools. Empirical evidence suggests that by integrating high-throughput phenotyping (HTP) data derived from UAS, pedigree, and genotypic data can improve genetic evaluation of selection candidates (Crossa et al., 2017). One strategy is to incorporate secondary traits measured with UAS as response variables into multi-trait models. In this predictive method, models are able to leverage information shared between related individuals and also between correlated traits.

Despite promising, there are many gaps in our knowledge on the application of these predictive methods in sugarcane. Only one study using the ssGBLUP method in sugarcane has been reported (B. J. Hayes et al., 2021). Moreover, few studies investigated the exploitation of pedigrees in sugarcane (Barbosa et al., 2005; Deomano et al., 2020; Inamori et al., 2024). Similarly, little information is known about the effectiveness of multi-trait pedigree- and genomic-based prediction models to estimate TCH across a larger numbers of diverse genotypes at early selection stages. Research done on this topic in sugarcane did not evaluate multi-trait models, nor integrating pedigree and genomic information, in addition, only a limited number of genotypes, established on larger plots were considered (Cholula et al., 2020; Khuimphukhieo et al., 2023).

1.2. Thesis objectives

The general objective of this thesis was to assess predictive methods that leverage different sources of information, to improve genetic evaluation of sugarcane genotypes. In the second chapter, we compared pedigree-based, genomic-based, and single-step (ssGBLUP) models for the genetic evaluation of a sugarcane population in which only a subset of individuals was genotyped. In the third chapter, we evaluated the integration of unoccupied aerial systems (UAS)-based red-green-blue (RGB) imaging with pedigree and genomic information using multi-trait prediction models to improve selection accuracy and efficiency for tonnes of cane per hectare (TCH).

1.3. References

Barbosa, M. H. P., de Resende, M. D. V., Bressiani, J. A., da Silveira, L. C. I., & Peternelli, L. A. (2005). Selection of sugarcane families and parents by reml/blup.

Bernardo, R. (2021). Predictive breeding in maize during the last 90 years. *Crop science*, *61*(5), 2872–2881.

Cholula, U., da Silva, J. A., Marconi, T., Thomasson, J. A., Solorzano, J., & Enciso, J. (2020). Forecasting yield and lignocellulosic composition of energy cane using unmanned aerial systems. *Agronomy*, *10*(5), 718.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., et al. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in plant science*, *22*(11), 961–975.

Cursi, D. E., Hoffmann, H. P., Barbosa, G., Bressiani, J. A., Gazaffi, R., Chapola, R. G., Fernandes Junior, A., Balsalobre, T. W. A., Diniz, C. A., Santos, J. M., et al. (2022). History and current status of sugarcane breeding, germplasm development and molecular genetics in brazil. *Sugar Tech*, *24*(1), 112–133.

Deomano, E., Jackson, P., Wei, X., Aitken, K., Kota, R., & Pérez-Rodríguez, P. (2020). Genomic prediction of sugar content and cane yield in sugar cane clones in different stages of selection in a breeding program, with and without pedigree information. *Molecular Breeding*, *40*, 1–12.

Endelman, J. B., Carley, C. A. S., Bethke, P. C., Coombs, J. J., Clough, M. E., da Silva, W. L., De Jong, W. S., Douches, D. S., Frederick, C. M., Haynes, K. G., et al. (2018). Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. *Genetics*, *209*(1), 77–87.

Forni, S., Aguilar, I., & Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution*, *43*, 1–7.

Goddard, M., & Hayes, B. (2007). Genomic selection. *Journal of Animal breeding and Genetics*, *124*(6), 323–330.

Hayes, B. J., Wei, X., Joyce, P., Atkin, F., Deomano, E., Yue, J., Nguyen, L., Ross, E. M., Cavallaro, T., Aitken, K. S., et al. (2021). Accuracy of genomic prediction of complex traits in sugarcane. *Theoretical and Applied Genetics*, *134*, 1455–1462.

Hayes, B. J., Visscher, P. M., & Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics research*, *91*(1), 47–60.

Henderson, C. R., et al. (1984). *Applications of linear models in animal breeding* (Vol. 462). University of Guelph Guelph.

Heslot, N., Jannink, J.-L., & Sorrells, M. E. (2015). Perspectives for genomic selection applications and research in plants. *Crop Science*, *55*(1), 1–12.

Hoarau, J., Dumont, T., Wei, X., Jackson, P., & D’hont, A. (2022). Applications of quantitative genetics and statistical analyses in sugarcane breeding. sugar tech 24: 320–340.

Inamori, M., Kimura, T., Mori, M., Tarumoto, Y., Hattori, T., Hayano, M., Umeda, M., & Iwata, H. (2024). Machine learning for genomic and pedigree prediction in sugarcane. *The Plant Genome*, e20486.

Khuimphukhieo, I., Marconi, T., Enciso, J., & da Silva, J. A. (2023). The use of uas-based high throughput phenotyping (htp) to assess sugarcane yield. *Journal of Agriculture and Food Research*, *11*, 100501.

Lu, G., Liu, P., Wu, Q., Zhang, S., Zhao, P., Zhang, Y., & Que, Y. (2024). Sugarcane breeding: A fantastic past and promising future driven by technology and methods. *Frontiers in Plant Science*, *15*, 1375934.

Moreira, J. R., & Pacca, S. A. (2020). The climate change mitigation potential of sugarcane based technologies for automobiles; co2 negative emissions in sight. *Transportation Research Part D: Transport and Environment*, *86*, 102454.

Munoz, P. R., Resende Jr, M. F., Huber, D. A., Quesada, T., Resende, M. D., Neale, D. B., Wegrzyn, J. L., Kirst, M., & Peter, G. F. (2014). Genomic relationship matrix for correcting pedigree errors in breeding populations: Impact on genetic parameters and genomic selection accuracy. *Crop Science*, *54*(3), 1115–1123.

Pereira, L. G., Cavalett, O., Bonomi, A., Zhang, Y., Warner, E., & Chum, H. L. (2019). Comparison of biofuel life-cycle ghg emissions assessment tools: The case studies of ethanol produced from sugarcane, corn, and wheat. *Renewable and Sustainable Energy Reviews*, *110*, 1–12.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, *91*(11), 4414–4423.

2. Chapter 2

Single-step genomic best linear unbiased predictions of sugarcane genotypes performance

2.1. Abstract

Genomic prediction has the potential to improve genetic progress in sugarcane. However, genomic information may not be available for all phenotyped individuals. The single-step genomic best linear unbiased prediction (ssGBLUP) method can provide an alternative genetic evaluation approach by combining nongenotyped and genotyped individuals. Here, we investigated ssGBLUP for genetic evaluation of a sugarcane population at an early selection stage. The pedigree contained 4,450 individuals, of which 3,704 were phenotyped and nongenotyped, and 377 were phenotyped and genotyped. We evaluated five models: a pedigree-based (PBLUP-2) and a genomic-based (GBLUP) model considering only genotyped individuals; a pedigree-based model considering all phenotyped individuals (PBLUP-1); and two ssGBLUP models considering all phenotyped individuals, one using the standard genomic relationship matrix (ssGBLUP-1) and one adjusted genomic relationship matrix (ssGBLUP-2). The models were evaluated in two cross-validation (CV) schemes: validation using genotyped individuals (CV1) and validation using non-genotyped individuals (CV2). In CV1, we found that ssGBLUP and PBLUP had overall superior performance compared to GBLUP. The two ssGBLUP models evaluated also gave similar performance. For stalk number and stalk diameter, PBLUP showed higher prediction accuracy than ssGBLUP. For CV2, we found no significant differences in performance between PBLUP and ssGBLUP models. Our results suggest that genetic evaluation using ssGBLUP models may be an alternative approach for sugarcane. Our results also showed that models including only pedigree information gave relatively high prediction accuracies, suggesting that pedigrees are an important potential source of genetic information, particularly for sugarcane and other crop species with complex polyploid genomes.

2.2. Background

Sugarcane (*Saccharum* spp.) is a large perennial grass widely cultivated in tropical and subtropical regions (Cursi et al., 2022; Hale et al., 2022). Sugarcane is grown primarily for sugar and ethanol. The sugarcane milling process produces a considerable amount of bagasse (residual biomass) that can be sent to incinerators to generate electrical energy (Bezerra & Ragauskas, 2016). Alternatively, mills can use bagasse to produce ethanol (Lopes et al., 2016). Sugarcane is considered a bioenergy crop because it is used as a feedstock for biofuel and bioelectricity production (Carvalho-Netto et al., 2014; Matsuoka et al., 2014).

The continuous development of high-performing sugarcane varieties is key for the industry (M. H. P. Barbosa et al., 2012). However, conventional genetic evaluation methods that rely solely on phenotypic selection have not been yielding sufficient genetic gain rates. This is partly due to the relatively long growing and breeding cycles, consisting of extensive field trials, involved in the process of developing new varieties. An alternative and promising approach to genetic evaluation is genomic prediction (Meuwissen et al., 2001).

Genomic prediction integrates high-throughput genotyping technologies with quantitative genetic analysis. In genomic prediction, genome-wide DNA markers are used simultaneously to infer the performance of selection candidates either as parents or per se, i.e., additive or total genetic values. A fundamental assumption is that at least one marker is in linkage disequilibrium with each genomic region affecting a target phenotype (Goddard & Hayes, 2007). The application of genomic prediction in plant breeding programs may allow more accurate predictions of the additive and total genetic value of selection candidates (Fugeray-Scarbel et al., 2021). Moreover, such predictions could be made as early as the seed or seedling stage. Hence, the increase in selection accuracy and reduction in breeding cycle length can ultimately be translated into increased genetic gain (Heffner et al., 2009). Several statistical models have been proposed for genomic prediction (Morota & Gianola, 2014). However, the most widely used model is the genomic best linear unbiased prediction (GBLUP) model and its extensions. The GBLUP model uses marker genotypes, such as single nucleotide polymorphisms (SNP), to compute a genomic relationship matrix (**G**) (VanRaden, 2008). It is equivalent to the animal model used in conventional pedigree-based selection or pedigree-based best linear unbiased prediction (PBLUP) where additive genetic effects (breeding values) are modeled by a covariance structure derived from a numerator relationship matrix (**A**) computed from pedigrees (Henderson, 1976; Piepho et al., 2008).

The GBLUP model assumes that all individuals are genotyped. However, information from non-genotyped relatives, including their phenotypic records and pedigree relatedness, can also be jointly exploited. One way to accomplish this task is to use the single-step genomic best linear unbiased prediction (ssGBLUP) approach (O. F. Christensen & Lund, 2010; Legarra et al., 2009; Misztal et al., 2009). In ssGBLUP, in addition to all phenotypic data, genomic (**G**) and pedigree-based (**A**) relationships are combined into a single hybrid covariance matrix (**H**). Since **H** is also considered a relationship matrix, the ssGBLUP approach can be implemented using standard mixed model equations. The ssGBLUP can better estimate genetic relationships, thereby providing more accurate predictions for all individuals in the pedigree. Likewise, increasing the sample size by including phenotypic information from nongenotyped individuals can contribute to increased accuracy (O. Christensen et al., 2012).

Genomic prediction has recently been increasingly applied to sugarcane (Gouy et al., 2013; Yadav, Wei, et al., 2021). However, the use of pedigree information in sugarcane is relatively limited with the exception of some studies based on a multiple kernel approach (Deomano et al., 2020; Inamori et al., 2024) and ssGBLUP (Hayes et al., 2021). In animal breeding, a number of studies have investigated the effectiveness of ssGBLUP and the approach is now routinely used in many livestock industries. In forest tree breeding, a growing body of research using ssGBLUP can be found (Cappa et al., 2019; Paludeto et al., 2021; Ratcliffe et al., 2017; Thavamanikumar et al., 2020; Walker et al., 2022). In crops, Ashraf et al. (2016) were the first to evaluate the performance of ssGBLUP using multi-environment wheat trial data, and Pérez-Rodríguez et al. (2017) used a larger reference set of wheat lines reporting the superiority of ssGBLUP over other prediction models. In addition, ssGBLUP has been used in the analyses of fruit trees (Imai et al., 2019), potato (Sood et al., 2020), rice (Morais et al., 2018; Morais Junior et al., 2018), and maize (de Oliveira et al., 2020).

Therefore, the objective of this study was to evaluate the utility of ssGBLUP for the genetic evaluation of a sugarcane segregating population in which only a subset of individuals was genotyped. We also compared PBLUP and GBLUP models to serve as baselines. We used data from an early selection stage generated by the sugarcane breeding program of the Universidade Federal de Viçosa, Brazil (PMGCA-UFV).

2.3. Material and methods

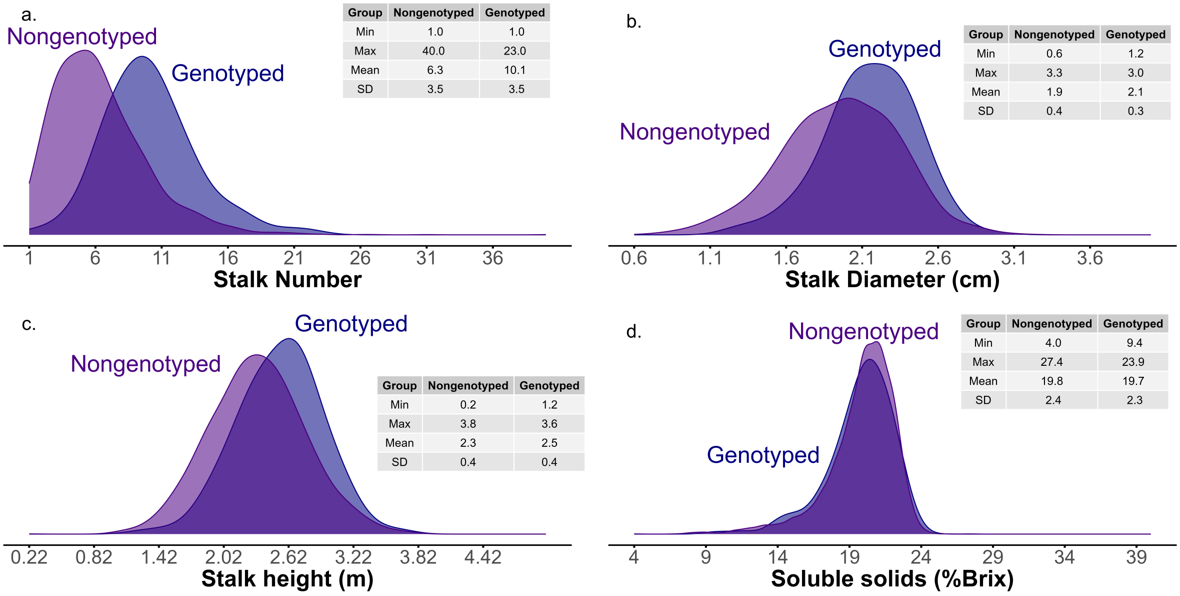
2.3.1. Data

The experimental data used in this study came from field trials established by PMGCA-UFV. An F1 breeding population was created by polycrosses performed at the Flowering and Crossing Station of Serra do Ouro, located in Murici, AL, Brazil. True seeds were transported to the Sugarcane Research and Breeding Center in Oratórios, MG, Brazil (20° 25’ S, 42° 48’ W; 494 m above sea level and typical eutrophic red latosol with clayey texture) and germinated in plastic trays.

In December 2010, seedlings from 124 half-sib families were planted in a field trial at the Sugarcane Research and Breeding Center under commercial growing conditions. Families were laid out in a randomized complete block design with six replicates. The resulting plots consisted of single furrows, 5 m long and 1.40 m apart, containing 10 seedlings planted as single stools. Each family was thus represented by 60 different genotypes, for a total of 7,440 individual plants. In May 2012, the following traits were measured on the first ratoon crop on a stool-mean basis: stalk number, stalk diameter (cm), stalk height (m), and soluble solids (Brix%). A total of 377 individuals were selected from the original population for genomic prediction analysis (Figure 1).

Genomic DNA was extracted from leaf tissue and samples were genotyped using a target sequencing system, specifically hybridization-based enrichment sequencing

Figure 1 – Density curve estimates representing the distribution of phenotypes (a. Stalk Number; b. Stalk Diameter; c. Stalk Height, and d. Soluble Solids) of the sugarcane population evaluated. Min: minimum; Max: maximum; SD: standard deviation.



Source: The author, 2024, p. 22

(Neves et al., 2013). In contrast to whole genome systems, target sequencing focuses on specific regions of the genome, allowing for a reduction in genome complexity (Mertes et al., 2011). Target genomic regions are captured for DNA library preparation and hybridized to biotinylated DNA single-stranded oligonucleotides (probes). The probes used in this study were synthesized by RAPiD Genomics (Gainesville, FL, US) using publicly available expressed sequence tags and whole-genome shotgun sequences. Subsequently, DNA libraries were sequenced and the raw sequence data were processed using the same criteria as in Gonçalves et al. (2021). After quality control, m = 124,307 SNP markers were available for subsequent analysis.

The pedigree information used in this study was compiled from a database maintained by PMGCA-UFV, a publicly available database (http://tropgenedb.cirad.fr/tropgene/JSP/index.jsp) developed by CIRAD (French Agricultural Research Center for International Development), and through personal communication with researchers. As a result, we traced relationships back to the founders for which information about further ancestors was unknown and individuals were assumed to be unrelated (Table 1). We considered individuals with at least three generations of records (Atkin et al., 2009).

Table 1 – Pedigree information of the sugarcane population used in our study.

Identification	Generation	No. of individuals
Nonphenotyped and Nongenotyped	Founders and Parents	369
Phenotyped and Nongenotyped	F1	3704
Phenotyped and Genotyped	F1	377
Total	-	4450

Source: The author, 2024, p. 23

2.3.2. Phenotypic data analysis

We adjusted phenotypes before fitting genomic prediction models. We derived BLUP values for each trait without including pedigree information (Piepho et al., 2008) to avoid double shrinkage of response variables used in subsequent analysis (Hayes et al., 2021; Yadav et al., 2023). We performed the analysis in the Selegen-REML/BLUP software using the following linear mixed model (Resende, 2016):

$$y_{ijk} = r_j + g_i + p_k + \epsilon_{ijk}, \tag{1}$$

where y_{ijk} is the phenotype of the *i*th individual measured in the *j*th replicate (block) and *k*th plot; r_j is the fixed effect of replicate (block) added to the general mean (μ); g_i is the random effect of the *i*th genotype; p_k is the random effect of plot; and ϵ_{ijk} is the residual term. The distributional assumptions for the random effects were as follows: g_i ~ N(0,σ_g^2); p_k ~ N(0,σ_p^2); and ϵ_{ijk} ~ N(0,σ_ϵ^2). Finally, we calculated deregressed BLUP values as follows:

$$deregressedBLUP = \frac{BLUP}{1 - \frac{PEV}{\hat{\sigma_g^2}}} \tag{2}$$

where PEV, and σ_g^2 are the prediction error variance, and genotype variance for each trait obtained from equation (1), respectively.

2.3.3. Quantitative genetic models

Here we describe the models used for the genetic evaluation of the sugarcane segregating population considered in this study. We used the BLUP values of the evaluated traits as response variables in all the models. We used a basic linear model of the form:

$$\mathbf{y}^* = \mathbf{1}\mu + \mathbf{Zu} + \boldsymbol{\epsilon}, \tag{3}$$

where $\mathbf{y}^*$ is a vector of adjusted phenotypes obtained from equation (2); **1** is a vector of ones; μ is the overall mean; **Z** is an incidence matrix associating phenotypes and random additive genetic effects; $\mathbf{u}$ is a vector of additive genetic effects; and ϵ is a vector of residuals. It was assumed that $u \sim N(0, \mathbf{K}\sigma_u^2)$ and $\epsilon \sim N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{K}$ is an $n \times n$ relationship matrix; n is the number of clones; $\mathbf{I}$ is an identity matrix; and σ_u^2 and σ_e^2 are the additive genetic variance and residual variance, respectively. The matrix $\mathbf{K}$ takes one of the five relationship matrices, $\mathbf{A}$, $\mathbf{A}_{22}$, $\mathbf{G}_v$, $\mathbf{H_v}$,and $\mathbf{H_a}$, defined below. Hereinafter, models using these relationship matrices are referred to as PBLUP-1, PBLUP-2, GBLUP, ssGBLUP-1, and ssGBLUP-2, respectively (Table 2).

Table 2 – Summary of models and respective relationship matrices.

Model	Relationship matrix	Matrix dimension	Information source
PBLUP-1	$\mathbf{A}$	4,081 × 4,081	Pedigrees
PBLUP-2	$\mathbf{A}_{22}$	377 × 377	Pedigrees
GBLUP	$\mathbf{G}_v$	377 × 377	Genotypes
ssGBLUP-1	$\mathbf{H_v}$ ($\mathbf{G}_v$ + $\mathbf{A}$)	4,081 × 4,081	Pedigrees + Genotypes
ssGBLUP-2	$\mathbf{H_a}$ ($\mathbf{G}_a$ + $\mathbf{A}$)	4,081 × 4,081	Pedigrees + Genotypes

Source: The author, 2024, p. 24

2.3.4. PBLUP models

For PBLUP, all phenotyped individuals were included in the analysis. We excluded missing plots due to seedling death and considered a minimum depth of three generations in the pedigree (progeny, parents, and grandparents), so we used 4,081 of the original 7,440 individual plants. This resulted in the 4,081 × 4,081 relationship matrix $\mathbf{A}$ (Table 2). The 377 × 377 relationship matrix $\mathbf{A}_{22}$ was constructed by subsetting the $\mathbf{A}$ matrix, considering only phenotyped and genotyped individuals. The $\mathbf{A}$ matrix was derived using the AGHmatrix R package (Amadeu et al., 2016).

2.3.5. GBLUP models

For GBLUP, only the phenotyped and genotyped individuals were considered. The genomic relationship matrix $\mathbf{G}_v$ was obtained as:

$$\mathbf{G}_v = \frac{\mathbf{W}\mathbf{W}'}{2\sum_{l=1}^{m} p_l(1-p_l)},$$

where $\mathbf{W}$ is the centered codominant diploidized SNP marker matrix with elements coded as 2 if homozygous for the reference allele, 1 if heterozygous, and 0 if homozygous for the alternative allele, and p_l is the allele frequency of the lth marker (VanRaden, 2008). Here, the reference is a monoploid sugarcane genome sequence (Garsmeur et al., 2018). We used the AGHmatrix R package to compute $\mathbf{G}_v$ (Amadeu et al., 2016).

2.3.6. ssGBLUP models

For ssGBLUP, all phenotyped individuals were included in the analysis, regardless of whether they were genotyped or not. Compatibility between genomic and pedigree information is an important subject when applying the ssGBLUP approach (Misztal et al., 2013). Differences in their genetic bases can lead to biased estimates and loss of accuracy. Also, this lack of compatibility can affect the estimation of variance components and result in poor convergence rates (C.-Y. Chen et al., 2011). Several approaches have been proposed to make pedigree- and marker-based relationship matrices compatible (Aguilar et al., 2010; Vitezica et al., 2011). One strategy is to scale $\mathbf{G}_v$ so that its diagonal (diag) and off-diagonal (off.diag) elements match those of the $\mathbf{A}_{22}$

matrix (O. Christensen et al., 2012):

$$\mathbf{G}_a = \beta\mathbf{G}_v + \alpha,$$

where β and α were estimated using the following system of linear equations:

$$\begin{cases} \text{Average}(\text{diag}(\mathbf{G}_v))\beta + \alpha = \text{Average}(\text{diag}(\mathbf{A}_{22})) \\ \text{Average}(\text{off.diag}(\mathbf{G}_v))\beta + \alpha = \text{Average}(\text{off.diag}(\mathbf{A}_{22})) \end{cases}$$

We then combined $\mathbf{A}$ with either $\mathbf{G}_v$ or $\mathbf{G}_a$ to derive the hybrid relationship matrix $\mathbf{H_v}$, and $\mathbf{H_a}$, respectively, for the ssGBLUP models:

$$\mathbf{H_o} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}(\mathbf{G_o} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G_o} \\ \mathbf{G_o}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G_o} \end{bmatrix},$$

where o takes v or a depending on whether $\mathbf{G}_v$ or $\mathbf{G}_a$ was used; $\mathbf{A}_{11}$ is the subset of $\mathbf{A}$ for nongenotyped individuals; $\mathbf{A}_{12}$ and $\mathbf{A}_{21}$ are the relationship between nongenotyped and genotyped individuals; $\mathbf{A}_{22}$ is the relationship between genotyped individuals; and $\mathbf{G_o}$ takes one of the genomic relationship matrices ($\mathbf{G}_v$ or $\mathbf{G}_a$) described above.

To gain a better understanding of how the ratio of genotyped to nongentoyped individuals impacts predictions, we randomly sampled and excluded nongenotyped individuals from the population and fitted ssGBLUP models. We tested the following ratios of genotyped to nongenotyped individuals: 9:100, 20:100, 30:100, and 40:100.

2.3.7. Assessment of prediction models

Our primary goal was to investigate the effectiveness of the ssGBLUP approach. Therefore, we compared the performance of ssGBLUP-1 ($\mathbf{H_v}$) and ssGBLUP-2 ($\mathbf{H_a}$) with PBLUP-1 ($\mathbf{A}$), PBLUP-2 ($\mathbf{A}_{22}$), and GBLUP ($\mathbf{G}_v$). Two cross-validation (CV) scenarios were evaluated. In the first scenario (CV1), we only considered the prediction of genotyped individuals, so we compared all PBLUP, GBLUP, and ssGBLUP models. In the second scenario (CV2), we considered the prediction of non-genotyped individuals and thus compared PBLUP-1, ssGBLUP-1 and ssGBLUP-2. We generated 20 partitions using repeated random sampling CV. For each partition, approximately 80 % of the individuals were assigned to the training set and the remaining 20 % were assigned to the test set. The test set consisted of 77 and 740 observations in the first and second scenarios, respectively. The size of the training set varied across the different modeling approaches and prediction scenarios (Figure 2). In CV1, the training set for GBLUP and PBLUP-2 consisted of 300 individuals, while the training set for PBLUP-1 and ssGBLUP consisted of 4,004 individuals. In CV2, PBLUP-1 and ssGBLUP were trained with 3,341 individuals. To evaluate the performance of the genomic prediction models, we used prediction accuracy, calculated as the Pearson correlation between adjusted phenotypes and predicted genetic values. Mean prediction accuracy estimates between the evaluated models were then evaluated using the Tukey test (α < 0.05).

Variance components were estimated using equation (3) and then used to calculate a narrow-sense genomic heritability estimate for each relationship matrix using the following formula:$\hat{h}^2 = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2+\hat{\sigma}_e^2}$

2.3.8. Implementation

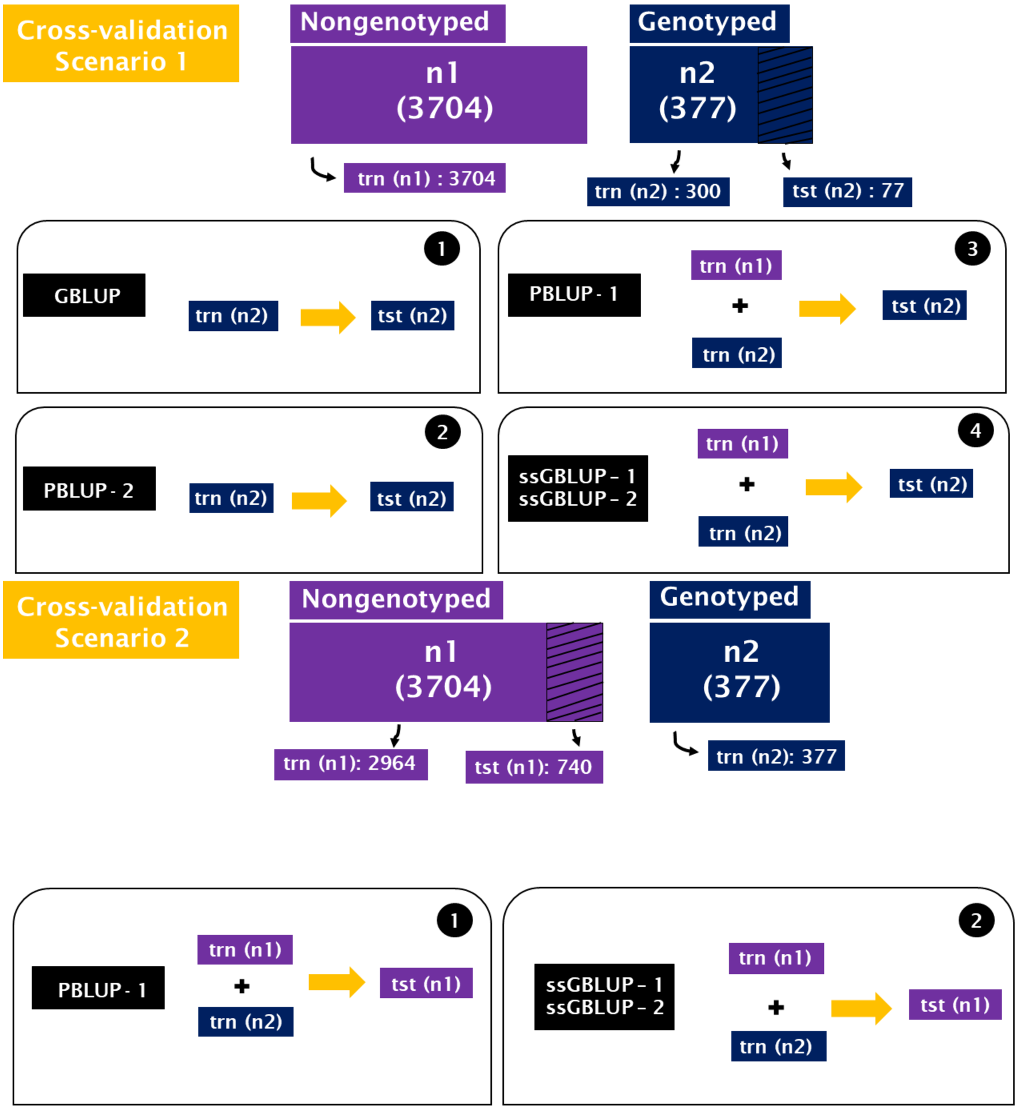
All models were implemented using a Markov chain Monte Carlo Bayesian approach with the R package BGLR (Pérez & de Los Campos, 2014). We assumed a Gaussian likelihood function and assigned a scaled inverse chi-squared distribution to the additive genetic and residual variance components to obtain the joint posterior distribution of the model unknown parameters:

$$p(\mu, \mathbf{u}, \sigma_u^2, \sigma_e^2|\mathbf{y}) \propto \prod_{i=1}^{n} N(\mu + \mathbf{u_i}, \frac{\sigma_e^2}{n_i})N(\mathbf{u_i}|0, \mathbf{K}\sigma_u^2)N(0, \mathbf{I}\sigma_e^2)$$

$$\chi^{-2}(\sigma_e^2|\nu_\epsilon, S_\epsilon) \quad \chi^{-2}(\sigma_u^2|\nu_\epsilon, S_\epsilon)$$

where χ^{-2} is the scaled inverse chi-squared distribution with v and S degrees of freedom and scaling parameter, respectively. In each analysis, 25,000 Markov chain Monte Carlo samples were generated and the first 10,000 samples were discarded as burn-in with a sampling interval of five. We verified the convergence of the chains using the Geweke diagnostic (Geweke, 1992) of sample residual estimates using the R package coda (Plummer et al., 2006).

Figure 2 – Schematic diagram showing cross-validation scenarios. n1: number of nongenotyped individuals; n2: number of genotyped individuals; trn: training; tst: test



Source: The author, 2024, p. 28

2.4. Results

2.4.1. Phenotypic data and genetic parameters estimates

The phenotypic distributions of all measured traits approximately followed the shape of a normal distribution (Figure 1). Due to the selection performed on the population, genotyped individuals had slightly higher mean phenotypic values than nongenotyped individuals, except for soluble solids. In terms of dispersion, little difference was observed, with similar standard deviation values between the two groups for all traits.

The genomic heritability estimates for the traits evaluated are shown in Table 3. Overall, moderate to high heritability estimates were observed, ranging from 0.68 to 0.91. For all traits, PBLUP-1 and ssGBLUP had the highest heritability estimates. PBLUP-2 showed the lowest heritability estimates for all traits. The GBLUP models showed genomic heritability estimates between PBLUP-1 and ssGBLUP.

Table 3 – Genomic heritability estimates based on pedigree (h_P^2), genomic (h_G^2) and the combination of both information (h_H^2) for stalk number (SN), stalk diameter (SD), stalk height (SH), and soluble solids (SS).

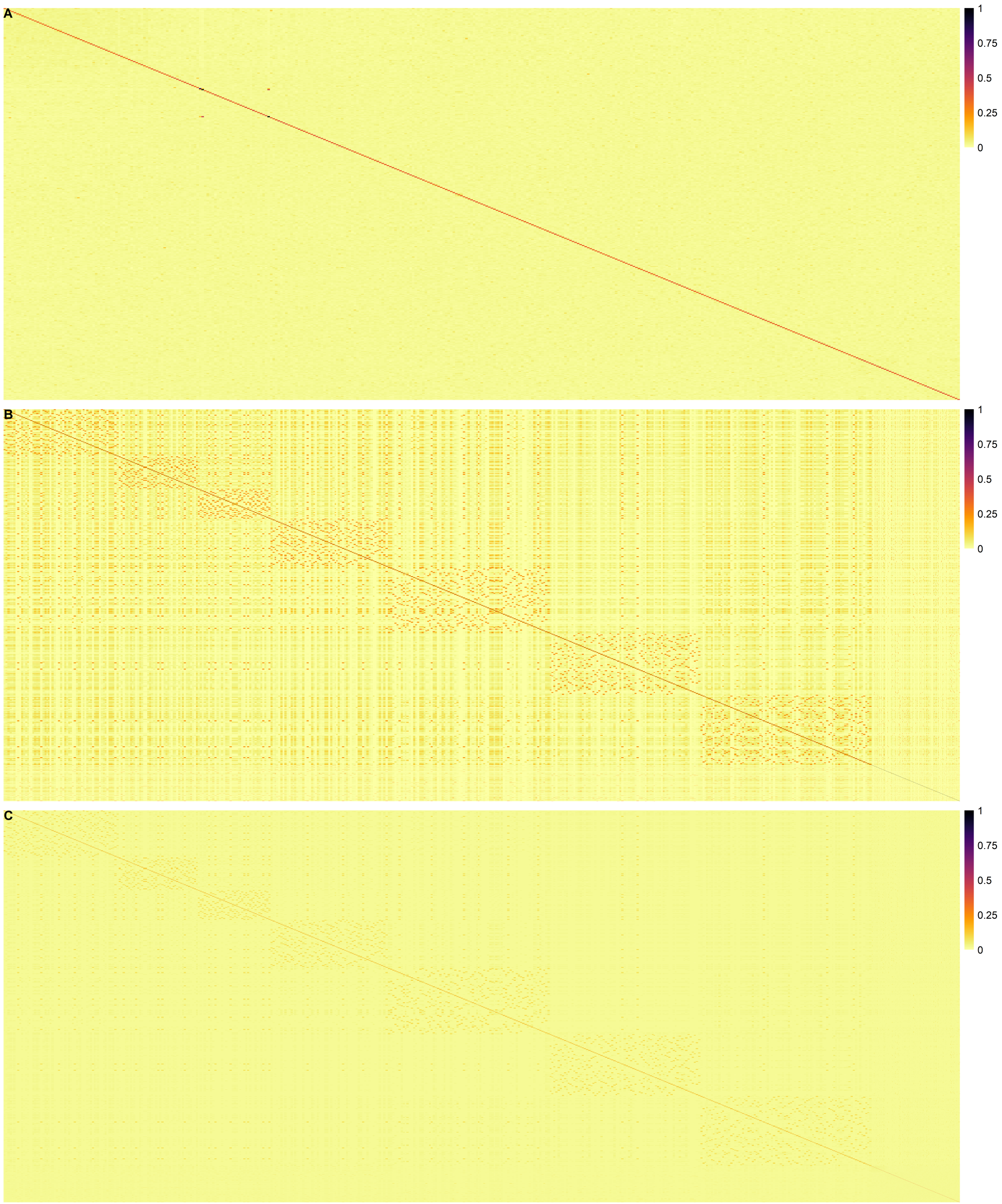
Model	SN			SD			SH			SS		
	h_P^2	h_G^2	h_H^2	h_P^2	h_G^2	h_H^2	h_P^2	h_G^2	h_H^2	h_P^2	h_G^2	h_H^2
PBLUP-1 [a]	0.91	-	-	0.88	-	-	0.82	-	-	0.85	-	-
PBLUP-2 [b]	0.80	-	-	0.76	-	-	0.69	-	-	0.68	-	-
GBLUP [c]	-	0.81	-	-	0.79	-	-	0.78	-	-	0.82	-
ssGBLUP-1 [d]	-	-	0.90	-	-	0.87	-	-	0.85	-	-	0.82
ssGBLUP-2 [e]	-	-	0.91	-	-	0.88	-	-	0.83	-	-	0.85

Notes:
[a] Pedigree-based model using $\mathbf{A}$ matrix (all phenotyped individuals considered).
[b] Pedigree-based model using $\mathbf{A}_{22}$ matrix (only phenotyped and genotyped individuals considered).
[c] Genomic-based model using $\mathbf{G}$ matrix (only phenotyped and genotyped individuals considered).
[d] Hybrid model using $\mathbf{H}_\mathbf{v}$ matrix(all phenotyped individuals considered).
[e] Hybrid model using the $\mathbf{H}_\mathbf{a}$ matrix (all phenotyped individuals considered).

Source: The author, 2024, p. 29

To examine the relationships among individuals in the sugarcane population considered, we normalized the elements of the $\mathbf{A}$, $\mathbf{G}_v$, and $\mathbf{H}_v$ matrices so that their off-diagonal elements are bounded between 0 and 1. The relationships among individuals were illustrated using heatmaps (Figure 3). The heatmaps showed that the relationships among individuals represented by the $\mathbf{G}_v$ matrix were modest. We found stronger relationships in the $\mathbf{A}$ matrix compared to those in $\mathbf{G}_v$ and $\mathbf{H}_v$.

Figure 3 – Hetmaps of the marker-based (A), pedigree-based (B), and hybrid (C) relationship matrices. The darker the color, the stronger the relationship.

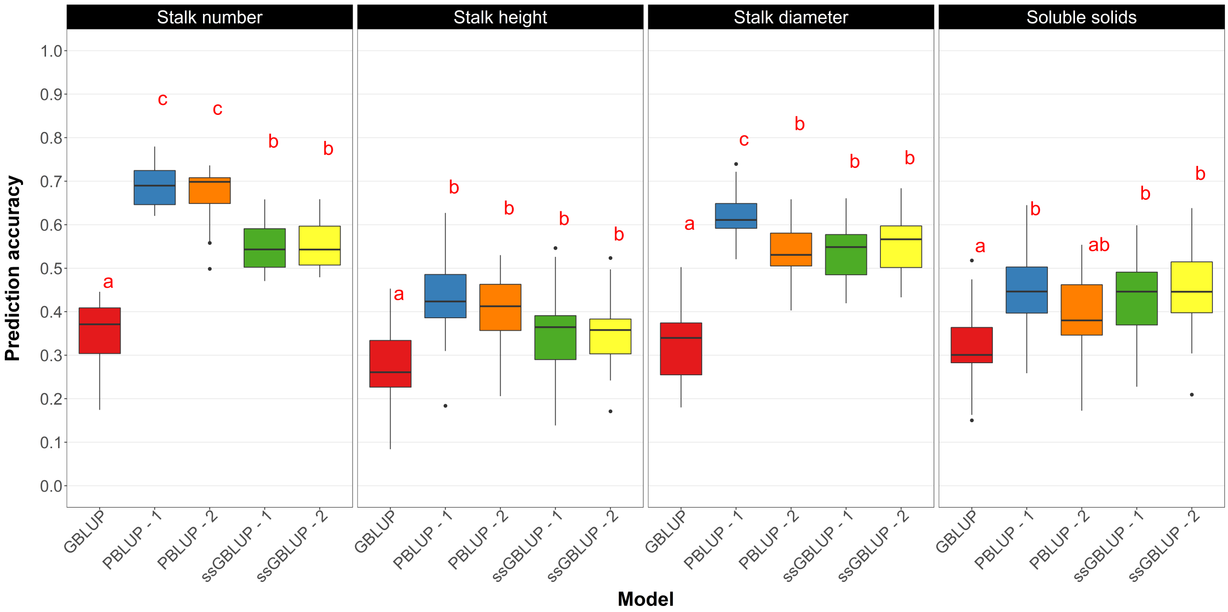


Source: The author, 2024, p. 30

2.4.2. Prediction models

Genomic prediction accuracies for CV1 and CV2 are shown in Figures 4 and 5, respectively. In CV1, the models were validated using genotyped individuals. The values of the alpha and beta parameters obtained after adjusting the G matrix were α = 0.013 and β = 2.023. Two ssGBLUP models, ssGBLUP-1 and ssGBLUP-2, were similar in predictive performance. Overall, GBLUP showed the lowest prediction accuracy for all traits. We found that ssGBLUP and PBLUP had superior performance compared to GBLUP models. For stalk height and soluble solids, the differences between ssGBLUP and PBLUP were not significant, but for stalk number and stalk diameter, the PBLUP models showed higher prediction accuracies.

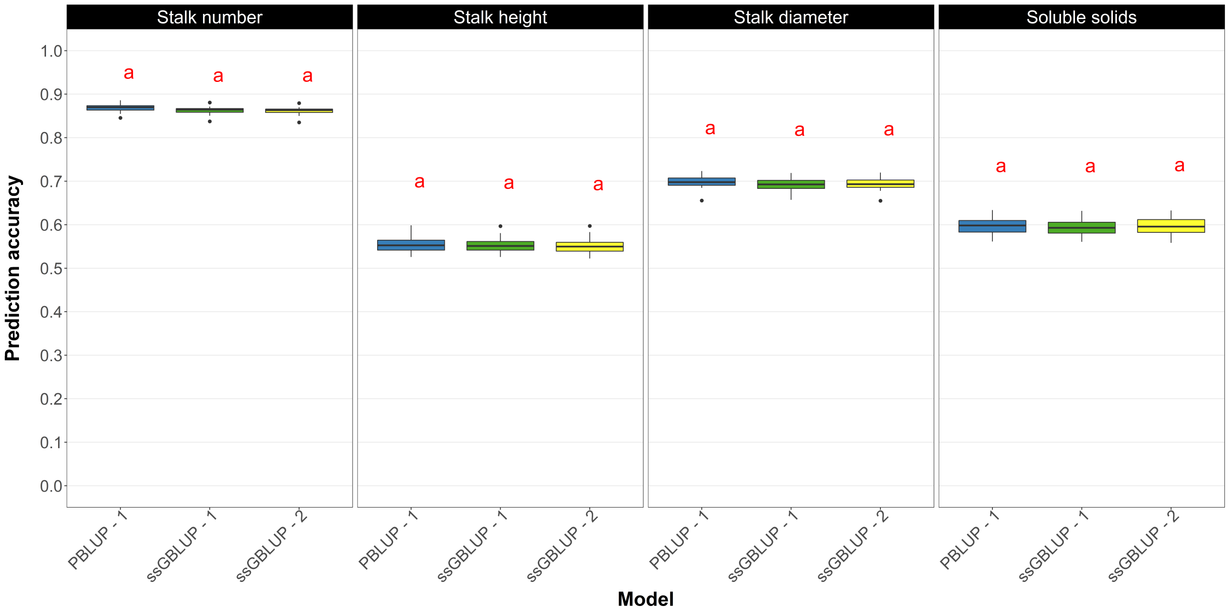
Figure 4 – Validation accuracy data distributions for prediction scenario 1 based on 20 cross-validation replicates. Means followed by a common letter are not significantly different by Tukey's test of multiple comparisons at the 5% level of significance



Source: The author, 2024, p. 31

In CV2, the models were validated using non-genotyped individuals. In this case, we observed an increase in accuracy and a decrease in prediction uncertainty. Also, for all traits, PBLUP and ssGBLUP were comparable, and the mean prediction accuracies were not significantly different.

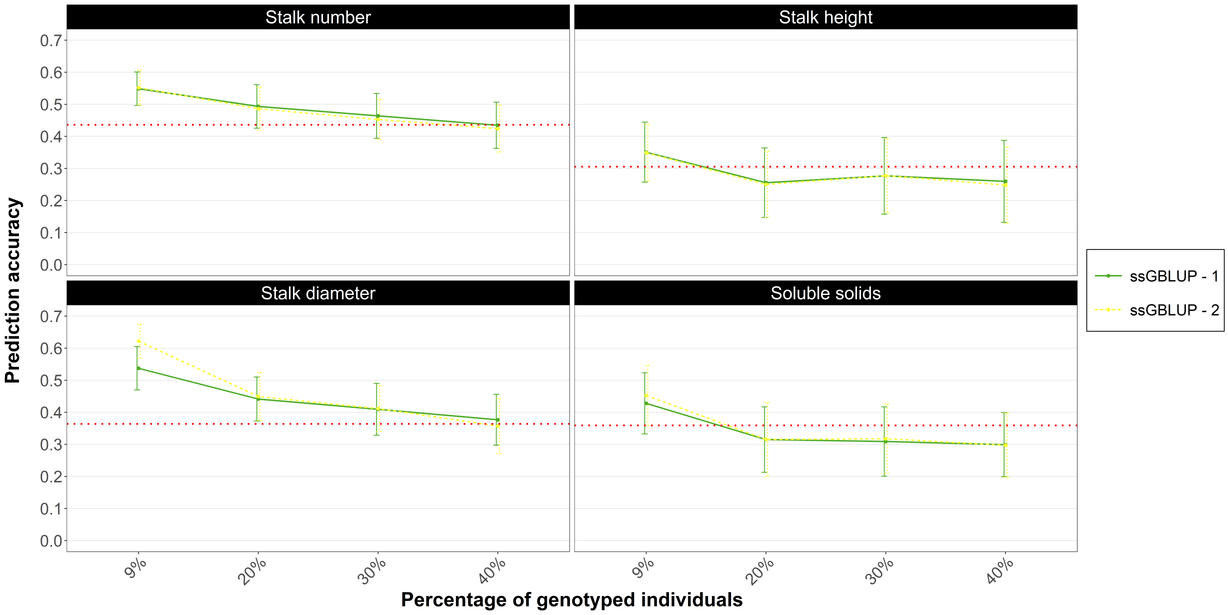
Figure 5 – Validation accuracy data distributions for prediction scenario 2 based on 20 cross-validation replicates. Means followed by a common letter are not significantly different by Tukey's test of multiple comparisons at the 5% level of significance



Source: The author, 2024, p. 32

We tested different ratios of genotyped individuals in the population and found that the accuracies of the ssGBLUP models tended to decrease and resulted in similar accuracies to the GBLUP models as the proportion of genotyped individuals progressively increased and the training set size decreased (Figure 6).

Figure 6 – Prediction accuracies of ssGBLUP models across different ratios of genotyped individuals. Red dashed horizontal lines indicate the mean prediction accuracy of GBLUP models.



Source: The author, 2024, p. 33

2.5. Discussion

In this study, we evaluated the performance of genetic prediction models using markers (GBLUP), pedigrees (PBLUP), and the combination of both sources of information (ssGBLUP). Ongoing research advances are increasingly allowing the use of marker information for plant breeding applications. Although phenotypic selection has been common in sugarcane, genomic prediction can have a major impact on genetic progress. Simulation studies have investigated the potential benefits of this approach as a breeding tool in sugarcane compared to phenotypic selection (Voss-Fels et al., 2021). Experimental studies on traits such as cane yield and its components, quality, sugar yield, and disease resistance have also been conducted with encouraging results (Batista et al., 2021; C. Chen et al., 2023; Deomano et al., 2020; Gonçalves et al., 2021; Inamori et al., 2024; Islam et al., 2021, 2022; Islam et al., 2023; O'Connell et al., 2022; Satpathy et al., 2022; Yadav, Wei, et al., 2021). However, genomic information may not be available for all phenotyped individuals due to budget constraints. In contrast, pedigree records are inexpensive and easier to obtain. There is growing evidence that the joint use of genomic and pedigree information can improve estimates of genetic relationships and the accuracy of prediction models (Crossa et al., 2014; Velazco et al., 2019). Thus, modeling strategies that can efficiently integrate different sources of genetic information, such as ssGBLUP, are highly desirable.

2.5.1. Comparison between GBLUP versus PBLUP

Although most studies differ in terms of target traits, statistical model, training population size, marker density, genotyping platform, genotyping protocol, and CV schemes, the overall GBLUP accuracies we found corroborate those reported in previous genomic prediction studies in sugarcane (Deomano et al., 2020; Hayes et al., 2021; Yadav, Wei, et al., 2021). For example, using an equivalent model, (Gouy et al., 2013) found accuracies of 0.31-0.47 for stalk number and stalk diameter. They also reported slightly higher accuracies for soluble solids, ranging from 0.42-0.55. Similar results were observed for stalk height (Aono et al., 2022). In contrast, lower accuracies were reported for stalk number, stalk diameter, stalk height, and soluble solids (Batista et al., 2021; Islam et al., 2022).

In our study PBLUP showed better prediction accuracies than GBLUP for all traits analyzed, which is different from what has been reported in the literature in other species (Albrecht et al., 2014; Ashraf et al., 2016; Auinger et al., 2016; Burgueño et al., 2012; Crossa et al., 2010; Endelman et al., 2018; Morais et al., 2018). This finding can be interpreted in terms of two main explanations.

The first possible reason is due to deficiencies in genomic information. Modern interspecific sugarcane hybrids have a large and complex nuclear genome ($\sim$ 10 Gb), which shows high ploidy levels with frequent occurrence of aneuploidy (Garcia et al., 2013; Thirugnanasambandam et al., 2018). Therefore, it is possible that the density of our marker panel was insufficient to capture all causal loci. However, previous studies evaluating target enrichment sequencing and other similar genotyping platforms with lower marker densities suggest that this may not be the most relevant factor in sugarcane genomic prediction given the high linkage disequilibrium present in current elite breeding populations (Deomano et al., 2020; Gemenet et al., 2020; Hayes et al., 2021; Islam et al., 2021; Raboin et al., 2008; Yang et al., 2019). It has been argued that other sequence variants, such as copy number variation, may help to further explain genetic variance (Schiessl et al., 2019; Weber et al., 2023). Regarding the genotyping protocol, although reduced representation DNA sequencing methods are powerful and more suitable for polyploid genomes, this type of data is likely to suffer from genotyping errors, such as low quality genotype calls, potentially leading to increased rates of missing data (Bourke et al., 2018; Endelman et al., 2018; Ferrão et al., 2021; Gemenet et al., 2020). The choice of allele dosage and ploidy parameterization models during genotype calling can also affect the genomic prediction of polyploids. For example, de C. Lara et al. (2019) found that models that included tetraploid dosage had slightly higher performance compared to pseudo-diploid models. Also, Yadav et al. (2023) reported small accuracy gains of 5–7% by using continuous genotyping parameterization. However, recent studies have shown that improvements with the inclusion of allele dosage information may depend on the ploidy level, trait, and genotype class frequencies in the population (Batista et al., 2021; de Bem Oliveira et al., 2019; Nyine et al., 2018). Another parameter that could be optimized is the sequencing depth or read depth (de Bem Oliveira et al., 2020; Wang et al., 2022).

The second possible reason is related to peculiarities of family or population structure. The breeding value of an individual can be factorized as the sum of the parental average (i.e., half of the breeding value of each female and male parent) and the Mendelian sampling term. One of the major advantages of using GBLUP is the ability to exploit within-family variance through efficient estimation of Mendelian sampling (H. Daetwyler et al., 2007). In contrast, PBLUP heavily relies on tracing gene flow through common ancestry by estimating the parental average component of breeding value, thereby emphasizing more the between-family variation (Werner et al., 2020). Because we used a CV design with random partitioning, the validation sets consisted of individuals from different families. This suggests that by using this CV design, both the parental average and the Mendelian sampling term were determinants of breeding value predictions (Werner et al., 2020).

Genetic diversity is often characterized by effective population size (Ne), which quantifies the rates of genetic drift and inbreeding. Hence, Ne serves as an indicator of changes in allele frequencies in a given population (Vencovsky & Crossa, 2003) and Ne can be predicted for a half-sib family (Resende & Barbosa, 2006; Vencovsky, 1978). To make adequate inferences based on half-sib families, a sample size of 300 is recommended because it represents 99 % of the maximum Ne (Peternelli & de Resende, 2015). In our study, the family size consisted of 60 individuals due to the scarcity of cropland and technical difficulties in setting up larger field trials. This sample size may not have allowed the full exploitation of all possible desirable genotype combinations in the segregating progeny, reducing the probability of the occurrence of top-ranking individuals superior to their parents partly due to Mendelian sampling deviations (Antonios et al., 2023; M. Barbosa et al., 2005; Bijma et al., 2020). Therefore, it is possible that this relatively modest family size was insufficient to exploit within-family haplotype variation, thereby reducing the advantages of GBLUP. A recent study also reached the same conclusions to explain the high accuracies obtained with pedigree-based linear kernel models for predicting disease resistance in wheat (Juliana et al., 2017). Similarly, analysis of simulated data showed an increase in the relative importance of Mendelian deviations as family size and heritability increased (Vela-Avitúa et al., 2015). However, another possibility is that this particular family or population structure combined with a larger sample size was better at estimating the breeding value of close relatives up and down the pedigree, thus favoring PBLUP over GBLUP (H. Daetwyler et al., 2007).

Another potential issue is a combination of the short breeding history of modern elite sugarcane germplasm and the breeding schemes used. It has been argued that modern sugarcane breeding populations exhibit substantial linkage disequilibrium (Barreto et al., 2019; Raboin et al., 2008; Yadav, Ross, et al., 2021). This high extent of linkage disequilibrium may be due to two main factors. First, the limited number of founder genitors involved during the initial breeding efforts coupled with the high similarity of the parental pools (Aitken et al., 2006; Todd et al., 2015; Yang et al., 2019; Zan et al., 2020). Second, the low number of crossing cycles leading to fewer opportunities for recombination events (Fickett et al., 2020; Yang et al., 2019). Together, these aspects can be associated with a narrow genetic base, implying low Ne (Marjanovic & Calus, 2020). Low Ne estimates are associated with loss of heterozygosity. For example, such a founder event, previously described in the breeding history of sugarcane, can create a population bottleneck, leading to a rapid decay of heterozygosity (Yang et al., 2018). The high genomic inbreeding coefficient estimates in modern sugarcane varieties obtained using runs of homozygosity further support this argument (Yang et al., 2020).

The degree of heterozygosity is considered an important factor that can influence the amount of Mendelian sampling variance. The two factors are positively proportional (Bonk et al., 2016). The PMGCA-UFV program practices family selection in early field trials, followed by recycling of advanced selection candidates as parents in new hybridization (Cursi et al., 2020). However, this can promote inbreeding due to co-selection of siblings resulting in loss of heterozygosity, especially in truncation selection schemes (H. Daetwyler et al., 2007). For example, the simulated individual best linear unbiased prediction (BLUPIS) method is designed to optimize selection between and within families. Using BLUPIS, families with genotypic values below a certain threshold are discarded (M. Barbosa et al., 2005; Resende & Barbosa, 2006). Furthermore, due to the complex reproductive biology of sugarcane, a high percentage of selfed progeny can occur after hybridization (Costa et al., 2014; Todd et al., 2020).

The loss of heterozygosity may be compensated by high ploidy levels (Comai, 2005; Costa et al., 2014), which may explain the lack of performance loss for soluble solids, stalk number, and stalk diameter reported in the literature (M. d. A. Silva & Gonçalves, 2011). Batista et al. (2021), who evaluated allele dosage information in ploidy levels ranging from 6 to 14, reported a low frequency of classes of heterozygous genotypes in an F1 biparental sugarcane population, possibly indicating low heterozygosity. The genetic variance among the progeny of modern sugarcane hybrids is expected to be conditioned by the proportion of the *Saccharum officinarum* to *Saccharum spontaneum*

(*S. spontaneum*) subgenomes, that each individual inherits. The greater the proportion of the *S. spontaneum* genome, the greater the variability (Matsuoka et al., 2014; Thirugnanasambandam et al., 2018; Yang et al., 2019). Moreover, theoretically, the contribution of Mendelian deviations in determining breeding values is expected to be greater as individuals are more distantly related to the base population (Walsh & Lynch, 2018). Hence, in addition to the possibility that Mendelian deviations may not be adequately expressed due to limited family size, Mendelian deviations may also be negligible as a consequence of a short breeding history and limited genetic diversity (Lauer et al., 2021).

Previous studies support this explanation. M. Barbosa et al. (2001) used real data to investigate the effect of sampling on the inference of sugarcane family means. The results found by the authors suggest low within-family variance, since no significant differences were found between rows within family plots for traits with low heritability, such as cane yield and its components (e.g., stalk number). Conversely, this is inconsistent with the high, significant estimates of within-plot (or within-family) variability reported in previous studies (Mbuma et al., 2020; Pisaroglo de Carvalho et al., 2014; F. L. d. Silva et al., 2015). However, this observation is controversial because it is often difficult to obtain reliable estimates of within-family variance in sugarcane (Atkin et al., 2009; M. Barbosa et al., 2001). Genotype by microenvironment interactions may account for a significant proportion of the phenotypic variation expressed by plants, particularly in family field trials, which consist of unreplicated individual genotypes planted in small plots subject to interplot competition effects such as differential light incidence (Brasileiro et al., 2016; Cursi et al., 2020; Jackson & McRae, 2001). Taken together, it seems reasonable to suggest that Mendelian deviations were small in our population, which was stratified into multiple segregating homogeneous families. The same explanation has been proposed by Juliana et al. (2018), who found equivalent and slightly superior prediction accuracies of pedigree-based compared to genomic-based models for yield in wheat.

Pedigree records can contain several errors. For example, individuals may be assigned to incorrect parents, and some parents may be unknown. These errors reflect the complex reproductive biology of sugarcane and the level of pollen control during hybridization. In our study, progeny were derived using polycrosses. In polycrosses, half-sibs may actually be full-sibs or even selfing may occur (Tew & Pan, 2010). Despite these challenges in using pedigree information, our results showed that PBLUP performed better than GBLUP for sugarcane traits. Similar results have been reported in previous studies (Beaulieu et al., 2014; Cappa et al., 2019; Z.-Q. Chen et al., 2018; Klápště et al., 2020), especially for perennial species (Beaulieu et al., 2022; Cros et al., 2014). For example, Nadeau et al. (2023) investigated the effect of sample size on genetic parameter estimates and breeding value accuracy in white spruce (*Picea glauca* [Moench] Voss) and argued that the equivalent performance of GBLUP and PBLUP was due to a small number of trees per family and lower within-family genetic variance. Also, Simiqueli et al. (2023) reported that PBLUP outperformed GBLUP models tested in different intra- and interspecific breeding populations of *Eucalyptus*.

2.5.2. ssGBLUP

In this study, our main objective was to investigate the utility of ssGBLUP for the genetic evaluation of sugarcane clones at an early stage of progeny testing. We evaluated two CV scenarios. The prediction of the performance of genotyped individuals (CV1) was consistently improved with ssGBLUP compared to GBLUP for all traits. Similarly, a sizeable increase in accuracy using ssGBLUP was reported for sugar content in citrus, but only modest improvements for fruit weight and acid content (Imai et al., 2019). In contrast, negligible improvements in accuracy were reported for advanced sugarcane clones using ssGBLUP compared to GBLUP for all traits evaluated (Hayes et al., 2021). In the present study, the overall performance of ssGBLUP was either comparable to or slightly lower than PBLUP. Previous studies have reported mixed results. For example, Sood et al. (2020) found improvements in accuracy using ssGBLUP compared to PBLUP. A similar result was obtained by Morais et al. (2018), but no clear superiority of ssGBLUP compared to GBLUP. Ashraf et al. (2016) reported that ssGBLUP provided higher prediction accuracy than PBLUP and GBLUP in wheat. For stalk number and stalk height, ssGBLUP showed lower performance compared to PBLUP, which is consistent with the findings of Pérez-Rodríguez et al. (2017) for grain yield in wheat lines and Cappa et al. (2019) for several traits, except lignin composition, in *Eucalyptus* hybrids.

Regarding the prediction of non-genotyped individuals (CV2), non-significant differences in accuracy were observed between ssGLBUP and PBLUP. This is in line with previous studies in *Eucalyptus* hybrids (Simiqueli et al., 2023) and in potato (Sood et al., 2020), which reported that ssGBLUP and PBLUP had similar prediction accuracies for non-genotyped individuals. Also, Imai et al. (2019) found that ssGBLUP and PBLUP were comparable for most traits, further corroborating our results.

We also considered the compatibility between pedigree- and genomic-based relationship matrices. It has been argued that pedigree and genomic relationship matrices are computed assuming genetic parameters from different base populations and hence are on different scales (O. Christensen et al., 2012). In general, we found only marginal and non-significant differences between ssGBLUP-1 and ssGBLUP-2. This may be due to the predominance of recent family relationships in our population, as opposed to more distant relationships (Atkin et al., 2009). This is particularly true for simple pedigrees, which mainly consist of parent-offspring relationships, and as a consequence of the short breeding history of elite sugarcane populations. In our sugarcane data, the $\mathbf{A2}$ and $\mathbf{G}_v$ matrices showed similar estimates of relationships (Figure 3).

The gains provided by ssGBLUP in CV1 can be partly explained by the fact that prediction accuracy is a function of the number of records or training set size (C.-Y. Chen et al., 2011; H. D. Daetwyler et al., 2008; Wientjes et al., 2013). Therefore, a higher prediction accuracy of ssGBLUP compared to GBLUP is expected with the inclusion of phenotypes from nongenotyped individuals (Ashraf et al., 2016; Cappa et al., 2019; Morais et al., 2018). This was further illustrated when we experimentally excluded information from nongenotyped individuals to increase the ratio of genotyped individuals (Figure 6). We found that as the number of nongenotyped individuals decreased, and in turn, the ratio of genotyped individuals increased, the accuracies of ssGBLUP roughly equaled those of GBLUP. The same trend was reported by Cappa et al. (2019), highlighting the benefits of additional phenotypic information.

With respect to CV2, an important advantage of ssGBLUP is that it can extend genomic information from genotyped individuals to nongentoyped relatives. Genomic information can provide two main benefits. First, it provides a more accurate estimate of relatedness between individuals as a proxy for the amount of genome segments shared, as opposed to what is expected based on pedigree relationships. Second, by eliminating the need for labor-intensive and time-consuming phenotyping of future selection candidates, it allows earlier selection decisions for traits that are difficult to measure. However, genotyping all selection candidates can still be prohibitively expensive, especially for small or fixed budget breeding programs, and particularly for crops with complex polyploid genomes that require large segregating populations to evaluate (Gemenet & Khan, 2017). In our dataset, the ssGBLUP and PBLUP models yielded rather similar accuracies. This result could be attributed to the small number of genotyped individuals relative to the nongenotyped (Sood et al., 2020). Indeed, it has been argued that lower proportions of genotyped individuals relative to the total population of phenotyped individuals may be associated with reduced performance of ssGBLUP compared to PBLUP and GBLUP (Song et al., 2019).

Overall, it is noteworthy that our results showed that ssGBLUP was at least as good as the best-performing model for most traits, regardless of the prediction scenario. This is in agreement with previous plant and animal breeding reports that investigated ssGBLUP and other tree improvement studies (Paludeto et al., 2021; Pérez-Rodríguez et al., 2017).

2.5.3. Practical considerations for sugarcane breeding programs

Our findings suggest that ssGBLUP can be an attractive predictive breeding approach for untested genotyped and nongenotyped selection candidates. As mentioned above, the PMGCA-UFV and most sugarcane breeding programs use family selection in early field trials. In our study, the units of test and selection were half-sib families. Theoretically, the additive genetic variance between and within half-sib families is one-quarter and three-quarters, respectively (Hallauer et al., 2010). Therefore, the response to selection is expected to be greater when selection is performed within families. In family trials, sugarcane breeders usually practice a combination of between-family and within-family selection to maximize genetic selection efficiency and ultimately response to selection. However, between-family selection is preferred because of the higher heritability at the family level for complex traits. Under these circumstances, conventional phenotypic selection without BLUP or using pedigree information (PBLUP) is likely to be sufficient to accurately predict the parental average component and rank families (H. Daetwyler et al., 2007). However, at the time of selection, phenotypic information of the progeny of individual selection candidates is not available, which precludes prediction of Mendelian sampling deviations and hinders within-family selection using BLUP or PBLUP analysis (H. Daetwyler et al., 2007; Hickey et al., 2014; Werner et al., 2020). Truncation selection schemes, such as BLUPIS, are designed to also weight the within-family genetic variance. However, this method is not optimal because the top-ranking families are subjected to mass selection (visual rating), which can be inefficient. This is because unreplicated plants tested in small plots and single environments are influenced by substantial genotype by environment interaction, which negatively affects the response to selection (Jackson & McRae, 2001). It can also potentially lead to the depletion of genetic diversity (Crossa et al., 2014; Hickey et al., 2014). In contrast, GBLUP can increase the efficiency of within-family selection by accurately predicting Mendelian deviations, regardless of the availability of phenotypic data from one's own selection candidates or their progeny (H. Daetwyler et al., 2007).

When family trials consist of full-sibs, it is reasonable to expect that the parental average, which is related to the between-family variance, and the Mendelian deviation, which is related to the within-family variance, are equally important factors influencing the breeding values of selection candidates. Regardless of the type of family used, ssGBLUP can be used to predict both components of the breeding value for all selection candidates (Pérez-Rodríguez et al., 2017). To fully benefit from the implementation of genomic prediction using the ssGBLUP approach will require greater control over pollination during hybridization including increased phenotyping efforts to better estimate marker effects and the development of ad hoc training and breeding populations. This may require large, genetically diverse, and thoroughly screened breeding materials. In addition, enhanced genotyping protocols are needed to cope with the complexity of the sugarcane genome, and to offset against high genotyping costs. Furthermore, our results highlight the advantage of exploiting pedigrees as a potential genetic information source, particularly for sugarcane and other crop species that exhibit complex polyploid genomes (Inamori et al., 2024).

2.5.4. Limitations and future directions

First, the CV prediction accuracies were based on a single generation, disregarding prediction scenarios with lower degrees of relatedness. We used the random sampling CV scheme, which gives equal weigh to within-family and between-family predictions (de Almeida Filho et al., 2019). However, datasets that allow the evaluation of forward validation scenarios should be considered in future investigations.

Second, we focused only on additive genetic effects. After hybridization, a forward selection scheme was applied. Since individual plants were vegetatively propagated all sources of genetic variance can be included. Therefore, considering only additive genetic effects may overlook the genetic merit of selection candidates. However, previous studies have reported that capturing non-additive genetic effects using machine learning only leads to an incremental increase in prediction in sugarcane (C. Chen et al., 2023; Inamori et al., 2024; Yadav, Wei, et al., 2021). Nevertheless, the mean cross-validation prediction accuracies we observed are in line with previous reports, indicating the robustness of pseudo-diploid additive-based prediction models (Ferrão et al., 2021). It would also be interesting to explore other sources of information beyond pedigrees and markers. Various omic technologies provide predictors that can also be used to infer relatedness between individuals (Bi et al., 2023).

Another potential advantage of the ssGBLUP is the ability to combine multi environment and unbalanced historical datasets, enabling the exploitation of genotype by environment interactions and connecting breeding programs. For example, Pérez-Rodríguez et al. (2017) used a multiplicative reaction norm model including the main and interaction effects of genotype and environment in combination with the ssGBLUP approach. Furthermore, Morais Junior et al. (2018) modeled genotype by environment interactions using multiplicative reaction norm models with high dimensional environmental covariates. The PMGCA-UFV is part of a nationwide network of sugarcane breeding programs (M. H. P. Barbosa et al., 2012; Cursi et al., 2022). Such a network could help leverage resources by effectively integrating all available pedigree, phenotypic, genotypic, and envirotypic data to predict the performance of untested sugarcane clones in different environments. However, when combining datasets, a potential shortcoming could be the quality of phenotypic data generated across selection stages due to different levels of experimental precision in replicated and unreplicated field trials (Hayes et al., 2021; Michel et al., 2020; Morais Junior et al., 2018; Sood et al., 2020). Nevertheless, Michel et al. (2020) reported advantages of the ssGBLUP when models were trained using genotyped individuals evaluated at advanced stages of selection in multi-environment trials and subsequently used to predict the performance of nongenotyped individuals in early preliminary yield trials. This highlights the importance of thorough phenotyping efforts to obtain accurate estimates of allele effects in genomic prediction models.

2.6. Conclusion

This study investigated the utility of combining pedigrees for genomic prediction in sugarcane. The ssGBLUP models showed relatively good performance compared to the baseline models for most traits, regardless of the prediction scenario evaluated. In conclusion, our study showed that ssGBLUP can be a convenient modeling approach for sugarcane breeding, as it provides the opportunity to integrate different sources of information, allowing for increased predictive power. Our results also suggest that the structure, size, and genetic composition of the training and breeding populations are key factors that need to be carefully considered by breeders for successful implementation of ssGBLUP.

2.7. References

Aguilar, I., Misztal, I., Johnson, D., Legarra, A., Tsuruta, S., & Lawlor, T. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of holstein final score. *Journal of dairy science*, *93*(2), 743–752.

Aitken, K. S., Jackson, P. A., & McIntyre, C. L. (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. https://doi.org/10.1007/s00122-006-0233-2

Albrecht, T., Auinger, H.-J., Wimmer, V., Ogutu, J. O., Knaak, C., Ouzunova, M., Piepho, H.-P., & Schön, C.-C. (2014). Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. *Theoretical and Applied Genetics*, *127*, 1375–1386.

Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A., Resende Jr, M. F., & Muñoz, P. R. (2016). Aghmatrix: R package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. *The plant genome*, *9*(3), plantgenome2016–01.

Antonios, S., Legarra, A., Pong-Wong, R., Astruc, J., Rodríguez-Ramilo, S., & Vitezica, Z. (2023). Partitioning of the genetic trends of french dairy sheep in mendelian samplings and long-term contributions. *Journal of Dairy Science*, *106*(9), 6275–6287. https://doi.org/10.3168/jds.2022-23009

Aono, A. H., Ferreira, R. C. U., Moraes, A. d. C. L., Lara, L. A. d. C., Pimenta, R. J. G., Costa, E. A., Pinto, L. R., Landell, M. G. d. A., Santos, M. F., Jank, L., et al. (2022). A joint learning approach for genomic prediction in polyploid grasses. *Scientific reports*, *12*(1), 12499.

Ashraf, B., Edriss, V., Akdemir, D., Autrique, E., Bonnett, D., Crossa, J., Janss, L., Singh, R., & Jannink, J.-L. (2016). Genomic prediction using phenotypes from pedigreed lines with no marker data. *Crop Science*, *56*(3), 957–964.

Atkin, F. C., Dieters, M. J., & Stringer, J. K. (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*, 555–565.

Auinger, H.-J., Schönleben, M., Lehermeier, C., Schmidt, M., Korzun, V., Geiger, H. H., Piepho, H.-P., Gordillo, A., Wilde, P., Bauer, E., et al. (2016). Model training across multiple breeding cycles significantly improves genomic prediction accuracy in rye (secale cereale l.) *Theoretical and applied genetics*, *129*, 2043–2053.

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

Barbosa, M., Peternelli, L., & Silveira, L. (2001). Plot size in sugarcane family selection experiments. *Cropp Breeding and Applied Biotechnology*, *1*(3), 271–276. https://doi.org/10.13082/1984-7033.v01n03a07

Barbosa, M., Resende, M., Bressiani, J., Silveira, L., & Peternelli, L. (2005). Selection of sugarcane families and parents by reml/blup. *Cropp Breeding and Applied Biotechnology*, *5*(4), 443–450. https://doi.org/10.12702/1984-7033.v05n04a10

Barreto, F. Z., Rosa, J. R. B. F., Balsalobre, T. W. A., Pastina, M. M., Silva, R. R., Hoffmann, H. P., de Souza, A. P., Garcia, A. A. F., & Carneiro, M. S. (2019). A genome-wide association study identified loci for yield component traits in sugarcane (saccharum spp.) (D. D. Fang, Ed.). *PLOS ONE*, *14*(7), e0219843. https://doi.org/10.1371/journal.pone.0219843

Batista, L. G., Mello, V. H., Souza, A. P., & Margarido, G. R. (2021). Genomic prediction with allele dosage information in highly polyploid species. *Theoretical and Applied Genetics*, 1–17.

Beaulieu, J., Doerksen, T., Clément, S., MacKay, J., & Bousquet, J. (2014). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity*, *113*(4), 343–352. https://doi.org/10.1038/hdy.2014.36

Beaulieu, J., Lenz, P., & Bousquet, J. (2022). Metadata analysis indicates biased estimation of genetic parameters and gains using conventional pedigree information instead of genomic-based approaches in tree breeding. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-06681-y

Bezerra, T. L., & Ragauskas, A. J. (2016). A review of sugarcane bagasse for second-generation bioethanol and biopower production. *Biofuels, Bioproducts and Biorefining*, *10*(5), 634–647.

Bi, Y., Yassue, R. M., Paul, P., Dhatt, B. K., Sandhu, J., Do, P. T., Walia, H., Obata, T., & Morota, G. (2023). Evaluating metabolic and genomic data for predicting grain traits under high night temperature stress in rice. *G3: Genes, Genomes, Genetics*, *13*(5), jkad052.

Bijma, P., Wientjes, Y. C. J., & Calus, M. P. L. (2020). Breeding top genotypes and accelerating response to recurrent selection by selecting parents with greater gametic variance. *Genetics*, *214*(1), 91–107. https://doi.org/10.1534/genetics.119.302643

Bonk, S., Reichelt, M., Teuscher, F., Segelke, D., & Reinsch, N. (2016). Mendelian sampling covariability of marker effects and genetic values. *Genetics Selection Evolution*, *48*(1). https://doi.org/10.1186/s12711-016-0214-0

Bourke, P. M., Voorrips, R. E., Visser, R. G., & Maliepaard, C. (2018). Tools for genetic studies in experimental populations of polyploids. *Frontiers in plant science*, *9*, 513.

Brasileiro, B. P., de Paula Mendes, T. O., Peternelli, L. A., da Silveira, L. C. I., de Resende, M. D. V., & Barbosa, M. H. P. (2016). Simulated individual best linear unbiased prediction versus mass selection in sugarcane families. *Crop Science*, *56*(2), 570–575. https://doi.org/10.2135/cropsci2015.03.0199

Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. *Crop Science*, *52*(2), 707–719.

Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Garcia, C. C., Mansfield, S. D., & Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in eucalyptus using phenotypes from non-genotyped trees by single-step gblup. *Plant Science*, *284*, 9–15.

Carvalho-Netto, O. V., Bressiani, J. A., Soriano, H. L., Fiori, C. S., Santos, J. M., Barbosa, G. V., Xavier, M. A., Landell, M. G., & Pereira, G. A. (2014). The potential of the energy cane as the main biomass crop for the cellulosic industry. *Chemical and Biological Technologies in Agriculture*, *1*(1), 1–8.

Chen, C., Powell, O., Dinglasan, E., Ross, E. M., Yadav, S., Wei, X., Atkin, F., Deomano, E., & Hayes, B. J. (2023). Genomic prediction with machine learning in sugarcane, a complex highly polyploid clonally propagated crop with substantial non-additive variation for key traits. *The Plant Genome*, e20390.

Chen, C.-Y., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. (2011). Effect of different genomic relationship matrices on accuracy and scale. *Journal of animal science*, *89*(9), 2673–2679.

Chen, Z.-Q., Baison, J., Pan, J., Karlsson, B., Andersson, B., Westin, J., García-Gil, M. R., & Wu, H. X. (2018). Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in norway spruce. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-5256-y

Christensen, O., Madsen, P., Nielsen, B., Ostersen, T., & Su, G. (2012). Single-step methods for genomic evaluation in pigs. *animal*, *6*(10), 1565–1571.

Christensen, O. F., & Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution*, *42*, 1–8.

Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, *6*(11), 836–846. https://doi.org/10.1038/nrg1711

Costa, P., Almeida, C., Silveira, G., Soares, B., Baffa, D., Peternelli, L., Bhering, L., & Barbosa, M. (2014). Selfing confirmation in sugarcane by using simple sequence repeat markers: An individual reciprocal recurrent selection scheme. *Genetics* *and Molecular Research*, *13*(4), 8962–8970. https://doi.org/10.4238/2014.october.31.11

Cros, D., Denis, M., Sánchez, L., Cochard, B., Flori, A., Durand-Gasselin, T., Nouy, B., Omoré, A., Pomiès, V., Riou, V., Suryana, E., & Bouvet, J.-M. (2014). Genomic selection prediction accuracy in a perennial crop: Case study of oil palm (elaeis guineensis jacq.) *Theoretical and Applied Genetics*, *128*(3), 397–410. https://doi.org/10.1007/s00122-014-2439-z

Crossa, J., Perez, P., Hickey, J., Burgueño, J., Ornella, L., Cerón-Rojas, J., Zhang, X., Dreisigacker, S., Babu, R., Li, Y., et al. (2014). Genomic prediction in cimmyt maize and wheat breeding programs. *Heredity*, *112*(1), 48–60.

Crossa, J., Campos, G. d. l., Peérez, P., Gianola, D., Burgueño, J., Araus, J. L., Makumbi, D., Singh, R. P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., & Braun, H.-J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, *186*(2), 713–724. https://doi.org/10.1534/genetics.110.118521

Cursi, D. E., Cox, M. C., de Oliveira Anoni, C., Hoffmann, H. P., Gazaffi, R., & Garcia, A. A. F. (2020). Comparison of different selection methods in the seedling stage of sugarcane breeding. *Agronomy Journal*, *112*(6), 4879–4897. https://doi.org/10.1002/agj2.20431

Cursi, D. E., Hoffmann, H. P., Barbosa, G., Bressiani, J. A., Gazaffi, R., Chapola, R. G., Fernandes Junior, A., Balsalobre, T. W. A., Diniz, C. A., Santos, J. M., et al. (2022). History and current status of sugarcane breeding, germplasm development and molecular genetics in brazil. *Sugar Tech*, *24*(1), 112–133.

Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome wide approach. *PloS one*, *3*(10), e3395.

Daetwyler, H., Villanueva, B., Bijma, P., & Woolliams, J. (2007). Inbreeding in genome□wide selection. *Journal of Animal Breeding and Genetics*, *124*(6), 369–376. https://doi.org/10.1111/j.1439-0388.2007.00693.x

de Almeida Filho, J. E., Guimarães, J. F. R., Fonseca e Silva, F., Vilela de Resende, M. D., Muñoz, P., Kirst, M., & de Resende Júnior, M. F. R. (2019). Genomic prediction of additive and non-additive effects using genetic markers and pedigrees. *G3: Genes, Genomes, Genetics*, *9*(8), 2739–2748.

de Bem Oliveira, I., Amadeu, R. R., Ferrão, L. F. V., & Muñoz, P. R. (2020). Optimizing whole-genomic prediction for autotetraploid blueberry breeding. *Heredity*, *125*(6), 437–448.

de Bem Oliveira, I., Resende, M. F. R., Ferrão, L. F. V., Amadeu, R. R., Endelman, J. B., Kirst, M., Coelho, A. S. G., & Munoz, P. R. (2019). Genomic prediction of autotetraploids; influence of relationship matrices, allele dosage, and continuous genotyping calls in phenotype prediction. *G3 Genes|Genomes|Genetics*, *9*(4), 1189–1198. https://doi.org/10.1534/g3.119.400059

de C. Lara, L. A., Santos, M. F., Jank, L., Chiari, L., Vilela, M. d. M., Amadeu, R. R., dos Santos, J. P. R., Pereira, G. d. S., Zeng, Z.-B., & Garcia, A. A. F. (2019). Genomic selection with allele dosage in panicum maximum jacq. *G3 Genes Genomes Genetics*, *9*(8), 2463–2475. https://doi.org/10.1534/g3.118.200986

de Oliveira, A. A., Resende Jr, M. F., Ferrão, L. F. V., Amadeu, R. R., Guimarães, L. J. M., Guimarães, C. T., Pastina, M. M., & Margarido, G. R. A. (2020). Genomic prediction applied to multiple traits and environments in second season maize hybrids. *Heredity*, *125*(1-2), 60–72.

Deomano, E., Jackson, P., Wei, X., Aitken, K., Kota, R., & Pérez-Rodríguez, P. (2020). Genomic prediction of sugar content and cane yield in sugar cane clones in different stages of selection in a breeding program, with and without pedigree information. *Molecular Breeding*, *40*, 1–12.

Endelman, J. B., Carley, C. A. S., Bethke, P. C., Coombs, J. J., Clough, M. E., da Silva, W. L., De Jong, W. S., Douches, D. S., Frederick, C. M., Haynes, K. G., et al. (2018). Genetic variance partitioning and genome-wide prediction with allele dosage information in autotetraploid potato. *Genetics*, *209*(1), 77–87.

Ferrão, L. F. V., Amadeu, R. R., Benevenuto, J., de Bem Oliveira, I., & Munoz, P. R. (2021). Genomic selection in an outcrossing autotetraploid fruit crop: Lessons from blueberry breeding. *Frontiers in plant science*, *12*, 676326.

Fickett, N. D., Ebrahimi, L., Parco, A. P., Gutierrez, A. V., Hale, A. L., Pontif, M. J., Todd, J., Kimbeng, C. A., Hoy, J. W., Ayala-Silva, T., Gravois, K. A., & Baisakh, N. (2020). An enriched sugarcane diversity panel for utilization in genetic improvement of sugarcane. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-70292-8

Fugeray-Scarbel, A., Bastien, C., Dupont-Nivet, M., Lemarié, S., Consortium, R., et al. (2021). Why and how to switch to genomic selection: Lessons from plant and animal breeding experience. *Frontiers in Genetics*, *12*, 1185.

Garcia, A. A., Mollinari, M., Marconi, T. G., Serang, O. R., Silva, R. R., Vieira, M. L., Vicentini, R., Costa, E. A., Mancini, M. C., Garcia, M. O., et al. (2013). Snp genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Scientific reports*, *3*(1), 3399.

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., Jenkins, J., Martin, G., Charron, C., Hervouet, C., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature communications*, *9*(1), 2638.

Gemenet, D. C., & Khan, A. (2017). Opportunities and challenges to implementing genomic selection in clonally propagated crops. *Genomic Selection for Crop Im-*

provement: New Molecular Breeding Strategies for Crop Improvement, 185–198.

Gemenet, D. C., Lindqvist-Kreuze, H., De Boeck, B., da Silva Pereira, G., Mollinari, M., Zeng, Z.-B., Craig Yencho, G., & Campos, H. (2020). Sequencing depth and genotype quality: Accuracy and breeding operation considerations for genomic selection applications in autopolyploid crops. *Theoretical and Applied Genetics*, *133*, 3345–3363.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics*, *4*, 641–649.

Goddard, M., & Hayes, B. (2007). Genomic selection. *Journal of Animal breeding and Genetics*, *124*(6), 323–330.

Gonçalves, M. T. V., Morota, G., Costa, P. M. d. A., Vidigal, P. M. P., Barbosa, M. H. P., & Peternelli, L. A. (2021). Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. *PLoS One*, *16*(3), e0236853.

Gouy, M., Rousselle, Y., Bastianelli, D., Lecomte, P., Bonnal, L., Roques, D., Efile, J.-C., Rocher, S., Daugrois, J., Toubi, L., et al. (2013). Experimental assessment of the accuracy of genomic selection in sugarcane. *Theoretical and applied genetics*, *126*, 2575–2586.

Hale, A. L., Todd, J. R., Gravois, K. A., Mollov, D., Malapi-Wight, M., Momotaz, A., Laborde, C., Goenaga, R., Kimbeng, C., Solis, A., et al. (2022). Sugarcane breeding programs in the usa. *Sugar Tech*, *24*(1), 97–111.

Hallauer, A. R., Carena, M. J., & Miranda Filho, J. d. (2010). *Quantitative genetics in maize breeding* (Vol. 6). Springer Science & Business Media.

Hayes, B. J., Wei, X., Joyce, P., Atkin, F., Deomano, E., Yue, J., Nguyen, L., Ross, E. M., Cavallaro, T., Aitken, K. S., et al. (2021). Accuracy of genomic prediction of complex traits in sugarcane. *Theoretical and Applied Genetics*, *134*, 1455–1462.

Heffner, E. L., Sorrells, M. E., & Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Science*, *49*(1), 1–12.

Henderson, C. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, 69–83.

Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B. M., Grondona, M., Zambelli, A., Windhausen, V. S., Mathews, K., et al. (2014). Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science*, *54*(4), 1476–1488.

Imai, A., Kuniga, T., Yoshioka, T., Nonaka, K., Mitani, N., Fukamachi, H., Hiehata, N., Yamamoto, M., & Hayashi, T. (2019). Single-step genomic prediction of fruit-quality traits using phenotypic records of non-genotyped relatives in citrus. *PLoS One*, *14*(8), e0221880.

Inamori, M., Kimura, T., Mori, M., Tarumoto, Y., Hattori, T., Hayano, M., Umeda, M., & Iwata, H. (2024). Machine learning for genomic and pedigree prediction in sugarcane. *The Plant Genome*, e20486.

Islam, M. S., McCord, P., Read, Q. D., Qin, L., Lipka, A. E., Sood, S., Todd, J., & Olatoye, M. (2022). Accuracy of genomic prediction of yield and sugar traits in saccharum spp. hybrids. *Agriculture*, *12*(9), 1436.

Islam, M. S., McCord, P. H., Olatoye, M. O., Qin, L., Sood, S., Lipka, A. E., & Todd, J. R. (2021). Experimental evaluation of genomic selection prediction for rust resistance in sugarcane. *The Plant Genome*, *14*(3), e20148.

Islam, M. S., Corak, K., McCord, P., Hulse-Kemp, A. M., & Lipka, A. E. (2023). A first look at the ability to use genomic prediction for improving the ratooning ability of sugarcane. *Frontiers in Plant Science*, *14*.

Jackson, P., & McRae, T. (2001). Selection of sugarcane clones in small plots: Effects of plot size and selection criteria. *Crop Science*, *41*(2), 315–322. https://doi.org/10.2135/cropsci2001.412315x

Juliana, P., Singh, R. P., Poland, J., Mondal, S., Crossa, J., Montesinos-López, O. A., Dreisigacker, S., Pérez-Rodríguez, P., Huerta-Espino, J., Crespo-Herrera, L., & Govindan, V. (2018). Prospects and challenges of applied genomic selection—a new paradigm in breeding for grain yield in bread wheat. *The Plant Genome*, *11*(3). https://doi.org/10.3835/plantgenome2018.03.0017

Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Rutkoski, J. E., Poland, J. A., Bergstrom, G. C., & Sorrells, M. E. (2017). Comparison of models and whole genome profiling approaches for genomic enabled prediction of septoria tritici blotch, stagonospora nodorum blotch, and tan spot resistance in wheat. *The Plant Genome*, *10*(2). https://doi.org/10.3835/plantgenome2016.08.0082

Klápště, J., Dungey, H. S., Telfer, E. J., Suontama, M., Graham, N. J., Li, Y., & McKinley, R. (2020). Marker selection in multivariate genomic prediction improves accuracy of low heritability traits. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.499094

Lauer, E., Holland, J., & Isik, F. (2021). Prediction ability of genome-wide markers in pinus taeda l. within and between population is affected by relatedness to the training population and trait genetic architecture (A. Lipka, Ed.). *G3 Genes Genomes Genetics*, *12*(2). https://doi.org/10.1093/g3journal/jkab405

Legarra, A., Aguilar, I., & Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *Journal of dairy science*, *92*(9), 4656–4663.

Lopes, M. L., Paulillo, S. C. d. L., Godoy, A., Cherubin, R. A., Lorenzi, M. S., Giometti, F. H. C., Bernardino, C. D., Amorim Neto, H. B. d., & Amorim, H. V. d. (2016). Ethanol production in brazil: A bridge between science and industry. *brazilian journal of microbiology*, *47*, 64–76.

Marjanovic, J., & Calus, M. P. L. (2020). Factors affecting accuracy of estimated effective number of chromosome segments for numerically small breeds. *Journal of Animal Breeding and Genetics*, *138*(2), 151–160. https://doi.org/10.1111/jbg.12512

Matsuoka, S., Kennedy, A. J., Santos, E. G. D. d., Tomazela, A. L., & Rubio, L. C. S. (2014). Energy cane: Its concept, development, characteristics, and prospects. *Advances in Botany*, *2014*.

Mbuma, N. W., Zhou, M. M., & van der Merwe, R. (2020). Family × environment interaction for sugarcane yield and its components in south africa. *Crop Science*, *60*(3), 1239–1253. https://doi.org/10.1002/csc2.20025

Mertes, F., ElSharawy, A., Sauer, S., van Helvoort, J. M., Van Der Zaag, P., Franke, A., Nilsson, M., Lehrach, H., & Brookes, A. J. (2011). Targeted enrichment of genomic dna regions for next-generation sequencing. *Briefings in functional genomics*, *10*(6), 374–386.

Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *genetics*, *157*(4), 1819–1829.

Michel, S., Löschenberger, F., Sparry, E., Ametz, C., & Bürstmayr, H. (2020). Multiyear dynamics of single-step genomic prediction in an applied wheat breeding program. *Agronomy*, *10*(10), 1591.

Misztal, I., Aggrey, S. E., & Muir, W. M. (2013). Experiences with a single-step genome evaluation. *Poultry science*, *92*(9), 2530–2534.

Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of dairy science*, *92*(9), 4648–4655.

Morais, O. P., Breseghello, F., Duarte, J. B., Coelho, A. S., Borba, T. C., Aguiar, J. T., Neves, P. C., & Morais, O. P. (2018). Assessing prediction models for different traits in a rice population derived from a recurrent selection program. *Crop Science*, *58*(6), 2347–2359.

Morais Junior, O. P., Duarte, J. B., Breseghello, F., Coelho, A. S., Morais, O. P., & Magalhaes Junior, A. M. (2018). Single-step reaction norm models for genomic prediction in multi-environment recurrent selection trials. *Crop Science*, *58*(2), 592–607.

Morota, G., & Gianola, D. (2014). Kernel-based whole-genome prediction of complex traits: A review. *Frontiers in genetics*, *5*, 363.

Nadeau, S., Beaulieu, J., Gezan, S. A., Perron, M., Bousquet, J., & Lenz, P. R. N. (2023). Increasing genomic prediction accuracy for unphenotyped full-sib families by modeling additive and dominance effects with large datasets in white spruce. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1137834

Neves, L. G., Davis, J. M., Barbazuk, W. B., & Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal*, *75*(1), 146–156.

Nyine, M., Uwimana, B., Blavet, N., Hřibová, E., Vanrespaille, H., Batte, M., Akech, V., Brown, A., Lorenzen, J., Swennen, R., & Doležel, J. (2018). Genomic prediction in a multiploid crop: Genotype by environment interaction and allele dosage effects on predictive ability in banana. *The Plant Genome*, *11*(2). https://doi.org/10.3835/plantgenome2017.10.0090

O'Connell, A., Deo, J., Deomano, E., Wei, X., Jackson, P., Aitken, K. S., Manimekalai, R., Mohanraj, K., Hemaprabha, G., Ram, B., et al. (2022). Combining genomic selection with genome-wide association analysis identified a large-effect qtl and improved selection for red rot resistance in sugarcane. *Frontiers in Plant Science*, *13*, 1021182.

Paludeto, J. G. Z., Grattapaglia, D., Estopa, R. A., & Tambarussi, E. V. (2021). Genomic relationship–based genetic parameters and prospects of genomic selection for growth and wood quality traits in eucalyptus benthamii. *Tree Genetics & Genomes*, *17*(4), 38.

Pérez, P., & de Los Campos, G. (2014). Genome-wide regression and prediction with the bglr statistical package. *Genetics*, *198*(2), 483–495.

Pérez-Rodríguez, P., Crossa, J., Rutkoski, J., Poland, J., Singh, R., Legarra, A., Autrique, E., Campos, G. d. l., Burgueño, J., & Dreisigacker, S. (2017). Single-step genomic and pedigree genotype× environment interaction models for predicting wheat lines in international environments. *The plant genome*, *10*(2).

Peternelli, L. A., & de Resende, M. D. V. (2015). Experimental designs for next generation phenotyping. In *Phenomics* (pp. 15–32). Springer International Publishing. https://doi.org/10.1007/978-3-319-13677-6_2

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

Pisaroglo de Carvalho, M., Gezan, S. A., Peternelli, L. A., & Pereira Barbosa, M. H. (2014). Estimation of additive and nonadditive genetic components of sugarcane families using multitrait analysis. *Agronomy Journal*, *106*(3), 800–808. https://doi.org/10.2134/agronj2013.0247

Plummer, M., Best, N., Cowles, K., Vines, K., et al. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R news*, *6*(1), 7–11.

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

Ratcliffe, B., El-Dien, O. G., Cappa, E. P., Porth, I., Klápště, J., Chen, C., & El-Kassaby, Y. A. (2017). Single-step blup with varying genotyping effort in open-pollinated picea glauca. *G3: Genes, Genomes, Genetics*, *7*(3), 935–942.

Resende, M. D. V. d. (2016). Software selegen-reml/blup: A useful tool for plant breeding. *Crop Breeding and Applied Biotechnology*, *16*, 330–339.

Resende, M. D. V. d., & Barbosa, M. H. P. (2006). Selection via simulated individual blup based on family genotypic effects in sugarcane. *Pesquisa Agropecuária Brasileira*, *41*(3), 421–429. https://doi.org/10.1590/s0100-204x2006000300008

Satpathy, S., Shahi, D., Blanchard, B., Pontif, M., Gravois, K., Kimbeng, C., Hale, A., Todd, J., Rao, A., & Baisakh, N. (2022). Evaluation of models for utilization in genomic prediction of agronomic traits in the louisiana sugarcane breeding program. *Agriculture*, *12*(9), 1330.

Schiessl, S.-V., Katche, E., Ihien, E., Chawla, H. S., & Mason, A. S. (2019). The role of genomic structural variation in the genetic improvement of polyploid crops. *The Crop Journal*, *7*(2), 127–140.

Silva, F. L. d., Barbosa, M. H. P., de Resende, M. D. V., Peternelli, L. A., & Pedrozo, C. u. (2015). Efficiency of selection within sugarcane families via simulated individual blup. *Crop Breeding and Applied Biotechnology*, *15*(1), 1–9. https://doi.org/10.1590/1984-70332015v15n1a1

Silva, M. d. A., & Gonçalves, P. d. S. (2011). Inbreeding in sugarcane varieties. *Ciência Rural*, *41*(4), 580–586. https://doi.org/10.1590/s0103-84782011005000028

Simiqueli, G. F., Resende, R. T., Takahashi, E. K., de Sousa, J. E., & Grattapaglia, D. (2023). Realized genomic selection across generations in a reciprocal recurrent selection breeding program of eucalyptus hybrids. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1252504

Song, H., Zhang, J., Zhang, Q., & Ding, X. (2019). Using different single-step strategies to improve the efficiency of genomic prediction on body measurement traits in pig. *Frontiers in genetics*, *9*, 730.

Sood, S., Lin, Z., Caruana, B., Slater, A. T., & Daetwyler, H. D. (2020). Making the most of all data: Combining non-genotyped and genotyped potato individuals with hblup. *The Plant Genome*, *13*(3), e20056.

Tew, T. L., & Pan, Y.-B. (2010). Microsatellite (simple sequence repeat) marker–based paternity analysis of a seven□parent sugarcane polycross. *Crop Science*, *50*(4), 1401–1408. https://doi.org/10.2135/cropsci2009.10.0579

Thavamanikumar, S., Arnold, R. J., Luo, J., & Thumma, B. R. (2020). Genomic studies reveal substantial dominant effects and improved genomic predictions in an open-pollinated breeding population of eucalyptus pellita. *G3: Genes, Genomes, Genetics*, *10*(10), 3751–3763.

Thirugnanasambandam, P. P., Hoang, N. V., & Henry, R. J. (2018). The challenge of analyzing the sugarcane genome. *Frontiers in plant science*, *9*, 616.

Todd, J., Glaz, B., Burner, D., & Kimbeng, C. (2015). Historical use of cultivars as parents in florida and louisiana sugarcane breeding programs. *International Scholarly Research Notices*, *2015*, 1–9. https://doi.org/10.1155/2015/257417

Todd, J., Pan, Y.-B., & Boykin, D. (2020). Fidelity of sugarcane crosses assessed with ssr markers. *Agronomy*, *10*(3), 386. https://doi.org/10.3390/agronomy10030386

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, *91*(11), 4414–4423.

Vela-Avitúa, S., Meuwissen, T., Luan, T., & Ødegård, J. (2015). Accuracy of genomic selection for a sib-evaluated trait using identity-by-state and identity-by-descent relationships. *Genetics Selection Evolution*, *47*(1), 9. https://doi.org/10.1186/s12711-014-0084-2

Velazco, J. G., Malosetti, M., Hunt, C. H., Mace, E. S., Jordan, D. R., & Van Eeuwijk, F. A. (2019). Combining pedigree and genomic information to improve prediction quality: An example in sorghum. *Theoretical and Applied Genetics*, *132*, 2055–2067.

Vencovsky, R. (1978). Effective size of monoecious populations submitted to artificial selection. *Brazilian journal of genetics*, *1*(3), 181–191.

Vencovsky, R., & Crossa, J. (2003). Measurements of representativeness used in genetic resources conservation and plant breeding. *Crop Science*, *43*(6), 1912–1921. https://doi.org/10.2135/cropsci2003.1912

Vitezica, Z., Aguilar, I., Misztal, I., & Legarra, A. (2011). Bias in genomic predictions for populations under selection. *Genetics Research*, *93*(5), 357–366.

Voss-Fels, K. P., Wei, X., Ross, E. M., Frisch, M., Aitken, K. S., Cooper, M., & Hayes, B. J. (2021). Strategies and considerations for implementing genomic selection to improve traits with additive and non-additive genetic architectures in sugarcane breeding. *Theoretical and Applied Genetics*, *134*, 1493–1511.

Walker, T. D., Cumbie, W. P., & Isik, F. (2022). Single-step genomic analysis increases the accuracy of within-family selection in a clonally replicated population of pinus taeda l. *Forest Science*, *68*(1), 37–52.

Walsh, B., & Lynch, M. (2018). *Evolution and selection of quantitative traits*. Oxford University Press.

Wang, L., Yang, J., Zhang, H., Tao, Q., Zhang, Y., Dang, Z., Zhang, F., & Luo, Z. (2022). Sequence coverage required for accurate genotyping by sequencing in polyploid species. *Molecular ecology resources*, *22*(4), 1417–1426.

Weber, S. E., Chawla, H. S., Ehrig, L., Hickey, L. T., Frisch, M., & Snowdon, R. J. (2023). Accurate prediction of quantitative traits with failed snp calls in canola and maize. *Frontiers in Plant Science*, *14*.

Werner, C. R., Gaynor, R. C., Gorjanc, G., Hickey, J. M., Kox, T., Abbadi, A., Leckband, G., Snowdon, R. J., & Stahl, A. (2020). How population structure impacts genomic selection accuracy in cross-validation: Implications for practical breeding. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.592977

Wientjes, Y. C., Veerkamp, R. F., & Calus, M. P. (2013). The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. *Genetics*, *193*(2), 621–631.

Yadav, S., Ross, E. M., Aitken, K. S., Hickey, L. T., Powell, O., Wei, X., Voss-Fels, K. P., & Hayes, B. J. (2021). A linkage disequilibrium-based approach to position unmapped snps in crop species. *BMC genomics*, *22*(1), 1–9.

Yadav, S., Ross, E. M., Wei, X., Liu, S., Nguyen, L. T., Powell, O., Hickey, L. T., Deomano, E., Atkin, F., Voss-Fels, K. P., et al. (2023). Use of continuous genotypes for genomic prediction in sugarcane. *The Plant Genome*, e20417.

Yadav, S., Wei, X., Joyce, P., Atkin, F., Deomano, E., Sun, Y., Nguyen, L. T., Ross, E. M., Cavallaro, T., Aitken, K. S., et al. (2021). Improved genomic prediction of clonal performance in sugarcane by exploiting non-additive genetic effects. *Theoretical and Applied Genetics*, *134*(7), 2235–2252.

Yang, X., Luo, Z., Todd, J., Sood, S., & Wang, J. (2020). Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (saccharum spp.) *The Plant Genome*, *13*(1), e20006.

Yang, X., Song, J., Todd, J., Peng, Z., Paudel, D., Luo, Z., Ma, X., You, Q., Hanson, E., Zhao, Z., et al. (2019). Target enrichment sequencing of 307 germplasm accessions identified ancestry of ancient and modern hybrids and signatures of adaptation and selection in sugarcane (saccharum spp.), a ‘sweet’crop with ‘bitter’genomes. *Plant Biotechnology Journal*, *17*(2), 488–498.

Yang, X., Song, J., Todd, J., Peng, Z., Paudel, D., Luo, Z., Ma, X., You, Q., Hanson, E., Zhao, Z., Zhao, Y., Zhang, J., Ming, R., & Wang, J. (2018). Target enrichment sequencing of 307 germplasm accessions identified ancestry of ancient and modern hybrids and signatures of adaptation and selection in sugarcane (saccharum spp.), a ‘sweet’ crop with ‘bitter’ genomes. *Plant Biotechnology Journal*, *17*(2), 488–498. https://doi.org/10.1111/pbi.12992

Zan, F., Zhang, Y., Wu, Z., Zhao, J., Wu, C., Zhao, Y., Chen, X., Zhao, L., Qin, W., Yao, L., Xia, H., Zhao, P., Yang, K., Liu, J., & Yang, X. (2020). Genetic analysis of agronomic traits in elite sugarcane (saccharum spp.) germplasm (N. Baisakh, Ed.). *PLOS ONE*, *15*(6), e0233752. https://doi.org/10.1371/journal.pone.0233752

3. Chapter 3

Sugarcane biomass yield prediction using high-throughput RGB image-extracted traits and multi-trait pedigree, and genomic prediction models

3.1. Abstract

In sugarcane breeding, the evaluation of tonnes of cane per hectare (TCH) is costly, time consuming and labor-intensive. Statistical methods that incorporate pedigree and genomic information can be particularly useful to make selection decisions for TCH. Unoccupied aerial systems (UAS) enable breeders to collect traits that are correlated with difficult to measure traits, such as TCH. The integration of pedigree and genomic information with UAS may further improve selection accuracy for TCH. Here our objectives were to estimate (i) genetic parameters and trends for TCH and Red-Green-Blue (RGB)-image extracted traits, to compare (ii) the performance of single-trait with multi-trait genomic and pedigree prediction models that incorporate RGB-image extracted traits. The performance of models was evaluated in terms of Pearson's correlation between adjusted phenotypes, and mean squared error (MSE) using three cross-validation schemes, which varied in the level of phenotypic information available: ST, where clones had been evaluated only for TCH; MT-1, where phenotypic data of secondary traits were combined with those of TCH in multi-trait models, and MT-2 where phenotypic data of secondary traits were available in both, training and testing sets. We used data of an augmented block design trial, consisting of 385 clones. Clones were phenotyped at the second ratoon stage for TCH, and for 12 RGB-image extracted traits collected in a single flight. In general, we found low genetic correlation between TCH and RGB-image extracted traits, and moderate narrow-sense heritability estimates for RGB-image extracted. Overall indirect response to selection of RGB-image extracted was higher compared to direct response to selection for TCH. Cross-validation prediction accuracies and MSE of single- and multi-trait models were very low and similar, regardless of the pair-wise combination of secondary and the primary trait and the cross-validation scheme considered. Our results suggest that accuracies of multi-trait models that incorporated RGB-image extracted traits did not improve compared to single-trait models for predicting TCH. Future research should investigate alternative sensor technologies and optimize UAS-based data collection.

3.2. Background

A key objective in sugarcane breeding is to increase cane (biomass) yield. There is evidence that gains from selection of cane yield are associated with increases in sucrose and fiber yields, which are also major economically important traits (Acreche et al., 2015; da Silveira et al., 2015; Jackson, 2005). However, phenotyping selection candidates for cane yield potential is challenging. Cane yield is a complex trait that shows quantitative genetic control, and hence, it is expected to be highly affected by environmental effects.

The task escalates at early selection stages due to the large number of selection candidates and the lack of propagation material, which prevents the implementation of replicated trials and larger experimental units, further contributing to decrease heritability estimates of cane yield potential (Yadav et al., 2020). Also, due to budget constraints, mechanical harvesters fitted with weighing machines are not always available to weigh whole plots (Zhou, 2014). To circumvent these issues, cane yield potential can be indirectly estimated based on data of yield-related traits such as stalk number, stalk weight, stalk height and stalk diameter sampled from family or clonal plots (Cursi et al., 2022; da Silveira et al., 2015; Zhou, 2014). Nevertheless, these manual measurements are cumbersome and prone to human error, lacking throughput and robustness (Araus et al., 2018). Moreover, estimated TCH using data of sampled yield-related traits may underestimate actual TCH of whole plots, potentially leading to the discard of superior genotypes (da Silveira et al., 2015). Collectively, the evaluation of cane yield potential is costly, time consuming and labor-intensive.

In this context, unoccupied aerial systems (UAS) such as image-based high-throughput phenotyping (HTP) tools integrated with unoccupied aerial vehicules (UAV) can be an attractive alternative for field phenotyping with plot-level resolution of extensive trials (Herr et al., 2023; Tattaris et al., 2016). In sugarcane, UAS-based imaging is being used to both measure and predict geometric (De Souza et al., 2017), physiological (Hoffman et al., 2024; Natarajan et al., 2019), quality (Barbosa Júnior et al., 2023; Chea et al., 2020; Cholula et al., 2020), agronomic yield-related (de Oliveira et al., 2022; Todd

& Johnson, 2021), and canopy spectral traits (Akbarian et al., 2023; Khuimphukhieo, Bhandari, et al., 2024). In particular, UAS based on Red-Green-Blue (RGB) imaging using standard digital cameras is a widely employed approach with many applications in plant breeding research (Araus et al., 2018; Luo et al., 2023). RGB imaging has advantages compared to other sensors including low cost, light weight, and high resolution (Araus et al., 2018; Yang et al., 2017). Indeed, a growing body of research supports the feasibility of estimating cane yield potential using RGB image-extracted traits (Cholula et al., 2020; Huang et al., 2024; Khuimphukhieo, da Silva, et al., 2024; Khuimphukhieo et al., 2023; Sanches et al., 2018; Som-ard et al., 2018; Sumesh et al., 2021; Todd et al., 2022).

Phenotypic data collected with HTP tools can be used to efficiently augment the information available of each selection candidate, and have mainly been incorporated into statistical models as response or predictor variables (Morota et al., 2022). In sugarcane, the majority of the UAS-based imaging studies used these HTP traits as predictors by using indirect selection (for example, Cholula et al., 2020) or index selection (Natarajan et al., 2019). Typically, these methods were employed within the framework of linear regression or machine learning models, which provide a streamlined modeling approach. However, other modeling strategies have been proposed. For example, Rutkoski et al., 2016 investigated the use of canopy spectral traits as response variables to improve accuracy in multi-trait pedigree and genomic prediction models for grain yield in wheat. Also, Mróz et al., 2024 utilized canopy spectral traits, including RGB bands, to build spectral reflectance-derived relationship matrices to exploit genotypic similarities among selection candidates to predict grain yield in spring wheat. Moreover, most of the available UAS-based imaging studies in sugarcane focused in predicting cane yield potential in large commercial fields, or in late selections stages with larger plots using few varieties.

Therefore, in the present study we evaluated the integration of UAS-based RGB imaging with pedigree and genomic prediction models to improve selection accuracy and efficiency in the development of superior sugarcane varieties. We used experimental data of small unreplicated plots from an early selection stage, and our specific objectives were: to estimate (i) genetic parameters and trends for TCH and RGB-image extracted traits, to compare (ii) the performance of single-trait with multi-trait genomic and pedigree prediction models that incorporate RGB-image extracted traits, and (iii) to provide knowledge on the application of UAS-based RGB imaging in sugarcane breeding programs.

3.3. Material and methods

3.3.1. Plant material and field trial

We evaluated a total of 385 clones selected from a original seedling population constituted of 98 half-sibs progeny previously established in a family trial, which is part of the sugarcane genetic breeding program conducted by the Universidade Federal de Viçosa (PMGCA). The field trial was planted in 2016 using an augmented randomized block design with 21 blocks, and three check varieties (RB867515, RB966928, and RB92579) replicated in each block. Experimental units consisted of two-row 3 m long furrows with 1.4 m inter-row spacing plots. The trial was grown following standard commercial agronomic practices during a 3-year crop cycle (plant cane, first and second ratoon) at the Sugarcane Genetic Breeding Research Station (CECA) of the PMGCA (20°25’ S, 42°48’ W, 494 m altitude), Oratórios, MG, Brazil.

3.3.2. Phenotypic data

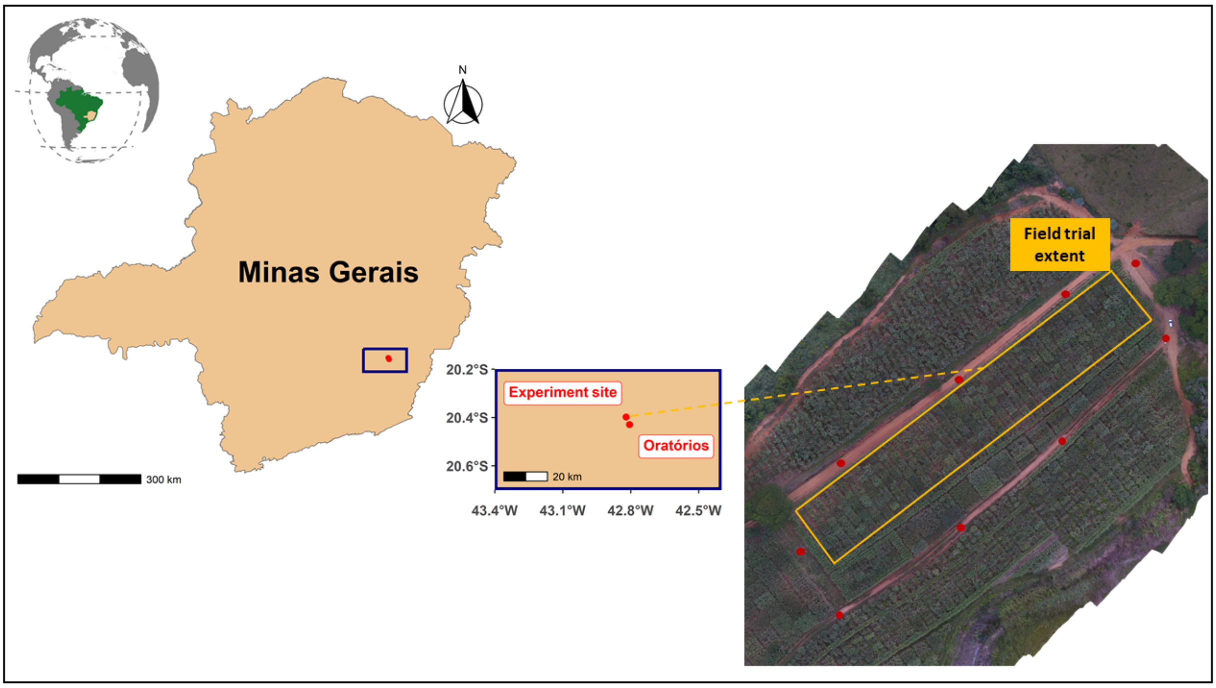
In 2019 (second ratoon), stalks were manually cut with machetes and cane yield potential, reflected by the tonnes of cane per hectare (TCH), was measured by mechanically weighing all stalks in each plot using a tractor-mounted boom pole fitted with a digital dynamometer. The whole plot weight records were converted to tonnes per hectare as:

$$TCH = (PW \times 10)/PA, \tag{4}$$

where PW is the plot weight (in kg) and PA is the plot area (in m^2).

We collected high-throughput phenotypic data in 2019 at a single timepoint in the end of the growing season (maturity stage). For this, we used a digital RGB camera (FC330, DJI, Shenzhen, Guangdong, China) attached to a quadcopter unoccupied aerial vehicle (Phantom 4 Pro, DJI, Shenzhen, Guangdong, China). Before the flight mission, we conducted a field survey in the study area. Then, a total of nine red ethylene-vinyl acetate (EVA) foam panels were distributed around and inside the field trial, which served as ground control point (GCP) targets (Figure 7). We used the DJI Ground Station and the DJI GS Pro app to program the flight mission. The autonomous flight mission was performed between 11:00-13:00 h, at a flight altitude of 50 m, and with 80 % forward and 60 % side image overlap. This resulted in a a ground sample distance (GSD) of 2.5 cm.

Figure 7 – Geographic location of the experiment site and orthomosaic image of field trial. Dark red dots in the orthomosaic image indicate the ground control points (GCPs).



Source: The author, 2024, p. 63

The GCPs were georeferenced using a global navigation satellite system (GNSS) receiver (T10, ComNav Technology Ltd., Shanghai, China) operating in real-time kinematic (RTK) mode to produce observations with a 0.05 m precision. This process enabled the orthorectification of the overlapping nadir images acquired to the South American geocentric reference system (SIRGAS 2000) datum with a projected coordinate system of UTM zone 23 S. After, GCPs were inported into Agisoft Metashape (Agisoft LLC, St. Petersburg, Russia) and processed using the structure-from-motion (SfM) algorithm for the three-dimensional reconstruction of the scene. With this data, we built the dense point cloud, and the digital elevation model, which allowed the generation of the orthomosaic image. Finally, we assessed geospatial accuracy using two of the nine GCPs as checkpoints.

The ’GeoTiff’ format orthomosaic file was imported into QGIS software (QGIS 3.4, Open-Source Geo-Spatial Foundation Project. https://qgis.osgeo.org). In QGIS, we created and adjusted corresponding polygon shapefiles to each plot (Figure 8). Then, image files were imported to R and we extracted the mean intensity pixel values of red, green and blue bands contained in the vector layer using the packages ”imager” and ”magick”. From the mean intensity pixels values of the respective visible bands we computed nine vegetation indices (VIs) according to the formulae shown in Table Table 4. These VIs have been widely adopted in previous studies and were shown to be closely associated with chlorophyll content, vigour, and biomass in several crops (Radočaj et al., 2023), including sugarcane (Cholula et al., 2020; Khuimphukhieo et al., 2023; Sanches et al., 2018). Also, due to the presumable ability of the red band to discriminate sugarcane clones (Peternelli & Andrade, 2023), potentially an indicator of differences in biomass content (Barzin et al., 2020).

Figure 8 – Example of polygon shapefile of a random plot.

Source: The author, 2024, p. 65

Table 4 – RGB-image extracted traits used in this study.

Band/Vegetation Index	Formula	Reference
Normalized Red band (r)	$\frac{R}{R+G+B}$	-
Normalized Green band (g)	$\frac{G}{R+G+B}$	-
Normalized Blue band (b)	$\frac{B}{R+G+B}$	-
Normalized Excess Red index ($NExR$)	$\frac{1.4r-G}{r+g+b}$	(Meyer et al., 1999)
Normalized Excess Green index ($NExG$)	$\frac{2g-r-b}{r+g+b}$	(Woebbecke et al., 1995)
Excess Green minus Excess Red index ($ExGR$)	$NExG-NExR$	(Meyer & Neto, 2008)
Visible Atmospherically Resistant index ($VARI$)	$\frac{g-r}{g+r-b}$	(Gitelson et al., 2002)
Normalized Green-Red difference index ($NGRDI$)	$\frac{g-r}{g+r}$	(Tucker, 1979)
Red Green Blue index ($RGBVI$)	$\frac{g^2-r\times b}{g^2+r\times b}$	(Bendig et al., 2015)
Modified Green-Red vegetation index ($MGRVI$)	$\frac{g^2-r^2}{g^2+r^2}$	(Bendig et al., 2015)
Green–Red ratio index ($GRRI$)	$\frac{g}{r}$	(Kanemasu, 1974)
Redness index (RI)	$\frac{r-g}{r+g}$	(Escadafal & Huete, 1991)

Notes: R, G, and B denote the values of red, green, and blue bands. r, g and b denote the normalized values of red, green, and blue bands.

Source: The author, 2024, p. 65

3.3.3. Genotypic data

The population of clones were genotyped using target enrichment sequencing, as described in more detail by Gonçalves et al., 2021. Briefly, markers with minor allele frequency lower than 5% were discarded. Also, markers with missing data were imputed from a binomial distribution density function using mean allele frequencies. After quality filtering, the resulted genotypic dataset was a matrix with order of 385 × 124,307 containing marker genotypes.

3.3.4. Phenotypic data analysis

The records of TCH and RGB-image extracted traits were analyzed using the lme4 package (Bates et al., 2005) in R (R Core Team, 2024). Best linear unbiased estimators (BLUES) were calculated by fitting the following linear mixed model:

$$y_{ij} = \mu + g_i + c_i + b_j + \epsilon_{ij}, \tag{5}$$

where y_{ij} is the phenotypic record of the *i*th clone measured in the *j*th block; μ is the overall mean; g_i is the fixed genotype effect of *i*th clone (BLUE), c_i is the random effect of check varieties distributed as $c_i \sim$ iid N(0,σ_c^2); b_j is the random effect of block distributed as $b_j \sim$ iid N(0,σ_b^2); and ϵ_{ij} is the associated residual effect distributed as $e_{ij} \sim$ iid N(0,σ_e^2).

3.3.5. Prediction models

In this study, single and multi-trait models using TCH as the primary trait were evaluated within the framework of a best linear unbiased prediction (BLUP) analysis. Accordingly, the prediction models considered three linear kernels (variance-covariance structures) that represent the genetic relationship among individuals. These linear kernels were constructed using different sources of information, including pedigrees, genotypes and visual spectral bands. All models were fitted using the R package sommer(Covarrubias-Pazaran 2016) through the “mmer” function.

3.3.6. Single-trait models

The general form of the single-trait model was:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u_i} + \boldsymbol{e}(i = 1, 2, 3), \tag{6}$$

where $\mathbf{y}$ is a 385 × 1 vector of BLUEs of TCH from equation (5); **1** is a vector of ones; μ is the overall mean; **Z** is an incidence matrix relating the elements of vector $\mathbf{y}$ to corresponding genetic effects; $\mathbf{u_i}$ is a vector of genetic effects; and e is the vector of residual effects. It was assumed that e ~ *N*(0,$\mathbf{I}\sigma_e^2$), u_1 ~ *N*(0,$\mathbf{A}\sigma_{u_1}^2$), and u_2 ~ *N*(0,$\mathbf{G}\sigma_{u_2}^2$), in which $\mathbf{I}$ is an identity matrix; $\sigma_{u_1}^2$, $\sigma_{u_2}^2$, and σ_e^2 are the corresponding pedigree-based and genomic-based additive genetic variances, and error variance, respectively; $\mathbf{A}$ is a 385×385 additive genetic relationship matrix calculated based on pedigrees records, and considering a minimum depth of three generations (progeny, parents, and grandparents); and $\mathbf{G}$ is the a 385×385 genomic relationship matrix computed following the first method of Van Raden (VanRaden, 2008). Models that incorporated $\mathbf{A}$ and $\mathbf{G}$ were denoted as ST-PBLUP, and ST-GBLUP, respectively. We used the AGHmatrix R package (Amadeu et al., 2016) to compute $\mathbf{A}$ and $\mathbf{G}$.

We also evaluated a third model (ST-KBLUP), which incorporated in the BLUP analysis a 385×385 genetic relationship matrix based on visual spectral bands extracted from RGB images. The visual spectral band reflectance-based relationship matrix $\mathbf{K}$ was obtained as :

$$\mathbf{K} = \frac{\mathbf{V}\mathbf{V}'}{n},$$

where $\mathbf{V}$ is a 385 $\times$ 3 scaled matrix of BLUEs of mean pixel values for each RGB channel for each genotype, $\mathbf{V}'$ is its transpose, and n is the number of visual spectral bands (n = 3). It was assumed that u_3 ~ *N*(0,$\mathbf{K}\sigma_{u_3}^2$), where $\sigma_{u_3}^2$ is the visual spectral band reflectance-based additive genetic variance.

3.3.7. Multi-trait models

The general form of the multi-trait model was:

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_j} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_j \end{bmatrix} + \begin{bmatrix} \mathbf{W_1} & 0 \\ 0 & \mathbf{W_j} \end{bmatrix} \begin{bmatrix} \mathbf{a_1} \\ \mathbf{a_j} \end{bmatrix} + \begin{bmatrix} \mathbf{e_1} \\ \mathbf{e_j} \end{bmatrix} \tag{7}$$

where $\mathbf{y^T} = [\mathbf{y_1}, ..., \mathbf{y_j}]$ is the vector of BLUEs obtained from equation (2) for the primary trait and secondary traits, respectively; $\mu^{\mathbf{T}} = [\mu_1, ..., \mu_j]$ is the vector of means; $\mathbf{a^T} = [a_1, ..., a_j]$ is the vector of additive genetic effects; and $\mathbf{e^T} = [e_1, ..., e_j]$ is the vector of residualeffects. It was assumed that $\mathbf{a^T} \sim$ *MVN*$(0,\mathbf{A} \otimes \mathbf{\Sigma_p})$ for multi-trait pedigree-based models (MT-PBLUP), or $\mathbf{a^T} \sim$ *MVN*$(0,\mathbf{G} \otimes \mathbf{\Sigma_g})$ for multi-trait genomic-based models (MT-GBLUP), and $\mathbf{e^T} \sim$ *MVN*$(0,\mathbf{R} \otimes \mathbf{I})$. Here, $\mathbf{\Sigma_p}$, $\mathbf{\Sigma_g}$, and $\mathbf{R}$, are $2x2$ unstructured variance-covariance matrices for the additive genetic, and residual effects of primary and secondary traits, respectively.

3.3.8. Genetic parameters and efficiency of selection

By fitting the model in equation 6, the variance components for each trait and relationship matrix were used to calculate narrow-sense heritability estimates as:

$$\hat{h}_a^2 = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_a^2 + \hat{\sigma}_e^2} \tag{8}$$

where $\hat{\sigma}_a^2$ is the additive genetic variance, and $\hat{\sigma}_e^2$ is the residual variance.

The variance components obtained from bivariate models in equation (4) were used to calculate the genetic correlation among traits as:

$$\hat{r}_{1j} = \frac{\hat{\sigma}_{a_{1j}}}{\sqrt{\hat{\sigma}_{a_1}^2 + \hat{\sigma}_{a_j}^2}} \tag{9}$$

where $\hat{r}_{1j}$ is the estimated genetic correlation between TCH and *j*th secondary trait; $\hat{\sigma}_{a_{1j}}$ is the covariance between TCH and *j*th secondary trait; $\hat{\sigma}_{a_1}^2$ and $\hat{\sigma}_{a_j}^2$ are the genetic variances of TCH and *j*th secondary trait, respectively. We also computed phenotypic correlations among traits using the Pearson's correlation coefficient.

To compare the efficiency of selection on single and multi trait models we estimated the direct and indirect response to selection (Morota et al., 2022). Here, for both parameters we did not considered the intensity of selection. We computed direct response to selection for each trait using the following equation:

$$R_x = \hat{h}_x \sigma_{gx} \tag{10}$$

where $\hat{h}_x$ is the square root of the narrow-sense heritability estimate of trait x, and σ_{gx} is the standard deviation of the additive genetic variance of trait x. We computed indirect response to selection as:

$$R_{XY} = \hat{h}_Y |r_{\hat{XY}}| \, \sigma_{gX} \tag{11}$$

where $\hat{h}^2$ is the narrow-sense heritability of secondary trait Y, $|r_{\hat{XY}}|$ is absolute value of the estimated genetic correlation between the primary trait X and secondary trait Y, and σ_{gX} is the square-root of the additive genetic variance of the primary trait X.

3.3.9. Prediction scenarios and assessment of models' performance

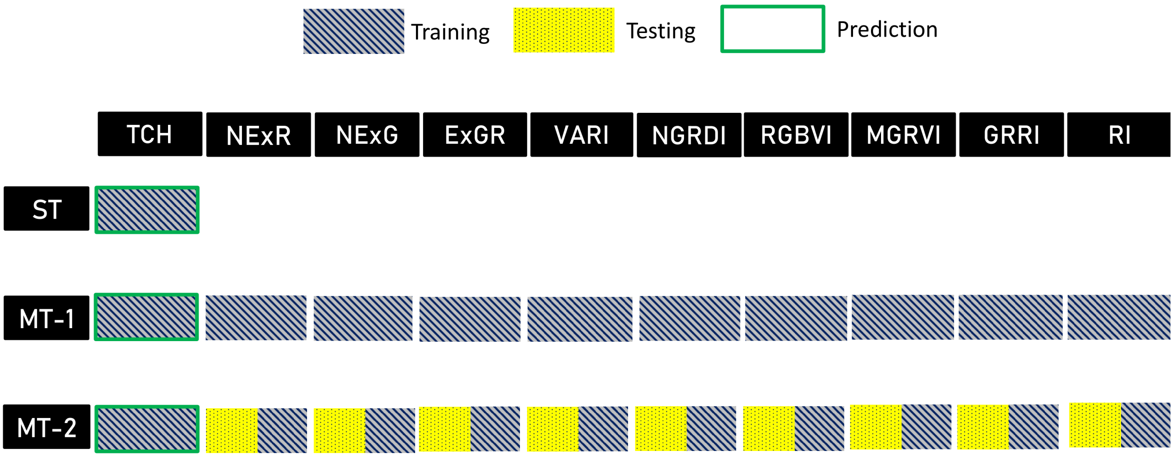
We evaluated models's performance using five-fold-cross validation. This procedure was used to estimate prediction accuracy and mean squared error (MSE). Prediction accuracy was computed as the average Pearson's correlation coefficient between observed and predicted values ($\hat{r} = cor(\hat{y}, y)$), and MSE as:

$$MSE = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n} \tag{12}$$

where $\hat{y}_i$ and y_i are the predicted value and observed value for the *i*th observation in the test set, and n is the number of observations. In five-fold-cross validation clones were randomly split into five folds, each consisted of 77 clones, of which four were used for training, and the remaining used to calculate prediction accuracy and RMSE estimates. This cross-validation procedure was repeated 50 times with different random partitions, where all models were analyzed in the same partition.

In total, we evaluated three single-trait (ST-PBLUP, ST-GBLUP, ST- KBLUP) and two multi-trait (MT-PBLUP, MT-GBLUP) models in three prediction scenarios (Figure 9). In the first prediction scenario (ST) we applied single-trait models, where clones had been evaluated only for TCH. In the second scenario (MT-1), phenotypic data of secondary traits were combined with those of TCH in multi-trait models. In the third scenario (MT-2), we also considered multi-trait models, but phenotypic data of secondary traits were available on both, training and testing sets. This scenario evaluates the prediction ability of models when selection candidates would have been evaluated at the time of predictions for secondary traits. In all scenarios, the primary trait was targeted for predictions.

Figure 9 – Graphic representation of the Cross-Validation schemes. ST: single-trait scenario; MT-1: first multi-trait scenario.; MT-2: second multi-trait scenario. TCH: tonnes of cane per hectare; ExR: Excess Red index; ExG: Excess Green index; ExGR Excess Green minus Excess Red index; VARI: Visible Atmospherically Resistant index; NGRDI: Normalized Green-Red difference index; RGBVI: Red Green Blue index; MGRVI: Modified Green-Red vegetation index; GRRI: Green–Red ratio index; RI: Redness index.



Source: The author, 2024, p. 70

3.4. Results

3.4.1. Heritability and genetic correlation

For the primary trait (TCH), heritability estimates showed the lowest values of 0.13, and 0.16 using genomic- and pedigree-based models, respectively (Table 5). The narrow-sense heritability estimates for secondary traits were all higher compared to TCH, and ranged from 0.27 to 0.80 for genomic-based models and from 0.49 to 0.73 for pedigree-based models. Overall, narrow-sense heritability estimates based on genomic information were lower compared to those based on pedigree information for most traits, including TCH.

Table 5 – Narrow-sense heritability (h^2), genetic correlation (r_g), direct response to selection (R_x) and indirect response to selection (R_{XY}) estimates of RGB-image extracted traits and TCH using genomic- (**G**) and pedigree-based (**A**) relationship matrices.

Trait	G				A			
	h^2	r_g	R_x	R_{XY}	h^2	r_g	R_x	R_{XY}
NExR	0.32	-0.27	-	0.04	0.52	-0.34	-	0.13
NExG	0.79	0.22	-	0.12	0.71	0.30	-	0.19
ExGR	0.55	0.25	-	0.08	0.65	0.33	-	0.18
VARI	0.80	0.21	-	0.11	0.73	0.29	-	0.18
NGRDI	0.52	0.25	-	0.08	0.64	0.33	-	0.17
RGBVI	0.79	0.22	-	0.12	0.70	0.30	-	0.18
MGRVI	0.40	0.26	-	0.06	0.56	0.34	-	0.14
GRRI	0.36	0.26	-	0.05	0.56	0.34	-	0.14
RI	0.27	-0.28	-	0.04	0.49	-0.34	-	0.12
TCH	0.13	-	0.14	-	0.16	-	0.16	-

Notes: NExR: Excess Red index; NExG: Excess Green index; ExGR Excess Green minus Excess Red index; VARI: Visible Atmospherically Resistant index; NGRDI: Normalized Green-Red difference index; RGBVI: Red Green Blue index; MGRVI: Modified Green-Red vegetation index; GRRI: Green–Red ratio index; RI: Redness index; TCH: tonnes of cane per hectare.
Source: The author, 2024, p. 71

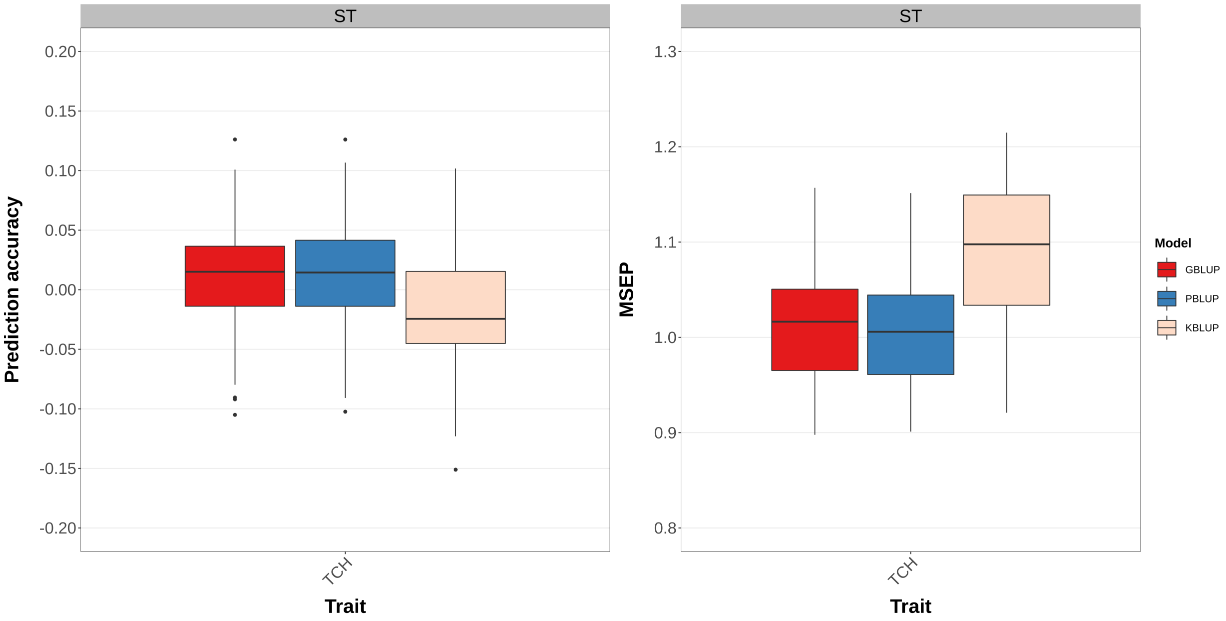
Genetic correlation estimates are provided in Table 5. Low positive and similar genetic correlations among secondary traits and TCH were observed using both, genomic- and pedigree-based models. Genetic correlation estimates ranged from -0.28 to 0.26

and from -0.34 to 0.34 for genomic- and pedigree-based models, respectively. With the exception of NExR, and RI, secondary traits were positively correlated with TCH, regardless of the relationship matrix used. Moreover, the strength of the genetic correlations were dependent on the relationship matrix used, where pedigree-based correlations had higher values.

3.4.2. Performance of single- and multi-trait models

Figure 10 shows the distribution of cross-validation prediction accuracies and the mean squared error of prediction for single-trait models. As shown in Figure 10, the cross-validation results using single-trait models gave near-zero and negative prediction accuracies. Prediction accuracies ranged from -0.11 to 0.12, from -0.10 to 0.13, and from -0.15 to 0.10, with means of 0.01, 0.01, and -0.02, for GBLUP, PBLUP and KBLUP models, respectively. The average MSEP ranged from 0.90 to 1.16, from 0.90 to 1.15, and from 0.92 to 1.21 for GBLUP, PBLUP and KBLUP models, respectively. The single-trait model that considered the relationship matrix derived from RGB bands (KBLUP) gave the lowest mean prediction accuracy and higher MSEP. In general, in terms of mean prediction accuracies and MSEP genomic-based (GBLUP) models performed similarly compared to pedigree-based models (PBLUP).

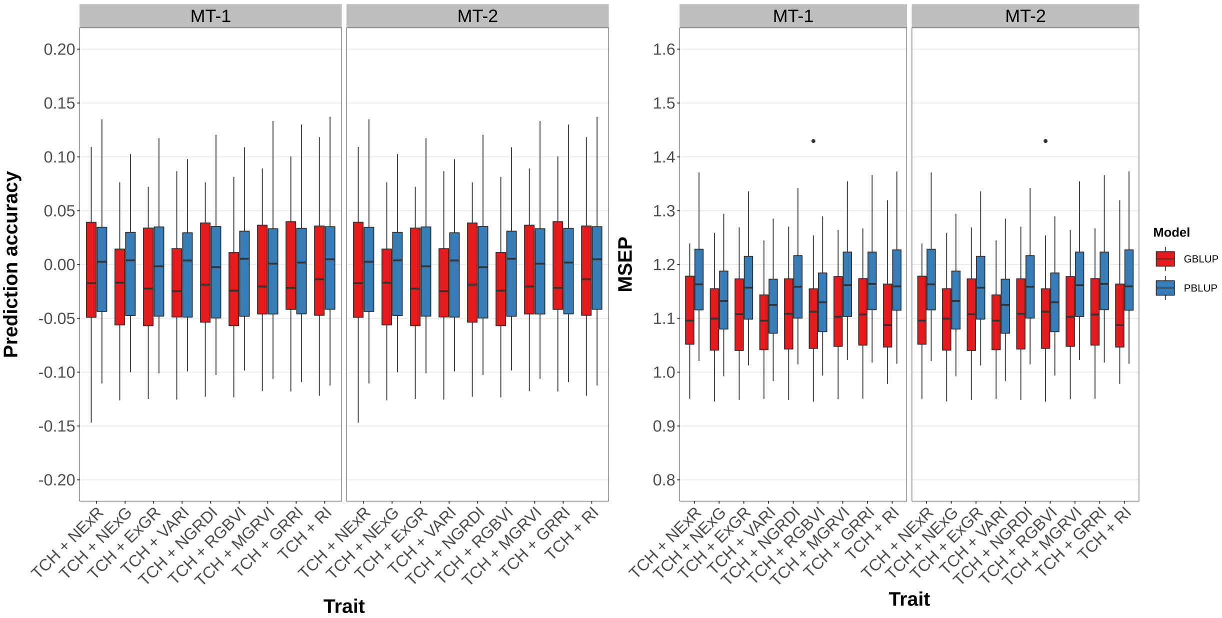
Figure 10 – Distribution of Cross-Validation prediction accuracies and mean square error of prediction (MSEP) of single-trait prediction models for tonnes of cane per hectare (TCH). ST: single-trait scenario.



Source: The author, 2024, p. 73

Figure 11 shows the distribution of cross-validation prediction accuracies and the mean squared error of prediction for multi-trait models evaluated in prediction scenarios MT-1 and MT-2. In both prediction scenarios evaluated (MT-1 and MT-2) low and similar prediction accuracies were found with negligible differences among MT-1 and MT-2. Also, in both MT-1 and MT-2, no single combination of secondary traits performed better for the prediction of TCH. In the MT-1 scenario, mean prediction accuracies ranged from -0.0195 to -0.0098, and from -0.0006 to -0.0118, with average MSEP ranging from 1.1052 to 1.8292, and from 1.1278 to 1.1720 for genomic and pedigree-based models, respectively. In the MT-2 scenario, mean prediction accuracies ranged from -0.0195 to -0.0098, and from -0.0075 to -0.0006 with average MSEP ranging from 1.1052 to 1.8292, and from 1.1278 to 1.1720 for genomic and pedigree-based models, respectively. Compared to single-trait models, the mean cross-validation precition accuracies and avarage MSEP of multi-trait models yielded similar results, regardless of the pairwise combination of secondary and the primary trait and the presence of secondary traits in training and testing sets.

Figure 11 – Distribution of Cross-Validation prediction accuracies and mean square error of prediction (MSEP) of multi-trait prediction models for tonnes of cane per hectare (TCH) using different combinations of secondary traits. MT-1: first multi-trait scenario.; MT-2: second multi-trait scenario. TCH: tonnes of cane per hectare; ExR: Excess Red index; ExG: Excess Green index; ExGR Excess Green minus Excess Red index; VARI: Visible Atmospherically Resistant index; NGRDI: Normalized Green-Red difference index; RGBVI: Red Green Blue index; MGRVI: Modified Green-Red vegetation index; GRRI: Green–Red ratio index; RI: Redness index.



Source: The author, 2024, p. 74

3.4.3. Efficiency of selection

Table 5 presents the estimates of expected direct response to selection (R_x) for each trait and of indirect response to selection (R_{XY}) for TCH with RGB-image extracted traits using $\mathbf{G}$ and $\mathbf{A}$ relationship matrices in multi-trait models. All secondary traits showed a higher indirect response to selection compared to direct selection for TCH. The secondary trait NExG showed the highest indirect response to selection. Also, ExGR VARI, RGBVI, and NGRDI had higher estimates of indirect response to selection compared to direct response to selection for TCH. Both direct, and indirect response to selection for primary and secondary traits were higher for models using the relationship matrix $\mathbf{A}$ as compared to $\mathbf{G}$.

3.5. Discussion

3.5.1. Single-trait models

The use of genotypic and pedigree data in prediction models has become of particular interest for plant breeders (Crossa et al., 2017; Dreisigacker et al., 2021). Several studies have demonstrated the potential advantage these information sources can have for predicting breeding values of selection candidates (Burgueño et al., 2012; Crossa et al., 2010; Juliana et al., 2018). In such models, predictions benefit from borrowing information shared between related individuals.

In this study, the predictive ability of both pedigree, and genomic-based single-trait models was substantially low, and similar. The genomic prediction results we found are in line with those of Islam et al. (2022), which reported very low genomic prediction accuracies for TCH. In contrast, the cross-validation genomic prediction accuracies we found are lower compared to other studies. For example, Yadav et al. (2021) reported prediction accuracies for TCH ranging from 0.218 to 0.325. In Hayes et al. (2021), despite the lowest accuracies of predicted genomic estimated breeding values being reported for TCH, the authors found higher avarage accuracies, of 0.3 across four locations. Deomano et al. (2020) reported higher pedigree and genomic prediction accuracies for TCH compared to our study.

These results can be attributed to the complex genetic architecture of TCH, which is influenced by a number factors such as climatic variables, biotic stresses, and inter-plot competition effects, particularly in small unreplicated plots, potentially resulting in large experimental error (Deomano et al., 2020; Natarajan et al., 2019). This was reflected by the low estimates of narrow-sense heritability we observed for TCH. Unlike most of the studies that assessed genomic prediction for TCH, here we used data of small unreplicated plots. Another possibility is that we employed models that considered only additive genetic effects, and these were insufficient to model the underlying genetic variance. The results of Hayes et al. (2021) and Deomano et al. (2020) suggest that models including only additive genetic effects may oversimplify the complexity of predicting TCH, which could be largely governed by non-additive genetic effects.

Another possible reason to explain the genomic prediction results might be the inadequate distribution of markers across the genome, especially for a large genome plant specie and a complex traits such as TCH, rendering markers unable to capture all QTLs (O. A. Montesinos-López et al., 2023). Also, our population was derived from a previously established family trial submitted to selection, this may have resulted in reduced genetic variation for TCH, diminishing the power of genomic predictions (Deomano et al., 2020; Juliana et al., 2018). In this study, pedigree and genomic information were not informative for predicting TCH. However, for less complex traits such as stalk number or soluble solids, pedigrees can be a valuable genetic information source, as demonstrated in the first chapter of this thesis and by Inamori et al. (2024).

In sugarcane breeding programs, the evaluation of TCH is a time-consuming and labor-intensive process. Among the factors responsible for this are the large number of progeny populations generated and the size of the crop, which can reach 4 m in height (Lu et al., 2024). In this sense, remote in-field measurements using UAS-based RGB imaging is an appealing tool, enabling sugarcane breeders to screen extensive field trials more efficiently at a relatively low cost (Natarajan et al., 2019; Todd et al., 2022). The usefulness of RGB-image extracted traits to estimate TCH and TCH-related traits has been investigated in previous studies (Sanches et al., 2018; Todd et al., 2022). In particular, spectral reflectance traits are frequently employed. Spectral reflectance traits are developed using straightforward mathematical formulae expressed as VIs (Liu et al., 2019). For example, Sumesh et al. (2021) found stalk density to be correlated with the Excess Green index. Also, Todd and Johnson (2021) utilized UAS-based RGB imaging to predict yield of sugarcane families for a second ratoon crop. Authors found that NGRDI and other spectral traits of the CIElab color space model significantly correlated with cane yield.

Besides VIs, raw data from individual wavelengths can be used to express the information of spectral reflectance traits (Mróz et al., 2024). Thus, we considered a third modeling strategy that incorporated RGB visual spectral bands to derive genotypic relationship matrix $\mathbf{K}$. Similarly to pedigree- and genomic-based accuracies, single-trait models that incorporated $\mathbf{K}$ gave a very poor performance. One possible explanation for this results is that we conducted only one flight during the crop season, which may have contained insufficient information to describe the genetic components of TCH. Mróz et al. (2024) combined data of multiple time points across crop years, potentially providing a more informative dataset. Indeed, the frequency of data collecting points, and the optimal date to measure spectral reflectance traits are known to be important aspects of UAS-based imaging (Herr et al., 2023). The results of Zhao et al. (2016) indicated that the best time to measure canopy reflectance for TCH estimation was during the tillering stage, which is an earlier growth stage compared to this study. In (Natarajan et al., 2019), authors considered average measurements across multiple time points over the crop season. They found that genetic variance and heritability of NDVI was higher in early growth stages compared to late stages, suggesting that as sugarcane plants reach physiological maturity, canopies close and spectral reflectance saturates, hindering the discrimination of clones. Similarly, Cholula et al. (2020) reported higher correlations of NDVI with cane yield roughly at the tillering stage. But for the excess green index, correlations with cane yield were only significant at the maturity stage. Sanches et al. (2018) identified the transition point between tillering and elongation (biomass accumulation) stage as the optimal timing to predict cane yield. The results of Hoffman et al. (2024) indicated that canopy temperature and NDVI could be used to identify high- and low-yielding sugarcane genotypes when measured earlier on in the growth cycle for well-watered environments. A potential weakness of our study was that since we collected data in the end of crop cycle, the height of plant canopies and overlap of leaves from neighboring plots may have interfered with image data quality (Gitelson et al., 2002). Indeed, sugarcane is known to have a complex canopy architecture (Barbosa Júnior et al., 2021).

3.5.2. Multi-trait models

Previous studies suggest that the integration of high-throughput phenotyping (HTP) data can further improve genomic prediction models, resulting in enhanced prediction accuracy (Morota et al., 2022; Persa et al., 2021). One of the applications of this strategy is to incorporate secondary traits measured with UAS as response variables into multi-trait models (Rutkoski et al., 2016). In this modeling strategy, the benefits stem from heritability and genetic correlation between primary and secondary traits (Calus & Veerkamp, 2011; Jia & Jannink, 2012). Multi-trait models are able to levarage information shared between individuals and also between traits (A. Montesinos-López et al., 2021). This can be an useful aspect for predictions across populations, cycles and environments. As we discussed in the previous chapter, there is a trade-off between the benefits and costs of genomic prediciton, particularly at early selection stages, where its application can be prohibitively expensive, due to the large number of selection candidates. Although pedigrees also have inherent weakness, such as being unable to capture within family variation, they represent a more affordable solution for predicting TCH at early selection stages. Hence, an alternative strategy would be to use multi-trait pedigree prediction models (Crossa et al., 2017).

Our results indicated that multi-trait models did not improve prediction accuracies compared to single-trait models, regardless of the cross-validation scheme. Moreover, we observed negligible differences of multi-trait models using $\mathbf{G}$ and $\mathbf{A}$ relationship matrices. These results can be attributable to the complexity of TCH and the broad range of heritability estimates of RGB-image extracted traits evaluated in the present study (Bhatta et al., 2020). This is in agreement with the results of Fernandes et al. (2018), who found that single- and multi-trait genomic models yielded similar results for predicting biomass yield in sorghum. Similarly, Lado et al. (2018) used cross-validation MT-1 and did not found multi-trait genomic to perform better than single-trait models for bread wheat baking quality traits. In contrast, the simulated results of Jia and Jannink (2012) showed that prediction accuracy for a low-heritability trait was significantly increased using multi-trait genomic prediction models. Rutkoski et al. (2016) found that by including secondary VIs prediction accuracies increased for grain yield in wheat using both, pedigree and genomic relationships. Similar results were found by Sun et al. (2017), who evaluated multi-trait pedigree and genomic models using canopy temperature and/or NDVI as secondary traits for predicting wheat grain yield.

One of the attractive features of UAS-based imaging is the ability to measure secondary traits that are correlated with difficult to measure primary traits, such as TCH, on large populations of selection candidates in a fast and cost-efficient manner. This feature allow prediction models to exploit not only information shared between related individuals, but also *per se* data of selection candidates at the time of selections (Rutkoski et al., 2016). Accordingly we tested a second prediction scenario using the cross-validation scheme MT-2. Here, we did not observed improvements in prediction accuracy of multi-trait models when RGB-image extracted traits were included in the testing set. Previous studies reported that significant improvements in predictions were observed when secondary traits were included on training and test set (Bhatta et al., 2020; Gaire et al., 2022; Lado et al., 2018; Muvunyi et al., 2022; Rutkoski et al., 2016; Sun et al., 2017; Velazco et al., 2019). A possible explanation could be the combination of low genetic correlations between TCH and RGB-image extracted traits and associated modest heritability estimates of the latter (Shahi et al., 2022).

Narrow-sense heritability of TCH was low using both $\mathbf{G}$ and $\mathbf{A}$. By contrast, all RGB-image extracted traits showed higher heritability than TCH, with estimates using $\mathbf{A}$ outperforming those using $\mathbf{G}$. However, these higher estimates of heritability did not translated into increased prediction accuracy of multi-trait models. This was likely due to experimental error variations as a result of using small unreplicated plots and a diverse set of genotypes, which may lead to biased estimates genetic parameters (Herr et al., 2023; Krause et al., 2020). The results found by Jia and Jannink (2012) suggest that genetic correlation might be more important than the heritability. Moreover, for complex traits, multi-trait models provided only negligible improvements relative to single-trait models. According to Montesinos-Lopez et al. (2022), multi-trait models can help increase the prediction accuracy when the degree of correlation is moderate to large. Despite the importance of the degree of heritability some authors argued that genetic correlations might be more a decisive factor (Krause et al., 2020). Therefore, the higher the absolute genetic correlation between traits, the greater the benefit of multi-trait models (A. Montesinos-López et al., 2021). In our dataset correlations between spectral reflectance traits and TCH were low. Khuimphukhieo et al. (2023) evaluated a series of RGB image-extracted traits to assess cane yield and quality traits. They found that most spectral reflectance traits measured between elongation and maturity growth stages were positively associated with TCH, including ExG, ExGR, MGRVI, NGRDI, RGBVI and VARI. In contrast, in the maturity phase correlations decreased and were equivalent the genetic correlation estimates of this study, corroborating our results and highlighting importance of defining the appropriate flight timing. Moreover, the findings of Zhao et al. (2016) indicated that TCH was better estimated using data collected before sugarcane canopy closure, as differences among clones for spectral reflectance traits are much smaller compared to early growth stages, leading to very low correlations.

Multi-trait models can provide more accurate estimates of genetic correlations between traits, thus improving selection decisions (O. A. Montesinos-López et al., 2019). Here, we found that the strength of the genetic correlations, direct, and indirect response to selection for primary and secondary traits were dependent on the relationship matrix used, where pedigree-based estimates yielded higher values. Krause et al. (2020) reported that HTP traits measured in small plots correlated with grain yield of larger plot sizes. This results suggests the indirect selection using UAS-based VIs could result in superior genetic gains compared to direct selection for TCH on small unreplicated plots.

In the present study, we restricted the analysis to solely bi-variate models. Sun et al. (2017) observed that including all secondary traits in multi-trait pedigree and genomic models did not improved predictive ability for grain yield compared to bi-variate models. Authors found that including a single secondary trait showing higher indirect selection efficiency gave the best results. In contrast, found multi-trait model that used both CT and NDVI in general had better prediction accuracy for those traits compared to model that used a single trait, i.e. either CT or NDVI, with a few exceptions (Lado et al., 2018; Sandhu et al., 2021). The results found by Gaire et al. (2022) corroborate that using only one or few highly heritable, and strongly correlated secondary traits to predict primary traits may be a more parsimonious approach (Sandhu et al., 2021). This is because the inclusion of more traits could come at the cost of increasing computational intensity and/or lead to problems with multicollinearity, and convergence to solve mixed model equations (O. A. Montesinos-López et al., 2019; Sandhu et al., 2021).

3.5.3. Future directions

The effectiveness of UAS-based imaging to optimize genetic evaluation for quality, yield, and yield-related traits in sugarcane breeding has been demonstrated in previous studies (Akbarian et al., 2023; Chea et al., 2020; de Oliveira et al., 2022; Poudyal et al., 2022). In future studies, the benefits of this tool could be greater by conducting multiple measurements over the growing season. Also, future investigations should assess different sensor technologies, such as multispectral, thermal, and hyperspectral cameras. These sensors could provide information on different wavelength that can be more informative for predicting TCH. Finally, future studies should consider the optimization of protocols for UAV flight planning and image processing, as these factors can have a significant impact on image data quality (Gano et al., 2024).

3.6. Conclusion

We found low genetic correlation between TCH and RGB-image extracted traits, and overall moderate narrow-sense heritability estimates for RGB-image extracted traits. Indirect response to selection using the indices ExGR VARI, RGBVI, and NGRDI was slightly higher compared to direct response to selection for TCH. In general, our results suggest that prediction accuracies of multi-trait models that incorporated RGB-image extracted traits did not improve compared to single-trait models for predicting TCH. Future research should investigate alternative sensor technologies and a temporal framework for UAS-based data collection.

3.7. References

Acreche, M. M., Saez, J. V., & Vera, J. C. (2015). Physiological bases of genetic gains in sugarcane yield in argentina. *Field Crops Research*, *175*, 80–86.

Akbarian, S., Jamnani, M. R., Xu, C., Wang, W., & Lim, S. (2023). Plot level sugarcane yield estimation by machine learning on multispectral images: A case study of bundaberg, australia. *Information Processing in Agriculture*.

Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A., Resende Jr, M. F., & Muñoz, P. R. (2016). Aghmatrix: R package to construct relationship matrices for autotetraploid and diploid species: A blueberry example. *The plant genome*, *9*(3), plantgenome2016–01.

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends in plant science*, *23*(5), 451–466.

Barbosa Júnior, M. R., Moreira, B. R. d. A., De Oliveira, R. P., Shiratsuchi, L. S., & Da Silva, R. P. (2023). Uav imagery data and machine learning: A driving merger for predictive analysis of qualitative yield in sugarcane. *Frontiers in Plant Science*, *14*, 1114852.

Barbosa Júnior, M. R., Tedesco, D., Corrêa, R. d. G., Moreira, B. R. d. A., Silva, R. P. d., & Zerbato, C. (2021). Mapping gaps in sugarcane by uav rgb imagery: The lower and earlier the flight, the more accurate. *Agronomy*, *11*(12), 2578.

Barzin, R., Pathak, R., Lotfi, H., Varco, J., & Bora, G. C. (2020). Use of uas multispectral imagery at different physiological stages for yield prediction and input resource optimization in corn. *Remote Sensing*, *12*(15), 2392.

Bates, D., et al. (2005). Fitting linear mixed models in r. *R news*, *5*(1), 27–30.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M. L., & Bareth, G. (2015). Combining uav-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation*, *39*, 79–87.

Bhatta, M., Gutierrez, L., Cammarota, L., Cardozo, F., Germán, S., Gómez-Guerrero, B., Pardo, M. F., Lanaro, V., Sayas, M., & Castro, A. J. (2020). Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (hordeum vulgare l.) *G3: Genes, Genomes, Genetics*, *10*(3), 1113–1124.

Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. *Crop Science*, *52*(2), 707–719.

Calus, M. P., & Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genetics Selection Evolution*, *43*, 1–14.

Chea, C., Saengprachatanarug, K., Posom, J., Wongphati, M., & Taira, E. (2020). Sugar yield parameters and fiber prediction in sugarcane fields using a multispectral camera mounted on a small unmanned aerial system (uas). *Sugar Tech*, *22*, 605–621.

Cholula, U., da Silva, J. A., Marconi, T., Thomasson, J. A., Solorzano, J., & Enciso, J. (2020). Forecasting yield and lignocellulosic composition of energy cane using unmanned aerial systems. *Agronomy*, *10*(5), 718.

Crossa, J., Campos, G. d. l., Pérez, P., Gianola, D., Burgueno, J., Araus, J. L., Makumbi, D., Singh, R. P., Dreisigacker, S., Yan, J., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, *186*(2), 713–724.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., et al. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in plant science*, *22*(11), 961–975.

Cursi, D. E., Hoffmann, H. P., Barbosa, G., Bressiani, J. A., Gazaffi, R., Chapola, R. G., Fernandes Junior, A., Balsalobre, T. W. A., Diniz, C. A., Santos, J. M., et al. (2022). History and current status of sugarcane breeding, germplasm development and molecular genetics in brazil. *Sugar Tech*, *24*(1), 112–133.

da Silveira, L. C. I., Brasileiro, B. P., Kist, V., Weber, H., Daros, E., Peternelli, L. A., & Barbosa, M. H. P. (2015). Selection strategy in families of energy cane based on biomass production and quality traits. *Euphytica*, *204*, 443–455.

de Oliveira, R. P., Barbosa Júnior, M. R., Pinto, A. A., Oliveira, J. L. P., Zerbato, C., & Furlani, C. E. A. (2022). Predicting sugarcane biometric parameters by uav multispectral images and machine learning. *Agronomy*, *12*(9), 1992.

De Souza, C. H. W., Lamparelli, R. A. C., Rocha, J. V., & Magalhães, P. S. G. (2017). Height estimation of sugarcane using an unmanned aerial system (uas) based on structure from motion (sfm) point clouds. *International journal of remote sensing*, *38*(8-10), 2218–2230.

Deomano, E., Jackson, P., Wei, X., Aitken, K., Kota, R., & Pérez-Rodríguez, P. (2020). Genomic prediction of sugar content and cane yield in sugar cane clones in different stages of selection in a breeding program, with and without pedigree information. *Molecular Breeding*, *40*, 1–12.

Dreisigacker, S., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Rosyara, U., Juliana, P., Mondal, S., Crespo Herrera, L., Velu, G., Singh, R. P., et al. (2021). Implementation of genomic selection in the cimmyt global wheat program, findings from the past 10 years. *Crop Breed Genetics and Genomics*.

Escadafal, R., & Huete, A. (1991). Etude des propriétés spectrales des sols arides appliquée à l'amélioration des indices de végétation obtenus par télédétection. *Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre*, *312*(11), 1385–1391.

Fernandes, S. B., Dias, K. O., Ferreira, D. F., & Brown, P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theoretical and applied genetics*, *131*, 747–755.

Gaire, R., de Arruda, M. P., Mohammadi, M., Brown-Guedira, G., Kolb, F. L., & Rutkoski, J. (2022). Multi-trait genomic selection can increase selection accuracy for deoxynivalenol accumulation resulting from fusarium head blight in wheat. *The plant genome*, *15*(1), e20188.

Gano, B., Bhadra, S., Vilbig, J. M., Ahmed, N., Sagan, V., & Shakoor, N. (2024). Drone-based imaging sensors, techniques, and applications in plant phenotyping for crop breeding: A comprehensive review. *The Plant Phenome Journal*, *7*(1), e20100.

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote sensing of Environment*, *80*(1), 76–87.

Gonçalves, M. T. V., Morota, G., Costa, P. M. d. A., Vidigal, P. M. P., Barbosa, M. H. P., & Peternelli, L. A. (2021). Near-infrared spectroscopy outperforms genomics for predicting sugarcane feedstock quality traits. *PLoS One*, *16*(3), e0236853.

Hayes, B. J., Wei, X., Joyce, P., Atkin, F., Deomano, E., Yue, J., Nguyen, L., Ross, E. M., Cavallaro, T., Aitken, K. S., et al. (2021). Accuracy of genomic prediction of complex traits in sugarcane. *Theoretical and Applied Genetics*, *134*, 1455–1462.

Herr, A. W., Adak, A., Carroll, M. E., Elango, D., Kar, S., Li, C., Jones, S. E., Carter, A. H., Murray, S. C., Paterson, A., et al. (2023). Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. *Crop Science*, *63*(4), 1722–1749.

Hoffman, N., Singels, A., & Joshi, S. (2024). Aerial phenotyping for sugarcane yield and drought tolerance. *Field Crops Research*, *308*, 109275.

Huang, Q., Feng, J., Gao, M., Lai, S., Han, G., Qin, Z., Fan, J., & Huang, Y. (2024). Precise estimation of sugarcane yield at field scale with allometric variables retrieved from uav phantom 4 rtk images. *Agronomy*, *14*(3), 476.

Inamori, M., Kimura, T., Mori, M., Tarumoto, Y., Hattori, T., Hayano, M., Umeda, M., & Iwata, H. (2024). Machine learning for genomic and pedigree prediction in sugarcane. *The Plant Genome*, e20486.

Islam, M. S., McCord, P., Read, Q. D., Qin, L., Lipka, A. E., Sood, S., Todd, J., & Olatoye, M. (2022). Accuracy of genomic prediction of yield and sugar traits in saccharum spp. hybrids. *Agriculture*, *12*(9), 1436.

Jackson, P. A. (2005). Breeding for improved sugar content in sugarcane. *Field Crops Research*, *92*(2-3), 277–290.

Jia, Y., & Jannink, J.-L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*, *192*(4), 1513–1522.

Juliana, P., Singh, R. P., Poland, J., Mondal, S., Crossa, J., Montesinos-López, O. A., Dreisigacker, S., Pérez-Rodríguez, P., Huerta-Espino, J., Crespo-Herrera, L., et al. (2018). Prospects and challenges of applied genomic selection—a new paradigm in breeding for grain yield in bread wheat. *The plant genome*, *11*(3), 180017.

Kanemasu, E. (1974). Seasonal canopy reflectance patterns of wheat, sorghum, and soybean. *Remote Sensing of Environment*, *3*(1), 43–47.

Khuimphukhieo, I., Bhandari, M., Enciso, J., & da Silva, J. A. (2024). Assessing drought stress of sugarcane cultivars using unmanned vehicle system (uas)-based vegetation indices and physiological parameters. *Remote Sensing*, *16*(8), 1433.

Khuimphukhieo, I., da Silva, J. A., Bhandari, M., & Enciso, J. (2024). Estimating sugarcane yield and its components using unmanned aerial systems (uas)-based high throughput phenotyping (htp). *Available at SSRN 4764376*.

Khuimphukhieo, I., Marconi, T., Enciso, J., & da Silva, J. A. (2023). The use of uas-based high throughput phenotyping (htp) to assess sugarcane yield. *Journal of Agriculture and Food Research*, *11*, 100501.

Krause, M. R., Mondal, S., Crossa, J., Singh, R. P., Pinto, F., Haghighattalab, A., Shrestha, S., Rutkoski, J., Gore, M. A., Sorrells, M. E., et al. (2020). Aerial high-throughput phenotyping enables indirect selection for grain yield at the early generation, seed-limited stages in breeding programs. *Crop Science*, *60*(6), 3096–3114.

Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., & Gutiérrez, L. (2018). Resource allocation optimization with multi-trait genomic prediction for bread wheat (triticum aestivum l.) baking quality. *Theoretical and Applied Genetics*, *131*, 2719–2731.

Liu, C., Pinto, F., Cossani, C. M., Sukumaran, S., & Reynolds, M. P. (2019). Spectral reflectance indices as proxies for yield potential and heat stress tolerance in spring wheat: Heritability estimates and marker-trait associations. *Frontiers of Agricultural Science and Engineering*, *6*(3), 296–308.

Lu, G., Liu, P., Wu, Q., Zhang, S., Zhao, P., Zhang, Y., & Que, Y. (2024). Sugarcane breeding: A fantastic past and promising future driven by technology and methods. *Frontiers in Plant Science*, *15*, 1375934.

Luo, T., Liu, X., & Lakshmanan, P. (2023). A combined genomics and phenomics approach is needed to boost breeding in sugarcane. *Plant Phenomics*, *5*, 0074.

Meyer, G. E., Hindman, T. W., & Laksmi, K. (1999). Machine vision detection parameters for plant species identification. *Precision agriculture and biological quality*, *3543*, 327–335.

Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and electronics in agriculture*, *63*(2), 282–293.

Montesinos-Lopez, O. A., Montesinos-López, A., Mosqueda-Gonzalez, B. A., Montesinos-López, J. C., & Crossa, J. (2022). Accounting for correlation between traits in genomic prediction. *Genomic Prediction of Complex Traits: Methods and Protocols*, 285–327.

Montesinos-López, A., Runcie, D. E., Ibba, M. I., Pérez-Rodríguez, P., Montesinos-López, O. A., Crespo, L. A., Bentley, A. R., & Crossa, J. (2021). Multi-trait genomic-enabled prediction enhances accuracy in multi-year wheat breeding trials. *G3*, *11*(10), jkab270.

Montesinos-López, O. A., Bentley, A. R., Saint Pierre, C., Crespo-Herrera, L., Rebollar-Ruellas, L., Valladares-Celis, P. E., Lillemo, M., Montesinos-López, A., & Crossa, J. (2023). Efficacy of plant breeding using genomic information. *The Plant Genome*, *16*(2), e20346.

Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Cuevas, J., Montesinos-López, J. C., Gutiérrez, Z. S., Lillemo, M., Philomin, J., & Singh, R. (2019). A bayesian genomic multi-output regressor stacking model for predicting multi-trait multi-environment plant breeding data. *G3: Genes, Genomes, Genetics*, *9*(10), 3381–3393.

Morota, G., Jarquin, D., Campbell, M. T., & Iwata, H. (2022). Statistical methods for the quantitative genetic analysis of high-throughput phenotyping data. In *High-throughput plant phenotyping: Methods and protocols* (pp. 269–296). Springer.

Mróz, T., Shafiee, S., Crossa, J., Montesinos-Lopez, O. A., & Lillemo, M. (2024). Multispectral derived genotypic similarities from budget cameras allow grain yield prediction and genomic selection augmentation in single and multi-environment scenarios in spring wheat. *Molecular Breeding*, *44*(1), 5.

Muvunyi, B. P., Zou, W., Zhan, J., He, S., & Ye, G. (2022). Multi-trait genomic prediction models enhance the predictive ability of grain trace elements in rice. *Frontiers in Genetics*, *13*, 883853.

Natarajan, S., Basnayake, J., Wei, X., & Lakshmanan, P. (2019). High-throughput phenotyping of indirect traits for early-stage selection in sugarcane breeding. *Remote Sensing*, *11*(24), 2952.

Persa, R., de Oliveira Ribeiro, P., & Jarquin, D. (2021). The use of high-throughput phenotyping in genomic selection context. crop breed app biotechnol 21: 1–11.

Peternelli, L. A., & Andrade, A. C. B. (2023). Insights and protocols for discrimination of sugarcane clones by dissimilarity measures on rgb and nir data. *PloS one*, *18*(7), e0288508.

Poudyal, C., Costa, L. F., Sandhu, H., Ampatzidis, Y., Odero, D. C., Arbelo, O. C., & Cherry, R. H. (2022). Sugarcane yield prediction and genotype selection using unmanned aerial vehicle-based hyperspectral imaging and machine learning. *Agronomy Journal*, *114*(4), 2320–2333.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Radočaj, D., Šiljeg, A., Marinović, R., & Jurišić, M. (2023). State of major vegetation indices in precision agriculture studies indexed in web of science: A review. *Agriculture*, *13*(3), 707.

Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., Reynolds, M., & Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genetics*, *6*(9), 2799–2808.

Sanches, G. M., Duft, D. G., Kölln, O. T., Luciano, A. C. d. S., De Castro, S. G. Q., Okuno, F. M., & Franco, H. C. J. (2018). The potential for rgb images obtained using unmanned aerial vehicle to assess and predict yield in sugarcane fields. *International journal of remote sensing*, *39*(15-16), 5402–5414.

Sandhu, K. S., Mihalyov, P. D., Lewien, M. J., Pumphrey, M. O., & Carter, A. H. (2021). Combining genomic and phenomic information for predicting grain protein content and grain yield in spring wheat. *Frontiers in plant science*, *12*, 613300.

Shahi, D., Guo, J., Pradhan, S., Khan, J., Avci, M., Khan, N., McBreen, J., Bai, G., Reynolds, M., Foulkes, J., et al. (2022). Multi-trait genomic prediction using in-season physiological parameters increases prediction accuracy of complex traits in us wheat. *BMC genomics*, *23*(1), 298.

Som-ard, J., Hossain, M. D., Ninsawat, S., & Veerachitt, V. (2018). Pre-harvest sugarcane yield estimation using uav-based rgb images and ground observation. *Sugar Tech*, *20*(6), 645–657.

Sumesh, K., Ninsawat, S., & Som-Ard, J. (2021). Integration of rgb-based vegetation index, crop surface model and object-based image analysis approach for sugarcane yield estimation using unmanned aerial vehicle. *Computers and Electronics in Agriculture*, *180*, 105903.

Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J.-L., & Sorrells, M. E. (2017). Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *The plant genome*, *10*(2), plantgenome2016–11.

Tattaris, M., Reynolds, M. P., & Chapman, S. C. (2016). A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Frontiers in plant science*, *7*, 1131.

Todd, J., & Johnson, R. (2021). Prediction of ratoon sugarcane family yield and selection using remote imagery. *Agronomy*, *11*(7), 1273.

Todd, J., Johnson, R., Verdun, D., & Richard, K. (2022). Identification of selection preferences and predicting yield related traits in sugarcane seedling families using rgb spectral indices. *Agriculture*, *12*(9), 1313.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, *8*(2), 127–150.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, *91*(11), 4414–4423.

Velazco, J. G., Malosetti, M., Hunt, C. H., Mace, E. S., Jordan, D. R., & Van Eeuwijk, F. A. (2019). Combining pedigree and genomic information to improve prediction quality: An example in sorghum. *Theoretical and Applied Genetics*, *132*, 2055–2067.

Woebbecke, D., Meyer, G., Von Bargen, K., & Mortensen, D. (1995). Shape features for identifying young weeds using image analysis. *Transactions of the ASAE*, *38*(1), 271–281.

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

Yadav, S., Wei, X., Joyce, P., Atkin, F., Deomano, E., Sun, Y., Nguyen, L. T., Ross, E. M., Cavallaro, T., Aitken, K. S., et al. (2021). Improved genomic prediction of clonal performance in sugarcane by exploiting non-additive genetic effects. *Theoretical and Applied Genetics*, *134*(7), 2235–2252.

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Frontiers in plant science*, *8*, 1111.

Zhao, D., Gordon, V. S., Comstock, J. C., Glynn, N. C., & Johnson, R. M. (2016). Assessment of sugarcane yield potential across large numbers of genotypes using canopy reflectance measurements. *Crop Science*, *56*(4), 1747–1759.

Zhou, M. (2014). Family evaluation for sugarcane yield using data estimated from stalk number, height, and diameter. *Journal of Crop Improvement*, *28*(3), 406–417.