



OPEN ACCESS

EDITED BY

Rodomiرو Ortiz,
Swedish University of Agricultural
Sciences, Sweden

REVIEWED BY

Just Jensen,
Aarhus University, Denmark
Maryke T. Labuschagne,
University of the Free State,
South Africa

*CORRESPONDENCE

Marcos Deon Vilela de Resende
✉ marcos.resende@embrapa.br
Eder Jorge de Oliveira
✉ eder.oliveira@embrapa.br

SPECIALTY SECTION

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

RECEIVED 15 October 2022

ACCEPTED 02 December 2022

PUBLISHED 16 December 2022

CITATION

de Andrade LRB, Sousa MBe, Wolfe M,
Jannink J-L, de Resende MDV,
Azevedo CF and de Oliveira EJ (2022)
Increasing cassava root yield: Additive-
dominant genetic models for
selection of parents and clones.
Front. Plant Sci. 13:1071156.
doi: 10.3389/fpls.2022.1071156

COPYRIGHT

© 2022 de Andrade, Sousa, Wolfe,
Jannink, de Resende, Azevedo and de
Oliveira. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Increasing cassava root yield: Additive-dominant genetic models for selection of parents and clones

Luciano Rogério Braatz de Andrade ¹,
Massaine Bandeira e Sousa ², Marnin Wolfe³,
Jean-Luc Jannink ^{4,5}, Marcos Deon Vilela de Resende ^{6,7,8*},
Camila Ferreira Azevedo ⁸ and Eder Jorge de Oliveira ^{2*}

¹Department of Crop Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil,

²Embrapa Mandioca e Fruticultura, Cruz das Almas, Bahia, Brazil, ³Department of Crop, Soil and
Environment Sciences, Auburn University, Auburn, AL, United States, ⁴Section on Plant Breeding
and Genetics, School of Integrative Plant Sciences, Cornell University, Ithaca, NY, United States,

⁵United States Department of Agriculture – Agriculture Research Service, Plant, Soil and Nutrition
Research, Ithaca, NY, United States, ⁶Department of Forestry Engineering, Universidade Federal de
Viçosa, Viçosa, Minas Gerais, Brazil, ⁷Embrapa Florestas, Colombo, Paraná, Brazil, ⁸Department of
Statistics, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

Genomic selection has been promising in situations where phenotypic assessments are expensive, laborious, and/or inefficient. This work evaluated the efficiency of genomic prediction methods combined with genetic models in clone and parent selection with the goal of increasing fresh root yield, dry root yield, as well as dry matter content in cassava roots. The bias and predictive ability of the combinations of prediction methods Genomic Best Linear Unbiased Prediction (G-BLUP), Bayes B, Bayes C π , and Reproducing Kernel Hilbert Spaces with additive and additive-dominant genetic models were estimated. Fresh and dry root yield exhibited predominantly dominant heritability, while dry matter content exhibited predominantly additive heritability. The combination of prediction methods and genetic models did not show significant differences in the predictive ability for dry matter content. On the other hand, the prediction methods with additive-dominant genetic models had significantly higher predictive ability than the additive genetic models for fresh and dry root yield, allowing higher genetic gains in clone selection. However, higher predictive ability for genotypic values did not result in differences in breeding value predictions between additive and additive-dominant genetic models. G-BLUP with the classical additive-dominant genetic model had the best predictive ability and bias estimates for fresh and dry root yield. For dry matter content, the highest predictive ability was obtained by G-BLUP with the additive genetic model. Dry matter content exhibited the highest heritability, predictive ability, and bias estimates compared with other traits. The prediction methods showed similar selection gains with approximately 67% of the phenotypic selection gain. By shortening the breeding cycle time by 40%, genomic selection may overcome phenotypic selection by 10%, 13%, and 18% for fresh root yield, dry root yield, and dry

matter content, respectively, with a selection proportion of 15%. The most suitable genetic model for each trait allows for genomic selection optimization in cassava with high selection gains, thereby accelerating the release of new varieties.

KEYWORDS

genomic selection, non-additive effects, dominance, breeding, breeding values, genotypic values

1 Introduction

Cassava (*Manihot esculenta* Crantz) has great social and economic importance for Brazilian agriculture, where nearly 18.2 million tons were produced across 1.2 million hectares in 2020 (FAO, 2022). Most of the planted area is within small farms where the product is destined for on-farm consumption or local sales. However, with the starch price rising, there is a trend of increasing industry involvement in intensive cassava production. Although almost the entire plant can be used for human and animal consumption, farmers have chiefly focused on root production.

Cassava can be propagated by seeds or vegetatively by stem pieces (cuttings), with the former generally limited to breeding programs for allele recombination and generation of new hybrid combinations and the latter the most common method used by farmers for multiplication and root production (Ceballos et al., 2012). Once the F₁ population is obtained, the hybrids are evaluated and selected regularly through several stages. Selection intensity and the evaluated traits depend on the amount of propagation material and the evaluation potential in different environments. According to Barandica et al. (2016), until the 21st century, hybrid selection in early-phase breeding programs was performed visually without extensive phenotypic data collection. Therefore, until relatively recently, inheritance knowledge about relevant traits was very limited (Calle et al., 2005; Zacarias and Labuschagne, 2010; Ceballos et al., 2012; Tumuhimbise et al., 2014; Oliveira et al., 2015a).

In several phases of the breeding program, vegetative propagation allows the maintenance of high heterozygosity and phenotypic plasticity expression for several traits (Oliveira et al., 2015a). In addition, it allows hybrids to be evaluated and selected in different locations and crop seasons (Barandica et al., 2016), thus allowing the separation of genetic and environmental effects, through the effects of the genotype by environment interactions (Ceballos et al., 2016a; Bakare et al., 2022). Due to vegetative propagation and the high heterozygosity of the parents (Ceballos et al., 2016a), genetic variability within families represents approximately 90% of total genetic variability (Ceballos et al., 2016b), supporting the idea that elite clones can be obtained within any family.

One hypothesis that may explain this high intra-family variability is the presence of non-additive genetic effects, especially for yield traits (Calle et al., 2005; Jaramillo et al., 2005; Zacarias and Labuschagne, 2010; Parkes et al., 2013; Tumuhimbise et al., 2014). While the non-additive effects hamper clone and parent selection, they allow for exploration of heterosis, as the best hybrids can be multiplied by vegetative propagation and then be release as new varieties (Parkes et al., 2013). However, the low correlation between root yield performance in the initial and final stages of the breeding program prevents the early and accurate selection of the best hybrids in clonal evaluation trials (Barandica et al., 2016). As a result, large seedling populations are evaluated annually and selected for the next stages (Ceballos et al., 2012), with the goal of identifying the most promising genotypes in advanced phases of the breeding program. This greatly increases the costs of the variety development pipeline, as phenotypic measurements demand suitable infrastructure, skilled labor, and consequently large amounts of financial resources.

Progress in genotyping, especially in reducing costs and increasing marker density, is revolutionizing marker applications in plant breeding (Ferguson et al., 2012). Since Meuwissen et al. (2001), there have been high expectations of genomic selection implementation in multiple breeding programs, due to possible selection gain in situations where traditional evaluation methods are expensive, laborious, and/or inefficient (Crossa et al., 2013). In genomic selection, breeding populations are phenotyped and genotyped with high genomic coverage markers in order to allow prediction methods to predict genomic estimated breeding values (GEBVs) of each clone (Ferguson et al., 2012). According to Crossa et al. (2013), genomic selection can predict clones' breeding values to accelerate recombination and their genotypic values as a means of targeting clones for advancement in the breeding pipeline.

For cassava, there is an expectation of genomic selection use for early selection in seedling trials as an alternative method to select traits that are difficult to measure or that demand high experimental accuracy (Oliveira et al., 2012), such as fresh root yield (FRY) and starch yield. In general, yield traits have predominantly non-additive effects (Jaramillo et al., 2005; Zacarias and Labuschagne, 2010; Parkes et al., 2013;

Tumuhimbise et al., 2014) and low correlation of the phenotypic values obtained at initial phases (seedling and clonal evaluation trials) with those of advanced trials (uniform yield trials) (Barandica et al., 2016). Another trait of great importance in cassava is the dry matter content (DMC) in roots; its genetic heritability has predominantly been associated with additive effects (Jaramillo et al., 2005; Parkes et al., 2013; Tumuhimbise et al., 2014; Wolfe et al., 2016a), and high correlation between the different breeding program stages (Barandica et al., 2016). As a result, clone and parent selection in the seedling trials is less accurate for yield traits than for DMC. However, early selection for DMC may also increase breeding efficiency, even though phenotyping in seedling trials is time-consuming and laborious. This is because seedling trials involve the evaluation of thousands of clones, and there is limited root production per clone, which prevents the use of a simple method of evaluation (specific gravimetry).

When only the additive effects are considered in the parent selection, the progeny mean is equal to the mean of the parents' breeding values; however, dominant effects prediction allows for heterosis exploration through parent complementarity (Almeida Filho et al., 2016). The genomic prediction of non-additive effects incorporated into genetic models increases the accuracy in parent and clone selection for low inheritance traits, as was observed in interspecific hybrid selection in *Eucalyptus* (Tan et al., 2018), intraspecific hybrids of *Pinus taeda* (Almeida Filho et al., 2016), maize (Lyra et al., 2019), inbred lines and crossbreed selection in Landrace and Yorkshire pigs (Esfandyari et al., 2016), and in clone selection of cassava (Wolfe et al., 2016a).

Genomic selection was also efficiently applied for predicting resistance to cassava mosaic disease, which displays a predominantly additive inheritance (Parkes et al., 2013; Tumuhimbise et al., 2014). In two years (annual breeding cycle), the allelic frequency of the marker with the greatest effect on cassava mosaic disease resistance rapidly increased from 44% to 66% (Wolfe et al., 2016b), much faster than the five or six years required in a conventional breeding cycle. Oliveira et al. (2012) noted that the two-year breeding cycle may have resulted in genetic gains higher than the conventional breeding cycle, of 56.9% and 39.92% for FRY and DMC, respectively.

Other important genomic selection goals are breeding population size reduction, time required to develop a new variety, and the ability to grow breeding populations outside the variety's recommended location, allowing selection for biotic and abiotic disturbances outside the endemic region (Ferguson et al., 2012). New prediction methodologies are consistently being published (Meuwissen et al., 2001; Park and Casella, 2008; Habier et al., 2011; Legarra et al., 2011; Azevedo et al., 2015; Wolfe et al., 2021). Application of the appropriate methodology to a trait of interest may increase selection gains and simultaneously reduce the work required in phenotypic evaluations, which are mostly high in cost and low in yield

(Ferguson et al., 2012). Wolfe et al. (2016a) have noted that a non-additive genomic relationship matrix may contribute to increased efficiency and yield in clone selection for traits with low heritability and/or that are difficult to measure.

Several studies have explored the efficiency of additive models of genomic selection. However, few have addressed the efficiency of dominant effects incorporated in genetic models for cassava breeding. Therefore, the objective of this work was to infer the efficiency of the G-BLUP, Bayes B, and RKHS genomic prediction methods with different genetic models for clone and parent selection to increase FRY, dry root yield (DRY) and DMC. Breeding program stages and genomic selection that may increase the efficiency of cassava breeding programs are also discussed.

2 Material and methods

2.1 Training population

The training population included 888 accessions belonging to the Cassava Germplasm Bank of Embrapa Cassava and Fruits (Cruz das Almas, Bahia, Brazil). This germplasm comprised 835 landraces and 53 improved varieties. One hundred and eighty accessions were characterized as sweet cassava (< 50 ppm of cyanogenic compounds), 136 as containing intermediary cyanide content (50–100 ppm cyanogenic compounds), 560 as bitter cassava (> 100 ppm cyanogenic compounds), and 12 as unclassified. These accessions were collected from all 26 Brazilian states, with every state represented by at least one genotype. The genotypes were evaluated in the cities of Cruz das Almas and Laje in the state of Bahia, Brazil, in 21 trials over a six-year period (2011 to 2016).

2.2 Phenotypic data collection

For most experiments, 15–20 cm stem cuttings were planted in double lines during the rainy season in the region (May–July). The experimental plot consisted of two rows of eight plants per row. The rows were 0.9 m apart, while plants in the same row were 0.8 m apart, with 11.52 m² per plot. All recommended cassava cultural practices were employed (as in Souza et al., 2006). Trials were harvested 11–12 months after planting. The traits measured to estimate genomic selection efficiency were: 1) fresh root yield (FRY) at plot level (16 plants) and then adjusted to t.ha⁻¹, 2) dry matter content in the roots (DMC), according to Kawano et al. (1987), where approximately 5 kg of roots were weighed in a hanging scale (WA) and then, the same sample was weighed with the roots submerged in water (WW). DMC was estimated utilizing the following formula: $DMC(\%) = (\frac{WA}{WA-WW} \times 158.3) - 142$ and 3) dry root yield (DRY) in t.ha⁻¹, estimated per plot by multiplying the FRY and DMC.

A joint analysis of 21 trials with complete randomized block design or augmented block design were used to obtain the phenotypic data. Three replicates were used in the complete randomized block design, while in the augmented block design, 10–16 replicates of the common checks were used, with equal distribution of accession number per block. Improved clones (9602-02, 9607-07, 9824-09, 9655-02) and improved varieties (BRS Dourada, BRS Gema de Ovo, and BRS Novo Horizonte) were used as checks in different field trials. More details from the phenotypic dataset could be seen in Table S1 and S2.

Due to unbalanced trials, we obtained the BLUP and deregressed BLUP (Garrick et al., 2009) for each clone. The BLUPs were obtained by the following mixed linear model: $y_{ijl} = \mu + c_i + \beta_j + r_{l(j)} + \epsilon_{ijl}$ in which y_{ijl} is the vector of phenotypic observations; c_i is the clone random effect with $c_i \sim N(0, I\hat{\sigma}_c^2)$; β_j is the combination of location and year, assumed as fixed effect; $r_{l(j)}$ is the replication nested within location and year, assumed as random effect with $r_{l(j)} \sim N(0, I\hat{\sigma}_r^2)$ and ϵ_{ijl} is the residual with $\epsilon_{ijl} \sim N(0, I\hat{\sigma}_e^2)$. The deregressed BLUPs were estimated by: *deregressed BLUP* = $\frac{BLUP}{1 - \frac{PEV}{\hat{\sigma}_e^2}}$ (Garrick et al., 2009), where the PEV is the prediction error variance of each clone and $\hat{\sigma}_e^2$ is the clonal variance component. The package lme4 (Bates et al., 2015) in R software version 3.5.2 (R Core Development Team, 2018) was used to obtain the BLUPs and deregressed BLUPs for each clone.

2.3 Genotyping and SNP quality control

DNA was extracted from cassava leaves following the CTAB (cetyltrimethylammonium bromide) protocol described by Doyle and Doyle (1987). To evaluate DNA integrity and standardize its concentration, 1.0% (w/v) agarose gels were stained with ethidium bromide (1.0 mg L⁻¹) for visual comparison of a series of DNA phage Lambda (Invitrogen) concentrations. The DNA samples were sent to the Genomic Diversity Facility at Cornell University (<http://www.biotech.cornell.edu/brc/genomic-diversity-facility>) for genotyping-by-sequencing (GBS) (Hamblin and Rabbi, 2014). Genotypic data were selected using a minimum call rate of 0.90 and the missing markers were imputed by Beagle 4.1 software (Browning and Browning, 2016). Finally, SNPs with minor allele frequency (MAF) > 0.05 were retained. After applying marker quality control, 48,655 SNPs were selected for genomic prediction.

2.4 Genomic selection methods and genetic models

The genomic best linear unbiased prediction (G-BLUP), Reproducing Kernel Hilbert Spaces (RKHS), and Bayes B prediction methods were evaluated, considering the additive (A) and additive-dominant (A+D) genetic models, except RKHS, which predicts genetic effects based on non-parametric

—and thus neither additive nor dominance—covariances. The additive-dominant genetic model of G-BLUP is expressed: $y_d = J\mu + Za + Hd + \epsilon$ where y_d is the deregressed BLUP vector; μ is the general mean; a is the additive effect vector, random $a \sim N(0, G\hat{\sigma}_a^2)$; d is the dominant deviation effect vector, random $d \sim N(0, D\hat{\sigma}_d^2)$; ϵ is the residual effect vector, $\epsilon \sim N(0, I\hat{\sigma}_e^2)$; Z and H are the incidence matrices for μ , a and d , respectively, as $COV(a, d) = 0$. The additive relationship matrix G was: $G = \frac{ZZ'}{\sum p_i(1-p_i)}$ in which Z is the marker matrix (-1, 0 and 1) and p_i is the major allele frequency of i marker. Two additive-dominant genetic models were tested for the G-BLUP method, the Classical (Vitezica et al., 2013) and the Genotypic (Su et al., 2012), differing in the parameterization of the genomic relationship matrix due to dominance. The Classical dominant relationship matrix was parameterized by the following Vitezica et al. (2013):

$$H = \begin{cases} \text{if } MM : -q^2 \\ \text{if } Mm : 2pq \\ \text{if } mm : -p^2 \end{cases}$$

$$D = \frac{HH'}{2 \sum p_i q_i (1 - p_i q_i)}$$

The Genotypic dominant relationship matrix was estimated by the following equation (Su et al., 2012):

$$H^* = \begin{cases} \text{if } MM : -2pq \\ \text{if } Mm : p^2 + q^2 \\ \text{if } mm : -2pq \end{cases}$$

$$D^* = \frac{H^*H^{*'}}{2 \sum p_i q_i (1 - p_i q_i)}$$

For the Bayes B method, the complete conditional prior distribution was used: $y_d | a_j, d_j, Z_{i \times j}, H_{i \times j} \sim N(\mu + \sum Z_{i \times j} a_j + \sum H_{i \times j} d_j, \hat{\sigma}_e^2)$ in which y_d is the deregressed BLUP vector; μ is the general mean; a_j and d_j are the additive and dominant marker effects, both random $a_j | \hat{\sigma}_{a_j}^2 \sim N(0, I\hat{\sigma}_{a_j}^2)$ $d_j | \hat{\sigma}_{d_j}^2 \sim N(0, I\hat{\sigma}_{d_j}^2)$ and $COV(a, d) = 0$. Z and H are the incidence matrix of a_j and d_j respectively.

The model of the RKHS method was: $y_d = J\mu + Xg + \epsilon$ where y_d is the deregressed BLUP vector; μ is the general mean; g is the genotypic effect vector, random $g \sim N(0, K\hat{\sigma}_g^2)$; ϵ is the residual effect vector, $\epsilon \sim N(0, I\hat{\sigma}_e^2)$; X and K are the incidence matrix of μ and g , respectively. K is a gaussian matrix estimated by: $K = \exp(\frac{-hD}{\text{median}(D)})$ h is the reduction coefficient to K values (in this work h was equal to 1), and D is the Euclidian distance of Z codified marker matrix (Gianola et al., 2006; Crossa et al., 2010).

The 5-fold cross-validation with three repetitions was performed to estimate the following parameters: 1) predictive ability ($\hat{r}_{\hat{y}_y} = \hat{C\hat{O}R}(\hat{P}red_{val}, BLUP_{val})$) in which $\hat{P}red_{val}$ are the genomic estimated breeding values (GEBVs) for additive genetic models, or genomic estimated genotypic values (GEGVs) for

additive-dominant and RKHS models, and $BLUP_{Val}$ are the BLUPs from the validation population; 2) bias ($\hat{b} = COV(Pred_{Train}, BLUP_{Train}) / \hat{\sigma}_{Pred_{Train}}^2$) in which $Pred_{Train}$ are the genomic estimated breeding values (GEBVs) for additive genetic models, or genomic estimated genotypic values (GEGVs) for additive-dominant genetic models, of the training population, $BLUP_{Train}$ are the BLUPs from the training population, $\hat{\sigma}_{Pred_{Train}}^2$ is the variance of the GEBVs for additive genetic models, or genomic estimated genotypic values (GEGVs) for additive-dominant genetic models of the training population; 3) broad-sense genomic heritability ($\hat{H}^2 = \hat{\sigma}_g^2 / (\hat{\sigma}_g^2 + \hat{\sigma}_e^2)$) in which $\hat{\sigma}_g^2$ is the genomic variance, $\hat{\sigma}_e^2$ is the residual variance; 4) narrow-sense genomic heritability ($\hat{h}^2 = \hat{\sigma}_a^2 / (\hat{\sigma}_g^2 + \hat{\sigma}_e^2)$) in which $\hat{\sigma}_a^2$ is the additive genomic variance, $\hat{\sigma}_g^2$ is the genomic variance, $\hat{\sigma}_e^2$ is the residual variance. For each replicate of the cross-validation process, the population was split into five equal folds. Five genomic predictions were performed per fold used as test set (no phenotypes) each fold was predicted by the remaining four-folds training set (with phenotypes).

The *sommer* R package (Covarrubias-Pazarán, 2016) was used to fit the G-BLUP and RKHS models, while the *BGLR* R package (Perez and De Los Campos, 2014) was used to fit the Bayes B model. All methods were performed using R software version 3.5.2 (R Core Development, 2018). For Bayes B method, we ran 20,000 Markov Chain Monte Carlo (MCMC) iterations with the burn-in of the initial 4,000 iterations and thinning of 10, we applied different priors for π for each trait and genetic model, these values were previously estimated by Bayes C π (Table S3).

The training-validation partitions of the population used in cross-validation were set up to be identical across prediction models, using the `set.seed()` function of R software version 3.5.2 (R Core Development, 2018). The residual variances of Markov Chain Monte Carlo (MCMC) of the Bayes B method were used to evaluate the MCMC convergence by the Raftery and Lewis's convergence diagnostic (Raftery and Lewis, 1992) applied in *coda* R package (Plummer et al., 2006).

2.5 Analysis of variance and Tukey's multiple comparison test

Analysis of variance was performed to estimate the effects of the genomic selection methods for predictive ability and bias estimates for DMC, FRY, and DRY. These analyses were performed using the *lme4* R package (Bates et al., 2015).

The following mixed model was used to estimate the efficiency of the genomic selection methods: $y_{ijk} = m_i + s_{jk} + e_{ijk}$ which y is the dependent variable, as predictive ability and bias; m_i is the mean of the genomic selection method I , assumed as fixed effect; s_{jk} is the effect of cross validation of the replication j and fold k , assumed as random effect $s \sim (0, \hat{\sigma}_{cv}^2)$ and e_{ijk} is the residual effect of the i genomic selection method of the j replication and k fold, $e \sim (0, \hat{\sigma}_e^2)$ The genomic prediction

means were submitted to the Tukey multiple comparison test implemented in the *emmeans* R package (Russel, 2018).

2.6 Cohen's Kappa coefficient

The Cohen's Kappa coefficient (Cohen, 1960) was used to analyze the coincidence of clone selection by the different genomic selection methods, considering a selection proportion (SP) amplitude ranging from 5–30%. The coincidence selection was performed using a binary code and the selected and unselected individuals received code "1" and "0", respectively. The Kappa coefficient and coincidences selection index were calculated using R.

3 Results

3.1 Efficiency of the genomic selection methods and genetic models

In general, the inclusion of the dominant genetic effects increased the genomic variance explained by the markers (Table 1), and reduced the genomic additive variance and residuals (Table 1 and Figure S1). Smaller changes in the broad-sense genomic heritability were observed for DMC, except for the Bayes B method, which demonstrated the highest broad-sense genomic heritability among the prediction methods with an additive-dominant genetic model.

Insert Table 1

A predominance of additive effects for DMC was identified with the G-BLUP method (Table 1 and Figure S1), while for FRY and DRY the dominant effects prevail. The Bayes B method showed the highest estimates of broad-sense genomic heritability and genomic variance components. However, the variation of the broad-sense genomic heritabilities between traits was smaller, suggesting a relatively large proportion of dominance variance. Even with the highest broad-sense genomic heritability, the Bayes B A+D method exhibited smaller narrow-sense genomic heritability than the G-BLUP A+D method, regardless of the dominant relationship matrix used (Table 1). However, all the additive-dominant genetic models overestimated the broad-sense genomic heritability because it was higher than the phenotypic heritability (0.337, 0.351, and 0.545 for FRY, DRY, and DMC, respectively).

The additive-dominant genetic models showed higher predictive ability than additive models and RKHS method for yield traits (FRY and DRY, Figure 1). The highest predictive ability was demonstrated by the G-BLUP A+D classical method (average of 0.484 for FRY and 0.492 for DRY), followed by Bayes B A+D (average of 0.479 for FRY and 0.488 for DRY). In addition, the predictions of dominant effects in genetic models for yield traits reduced the bias estimate, with the smaller bias at

TABLE 1 Means of the genetic parameters estimated by different genomic prediction methods for fresh root yield (FRY), dry root yield (DRY), and dry matter content (DMC) in roots of cassava.

Traits / Prediction methods	Genetic parameters					
Fresh root yield	\hat{h}^2	\hat{H}^2	$\hat{\sigma}_a^2$	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_e^2$
G-BLUP A ¹	0.347	–	17.0	–	17.0	32.0
G-BLUP A+D Classical ²	0.139	0.386	6.4	11.5	17.9	28.5
G-BLUP A+D Genotypic ³	0.053	0.400	2.6	16.8	19.4	29.1
RKHS	–	0.520	–	–	31.6	29.0
Bayes B A ⁴	0.582	–	43.9	–	43.9	31.2
Bayes B A+D ⁵	0.257	0.734	26.2	49.2	75.4	26.9
Dry root yield	\hat{h}^2	\hat{H}^2	$\hat{\sigma}_a^2$	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_e^2$
G-BLUP A ¹	0.332	–	1.39	–	1.39	2.81
G-BLUP A+D Classical ²	0.175	0.369	0.71	0.79	1.49	2.55
G-BLUP A+D Genotypic ³	0.096	0.381	0.40	1.20	1.60	2.59
RKHS	–	0.504	–	–	2.61	2.57
Bayes B A ⁴	0.571	–	3.69	–	3.69	2.74
Bayes B A+D ⁵	0.262	0.728	2.32	4.15	6.46	2.39
Dry matter content	\hat{h}^2	\hat{H}^2	$\hat{\sigma}_a^2$	$\hat{\sigma}_g^2$	$\hat{\sigma}_d^2$	$\hat{\sigma}_e^2$
G-BLUP A ¹	0.517	–	2.10	–	2.10	
G-BLUP A+D Classical ²	0.477	0.522	1.92	0.18	2.10	1.92
G-BLUP A+D Genotypic ³	0.457	0.525	1.86	0.27	2.13	1.92
RKHS	–	0.504	–	–	2.61	2.57
Bayes B A ⁴	0.673	–	4.04	–	4.04	1.95
Bayes B A+D ⁵	0.325	0.792	2.73	3.99	6.72	1.75

\hat{h}^2 , narrow-sense genomic heritability; \hat{H}^2 , broad-sense genomic heritability; $\hat{\sigma}_a^2$, additive genomic variance; $\hat{\sigma}_g^2$, dominant genomic variance; $\hat{\sigma}_d^2$, genomic variance; $\hat{\sigma}_e^2$, residual variance. ¹G-BLUP with additive model; ²G-BLUP with additive-dominant model, classical dominant relationship matrix (Vitezica et al., 2013); ³G-BLUP with additive-dominant model, genotypic dominant relationship matrix (Su et al., 2012); ⁴Bayes B with additive model; ⁵Bayes B with additive-dominant model.

Bayes B method (Figure 1). The RKHS method showed the highest bias estimates for all traits.

3.2 Analysis of variance and Tukey's multiple comparison test of the different genomic selection methods

Significant differences between the genomic selection methods with different genetic models were identified for predictive ability and bias for all agronomic traits except the predictive ability of DMC (Table 2). Although there were no significant differences in the predictive ability between the genomic selection methods with additive-dominant models, the G-BLUP A+D classical method showed the highest predictive ability for FRY (0.483) and DRY (0.492) (Table 2).

Bayes B A+D and RKHS methods did not show significant differences for predictive ability in comparison with G-BLUP A+D classical method for DRY. On the other hand, for FRY only Bayes B A+D and G-BLUP A+D genotypic methods did not show significant differences with the G-BLUP A+D classical method.

Among the methods with non-additive effects, the G-BLUP A+D classical was significantly different from the RKHS method for DRY but not for FRY. As the RKHS method can predict additive and partial epistatic effects (Gianola et al., 2006; Crossa et al., 2010), it is possible that the epistatic effects were more important for FRY than DRY, as the RKHS method did not show a significant difference with the additive genetic models G-BLUP A and Bayes B A (Table 2).

DMC showed the highest phenotypic heritability and predictive ability of traits. However, there was no improvement

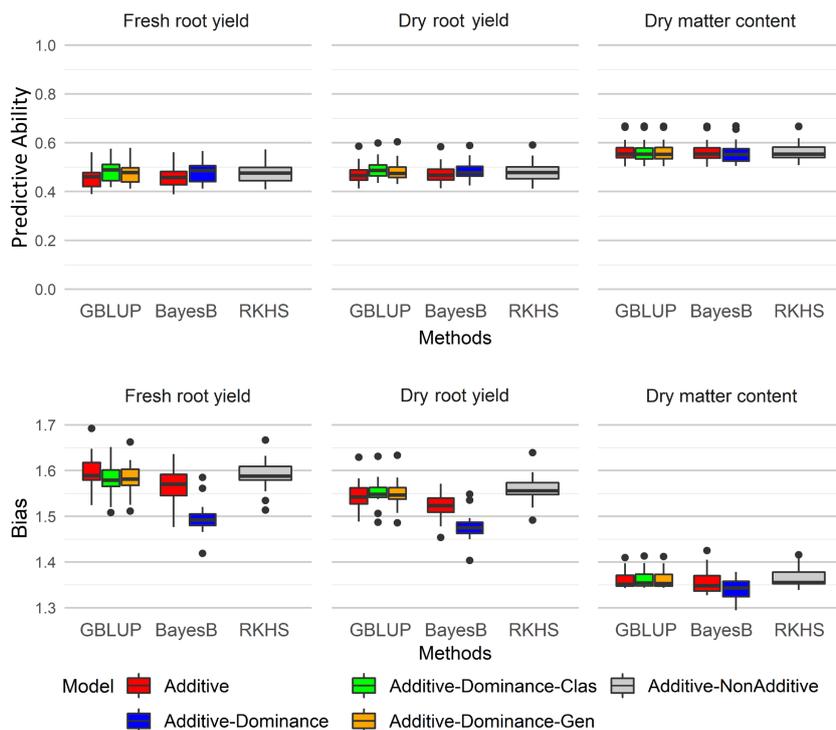


FIGURE 1
 SP: selection proportion; SD GS: genomic selection differential; SD PS: phenotypic selection differential; GB/PB: ratio between the breeding cycle assisted by genomic selection and conventional breeding cycle; Efficiency = $SD\ GS / [SD\ PS \times (GB/PB)]$ Boxplots of predictive ability and bias for different genomic selection methods (G-BLUP, Bayes B, and RKHS) with additive and additive-dominant genetic models for fresh root yield (FRY), dry root yield (DRY), and dry matter content (DMC). GBLUP, genomic best linear unbiased prediction; RKHS, reproducing kernel Hilbert spaces.

in predictive ability when the additive-dominant genetic models were used to predict this trait, which reinforced the theory that DMC in cassava has a high influence from additive effects. On the

other hand, for FRY and DRY, the additive-dominant models demonstrated increased predictive ability, suggesting a greater importance of dominant effects for these traits in cassava.

TABLE 2 Analysis of variance (ANOVA) and Tukey’s multiple comparison test ($p \leq 0.05$) for prediction parameters of different genomic selection methods for fresh root yield (FRY), dry root yield (DRY), and dry matter content (DMC) in cassava.

ANOVA	DF	Fresh root yield		Dry root yield		Dry matter content	
		$\hat{r}_{\hat{y}y}$	\hat{b}	$\hat{r}_{\hat{y}y}$	\hat{b}	$\hat{r}_{\hat{y}y}$	\hat{b}
Methods	5	21.91*	51.76*	10.95*	50.35*	2.07	9.28*
Tukey multiple comparison test							
Bayes B A		0.458C	1.568B	0.474C	1.522B	0.566A	1.357B
Bayes B A+D		0.479AB	1.497A	0.488AB	1.477A	0.561A	1.340A
G-BLUP A		0.457C	1.598C	0.474C	1.547C	0.567A	1.360B
G-BLUP A+D Classical		0.483A	1.580BC	0.492A	1.552C	0.564A	1.362B
G-BLUP A+D Genotypic		0.474B	1.582BC	0.485AB	1.550C	0.565A	1.361B
RKHS		0.476AB	1.590C	0.482BC	1.560C	0.567A	1.366B

$\hat{r}_{\hat{y}y}$, predictive ability; \hat{b} , bias, DF, degrees of freedom. *significant by chi-square test ($p \leq 0.05$). Upper case letters means significant difference between genomic selection methods for the Tukey multiple comparison test ($p \leq 0.05$): predictive ability; bias, DF: degrees of freedom. *significant by chi-square test ($p \leq 0.05$). Upper case letters means significant difference between genomic selection methods for the Tukey multiple comparison test ($p \leq 0.05$).

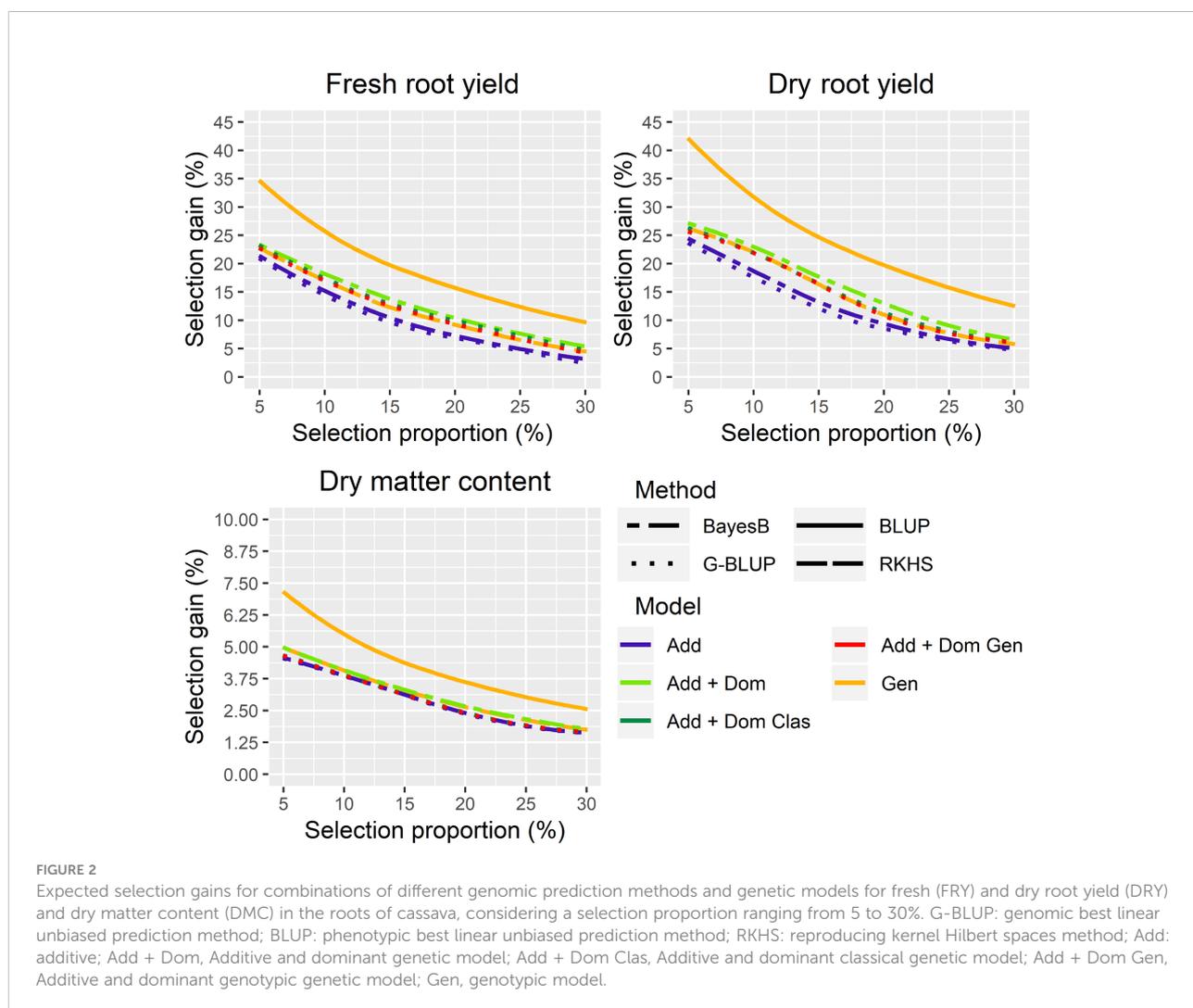
3.3 Expected genetic gains from different genomic prediction methods through different selection proportion

Although significant differences were detected between genomic prediction methods with different genetic models by ANOVA and Tukey's mean test (Table 2), the expected genetic gains for genomic prediction were still smaller than those obtained by phenotypic selection, with expected selection gains equivalent to 67.5%, 67.1%, and 69.4% of the phenotypic selection for FRY, DRY, and DMC, respectively (Figure 2). Although selection gains with genomic predictions were similar for all traits, the non-additive genetic models, such as Bayes B A+D, RKHS, and G-BLUP A+D classical and genotypic, increased the gain by an average of 0.69 t/ha for FRY and 0.24 t/ha for DRY in comparison with the additive genetic models. For DMC, the differences between the selection gains of genomic prediction methods were lower (average of 0.04%), because there

was no significant difference between the clone prediction methods for this trait (Table 2). Moreover, the selection differential for DMC in the roots was lower than for yield traits due to the smaller trait amplitude (17–38%).

There was a great uniformity in the differences between the selection gains of the phenotypic BLUP and the predicted gains in the different selection proportions, with a mean difference of selection gain of 6.18% and 7.79% of the Bayes B A+D model for FRY and DRY, respectively (Figure 2 and Table S2). For DMC, there were lower gains differences between the phenotypic selection and genomic prediction, with the largest difference observed in the G-BLUP A method (average of 1.40% of genetic gain) compared to others (Figure 2 and Table S4).

The genomic expected selection gain and its relative efficiency to phenotypic expected selection gain were calculated. According to Oliveira et al. (2012), the conventional breeding cycle of cassava is at least four years due to the need to include phenotypic information from a minimum of four breeding phases (clonal evaluation trial,



preliminary yield trial, advanced yield trial, and uniform yield trial). The efficiency and the selection gains per time unit to simulate early selection assisted by genomic selection were calculated. The efficiency was determined by comparing time required to recombine the selected clones as parents in a conventional breeding program vs. one assisted by genomic selection.

Genomic selection based on the G-BLUP A+D classical method for FRY and G-BLUP A for DMC was more efficient than phenotypic selection when the breeding cycle was ≤ 0.60 of the conventional breeding cycle (Table 3 and Figure S2). However, for DRY the genomic selection was more efficient than phenotypic selection only with a selection proportion of 5–15%.

Reducing breeding cycle time by 60% using genomic selection could result in gains of 65%, 69%, and 77% over those provided by phenotypic selection for FRY, DRY, and DMC, respectively, in a selection proportion of approximately 15% of the best clones (Table 3). If the breeding cycle was reduced to 20% of the conventional breeding cycle (four years to ten months), the genetic gains would be 163%, 170%, and 183% over those provided by phenotypic selection for FRY, DRY, and DMC, respectively.

The selection proportion affected significantly the relative efficiency of the genomic prediction only in breeding cycle time reductions biggest than 35% of the conventional Cassava breeding cycle (Table 3).

4 Discussion

4.1 Phenotypic and genomic heritability and its implications for genomic selection

According to Oliveira et al. (2015b), heritability estimates can assist selection strategies in increasing genetic gain, as well as defining the breeding method and experimental design. Given the broad- and narrow-sense genomic heritability, the G-BLUP A+D classical method showed that cassava yield traits demonstrate a predominance of dominant effects. In addition, the broad-sense genomic heritability of G-BLUP A+D was closer to the phenotypic heritability values (0.337 for FRY, 0.351 for DRY, and 0.545 for DMC [Table 1]). Stability of FRY and DRY are important agronomic attributes for any cassava variety to ensure high market competitiveness in the starch industry, especially as there is a minimum acceptable DMC threshold for processing the raw material. Roots with DMC index below this threshold are not processed by the starch industry due to the high industrial cost and low starch yield.

Knowledge about trait heritability and variation gained during field evaluation in different environments may assist in optimizing selection of cassava breeding programs, with the goal of developing new cassava varieties with higher starch yield stability. Optimizing the selection proportion and evaluated traits in each breeding phase can maximize the probability of selecting the best clone. This is because low heritability traits such as FRY and starch yield are generally evaluated in the final breeding phases due to greater stem cutting availability (more plants per plot across multiple locations).

TABLE 3 Relative efficiency of genomic selection compared to phenotypic selection using different selection proportions with the G-BLUP A+D classical method for fresh root yield (FRY), dry root yield (DRY), and dry matter content (DMC) in cassava.

Method	Fresh root yield						Dry root yield						Dry matter content					
	G-BLUP		Additive		Dominant Classical		G-BLUP		Additive		Dominant Classical		G-BLUP			Additive		
SP (%)	5%	10%	15%	20%	25%	30%	5%	10%	15%	20%	25%	30%	5%	10%	15%	20%	25%	30%
SD GS	23.8	17.2	13.0	9.9	7.4	4.7	25.8	22.0	16.7	11.3	8.4	5.8	4.6	3.9	3.1	2.4	1.9	1.6
SD PS	35.1	25.4	19.8	15.7	12.4	9.6	42.8	31.7	24.7	19.7	15.8	12.5	7.3	5.5	4.4	3.6	3.0	2.5
GB/PB	Efficiency						Efficiency						Efficiency					
1.00	0.68	0.68	0.66	0.63	0.60	0.49	0.60	0.69	0.68	0.57	0.53	0.47	0.64	0.71	0.71	0.67	0.62	0.62
0.80	0.85	0.84	0.82	0.79	0.75	0.61	0.75	0.87	0.84	0.72	0.66	0.59	0.79	0.88	0.89	0.83	0.78	0.78
0.65	1.04	1.04	1.01	0.97	0.92	0.75	0.93	1.07	1.04	0.88	0.82	0.72	0.98	1.09	1.09	1.02	0.96	0.96
0.60	1.13	1.13	1.10	1.06	1.00	0.81	1.00	1.16	1.13	0.95	0.88	0.78	1.06	1.18	1.18	1.11	1.04	1.04
0.40	1.70	1.69	1.65	1.58	1.50	1.22	1.51	1.74	1.69	1.43	1.33	1.17	1.59	1.77	1.77	1.66	1.56	1.56
0.20	3.40	2.70	2.63	2.53	2.40	1.95	2.41	2.78	2.70	2.29	2.12	1.87	2.54	2.83	2.83	2.66	2.49	2.49

SP, selection proportion; SD GS, genomic selection differential; SD PS, phenotypic selection differential; GB/PB, ratio between the breeding cycle assisted by genomic selection and conventional breeding cycle; Efficiency = SD GS/[SD PS×(GB/PB)].

Wolfe et al. (2016b) also related the predominance of additive and dominant deviation effects for DRY and FRY, respectively. They found similar estimates of broad- and narrow-sense heritability for the first genomic selection cycle of IITA population using the G-BLUP A+D method (0.12 and 0.35 for narrow- and broad-sense heritability, respectively, for FRY, and 0.47 and 0.52 for narrow and broad-sense heritability, respectively, for DMC). Wolfe et al. (2016b) found that the appropriate genetic model for DMC was the additive-dominant, while in the present study the additive-dominant models obtained similar results to the other models'. The reduction of the variance explained by the additive component was noted previously in cassava (Wolfe et al., 2016b) and other species such as *Pinus taeda* L. (Muñoz et al., 2014) and hybrids of *Eucalyptus urophylla* and *E. grandis* (Bouvet et al., 2016). Several authors reported that during prediction using additive genetic models, part of the dominant deviation was predicted along with the additive effects; however, when using additive-dominant genetic models, this dominant deviation predicted by the additive effects is then computed by the dominant variance (Zuk et al., 2012; Vitezica et al., 2013; Muñoz et al., 2014; Wolfe et al., 2016a). According to Vitezica et al. (2013) genetic models with assumptions of additive and dominant deviation effects result in better genomic predictions.

4.2 Efficiency of cassava selection considering different genomic prediction methods and genetic models

There were significant differences in predictive ability between the methods for FRY and DRY, mainly due to different genetic models (additive and non-additive). The additive-dominant genetic models showed higher predictive ability than additive genetic models for FRY and DRY. Among the genomic selection methods, the G-BLUP A+D classical (Vitezica et al., 2013) had high predictive ability and low bias, statistically similar to other additive-dominant genetic models. Therefore, the additive-dominant genetic models allow for exploration of part of the non-additive effects by increasing cassava clone selection accuracy. Other authors have evaluated relationship matrices with classical and genotypic dominant models and verified the lack of differences in the genomic predictions of these matrices, although the broad-sense heritability has been somewhat lower in the matrix (H^*) of genotypic dominant (Vitezica et al., 2013; Wolfe et al., 2016a). However, the correlation between the additive and dominant parameters was higher in the G-BLUP genotypic method in comparison with the classical one (Wolfe et al., 2016a), which corroborates the correlations found in this work.

Dominance effects occurs due to the interaction between alleles at the same locus and its main benefits are expected in crossbreeding, since dominance has been suggested as one of the

genetic mechanisms explaining heterosis (Shull, 1908). Indeed, hybrid vigor for yield components in cassava over better-parent values has been reported (Parkes, et al., 2020), indicating that heterosis should be explored in order to develop superior cassava genotypes. Therefore, genomic predictions for traits such as FRY and DRY must be based on the assumption that non-additive effects are an important component that should be considered in the predictions to optimize crossing designs, such as in mate-pair allocation (Almeida Filho et al., 2019). As a further step, the role of dominance effects on the genetic architecture of FRY and DRY should be evaluated in the breeding population generated in this study.

For DMC, there was no significant difference between genetic models for predictive ability, although the additive genetic models showed better prediction abilities than additive-dominant models. Similar results were reported in the first genomic selection cycle of IITA cassava population (Wolfe et al., 2016a). According to Denis and Bouvet (2013) the decision of which genetic model to use in genomic selection depends on the training population as well the traits under selection. Specifically, in cassava, the genomic prediction of FRY and DRY would be more efficient if applying additive-dominant genetic models, while for DMC the additive models are satisfactory. Among genetic models, additive gene action had highest response to selection. Therefore, in the case of DMC, the population improvement focused on genetic additive effects can achieve large medium-to-long term genetic gains.

Incorporating non-additive effects into the genetic model reduces the additive genomic variance and the bias of the GEBVs, as well increasing the accuracy for selecting the best parent (Vitezica et al., 2013; Wolfe et al., 2016a). On the other hand, some simulated studies did not find any differences in predictive ability of GEBVs between additive and additive-dominant genetic models (Almeida Filho et al., 2016; Heidaritabar et al., 2016). Therefore, it is expected that non-additive effects prediction may increase the genetic gains for yield traits of new cassava varieties in the breeding programs (Muñoz et al., 2014; Wolfe et al., 2016a).

The expected selection gains for FRY and DRY were high due to the training population being composed of germplasm accessions with high genetic variability (Figure S2) for several traits, including yield traits (Oliveira et al., 2015a; Oliveira et al., 2016). Within a group of clones that deviated from the FRY and DRY mean (Figure S2), some were from high-yield, improved varieties (FRY potential of > 30 t/ha) such as BRS Novo Horizonte, BRS Poti Branca, BRS Kiriris, and BRS Tapioqueira.

According to the Bayes B A+D method, if the cassava breeding cycle was reduced by 40%, the genomic selection gains would be on average 12.48% and 11.92% higher than phenotypic selection gains for FRY and DRY, respectively. A similar observation was made for DMC (22.16%). Oliveira et al. (2012) reported that with a 25% reduction in breeding cycle time, the relative efficiency of RR-BLUP genomic prediction was

4.6%, 15.96%, and -7.05% for FRY, starch yield, and DMC, respectively. According to these authors, higher selection gains may be achieved by reducing the time required to identify and recombine the parents in the breeding cycle. These results may assist in the planning of genomic selection implementation to increase the frequency of new cassava varieties with good agronomic traits and adaptations to new biotic and abiotic stress challenges.

Reducing the selection proportion is not feasible as reducing breeding cycle time to improve genetic gain, but it may be the next milestone to improve genetic gain in cassava breeding, by increasing the number of clones evaluated in earlier stages by genomic prediction.

4.3 Potential application of genomic selection in cassava breeding

A previously recommended method for clone and parent selection in seedling trial phases was assessment of the harvest index (the ratio of FRY and the biomass yield), used for FRY indirect selection in seedling nursery trials and clonal evaluation trials (Kawano et al., 1998) due to its high correlation (0.730). However, when analyzing the historical data (2000–2013) of the cassava breeding program at the International Center for Tropical Agriculture (CIAT), Barandica et al. (2016) found very low correlation between FRY and harvest index (0.11). In addition, the harvest index assessment is more labor-intensive than the FRY evaluation alone. According to Barandica et al. (2016) the correlation of FRY between the clonal evaluation trials and the uniform yield trials was 0.29, while in the present study the correlation between the GEBVs and the uniform yield trials for the G-BLUP A+D classical method was 0.483. In future studies, the efficiency of genomic selection in the seedling trial phase for FRY could be better understood by determining realized genetic gains.

Ceballos et al. (2016a) stated that one issue in the selection of good parents is the high intra-family genetic variability due to high heterozygosity. Thus, new cassava varieties may derive from crosses between parents with low agronomic performance. Indeed, Kawano et al. (1998) evaluated almost 327,000 clones from 4,130 crosses during 14 years of research, and among all those evaluated clones only three were officially released as new varieties.

Commonly the standard methods used for parent selection are the *per se* performance and, less commonly, general combining ability (Ceballos et al., 2004). Unfortunately, there is no linear relationship between the *per se* performance and the progeny's breeding values due to dominant deviation (Ceballos et al., 2004). In addition, the diallel analysis in cassava breeding programs is problematic because the crosses in cassava are laborious and usually imbalanced due to issues of flowering synchronization (Ceballos et al., 2017), the considerable

unpredictability of the flowering season (between four to ten months after planting), and the time for seed maturity after harvest demanding at least one year to obtain the seeds of controlled crossings (Ceballos et al., 2004). Su et al. (2012) reported that genetic models with additive and non-additive effects prediction might allow for exploitation of specific combining ability. Therefore, applying genomic selection with genetic models that consider both genetic effects may be a faster alternative for selecting clones for advancement in the breeding pipeline, parents for crossings, inheritance studies, and variation of traits at the different stages of the cassava breeding program.

Another strategy for selecting promising parents is the pedigree-based best linear unbiased prediction (P-BLUP) method for predicting breeding values (Ceballos et al., 2016a). This strategy attempts to estimate breeding values after obtaining clone phenotypic data. According to Piepho et al. (2008) this method allows for dissection of the genotypic value in additive and non-additive effects, and it can be approached by identity-by-descent (P-BLUP) or identity-by-state (G-BLUP) information. However, this method requires a large amount of kinship information, which is not always available once several crosses have been carried out between germplasm accessions with no kinship data available. Nevertheless, the lack of kinship/pedigree information can be efficiently compensated for by identity-by-state (IBS) performed using an additive relationship matrix proposed by VanRaden (2007), as Hayes et al. (2009) considered the additive genomic relationship matrix as accurate as the kinship matrix. Bouvet et al. (2016) found that the G-BLUP prediction method had higher predictive ability than P-BLUP in several genetic models. According to Zhang et al. (2015) the GEBVs may be even more accurate when using a genetic architecture-enhanced relationship matrix for each trait, with the parametrization of relationship matrix composed by markers with high effect for the trait.

Using G-BLUP for breeding value estimation at preliminary, advanced and uniform yield trials, we assume that there is a genetic correlation between clones due to relationship-by-state (Piepho et al., 2008). Since cassava is vegetatively propagated, the additive genomic matrix may be used as a genetic covariance matrix for selecting promising parents by applying mixed models in the different breeding phases (such as the clonal evaluation trial, preliminary yield trial, advanced yield trial, and uniform yield trial). As the correlation between breeding values vs. genotypic values is not perfect (0.716 of selection coincidence at 13.3% selection proportion for FRY, and 0.690 of selection coincidence at 8.3% selection proportion for DRY), the coincidence in the selection of clones to be used as parents and for advancement in the breeding program tends to be low. Therefore, by using the genetic covariance matrix in mixed models, the selection of parents with high potential to generate promising clones would be performed based on their breeding values even if the clones had low genotypic value and/or low *per se* performance. This strategy can increase the parent selection

accuracy, estimate the narrow-sense heritability, and predict the GEBVs and GEGVs across the field trials, assisting parent and clone selection, respectively.

5 Conclusions

The genetic variances for FRY and DRY were largely derived from dominance deviations, while DMC was predominantly additive. Identification of the best genetic model allows breeders to achieve higher genetic gains in the cassava breeding program. Genomic selection can be used to assist in breeding value prediction and the selection of outstanding parents at early breeding steps, as well as to identify and select the genotypic value of good clones for advancement in the breeding pipeline. Genomic selection may achieve higher genetic gains by reducing the breeding cycle time by at least 40%.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://figshare.com/>, doi.org/10.6084/m9.figshare.21330972.

Author contributions

Study conception and design: LA, MR, EO. Data collection: LA, MS, EO. Analysis and interpretation of results: LA, MW, J-LJ, MR, EO. Draft manuscript preparation: LA, MS, MW. Final manuscript revision: MS, MW, CA, MR, EO. All authors reviewed the results and approved the final version of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

MS: Funarbe (Fundação Arthur Bernardes). Grant number: 4986. LA: Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Grant number: 001 - 88882.157009/2017-01. EO: CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). Grant number: 442050/2019-4 and 303912/2018-9. FAPESB (Fundação de Amparo à Pesquisa do Estado da Bahia). Grant number: Pronem 15/2014. This work was partially funded by UK's Foreign, Commonwealth & Development Office (FCDO) and the Bill & Melinda Gates Foundation. Grant INV-007637. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has

already been assigned to the Author Accepted Manuscript version that might arise from this submission. The funder provided support in the form of fellowship and funds for the research, but did not have any additional role in the study design, data collection and analysis, decision to publish, nor preparation of the manuscript.

Acknowledgments

Last, we would like to pay our gratitude and our respects to our colleague, Professor Fabyano Fonseca e Silva. After helping the genomic analysis and performed a deep revision of the manuscript, Professor Fabyano passed away in August, 2021.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1071156/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Variance components and standard deviation of genomic effects predicted by the genetic models of different genomic prediction method for fresh (FRY) and dry root yield (DRY), dry matter content (DMC) in cassava. Bayes B A: Bayes B method with additive genetic model; Bayes B A+D: Bayes B method with additive-dominant genetic model; G-BLUP A: G-BLUP method with additive genetic model; G-BLUP A+D Cla: G-BLUP method with additive-dominant classical genetic model; G-BLUP A+D Gen: G-BLUP method with additive-dominant genotypic genetic model; RKHS: reproducing kernel Hilbert spaces..

SUPPLEMENTARY FIGURE 2

Genetic gains by reducing the cassava breeding cycle takes into account the genomic (GB) and phenotypic breeding (PB) using the genomic prediction method and genetic models with higher predictive ability. G-BLUP A+D classical method for fresh root yield (FRY) and dry root yield (DRY) and G-BLUP A for dry matter content (DMC).

References

- Almeida Filho, J. E., Guimarães, J. F. R., Silva, F. F., Resende, M. D. V., Muñoz, P., Kirst, M., et al. (2016). The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity* 117, 33–41. doi: 10.1038/hdy.2016.23
- Almeida Filho, J. E., Guimarães, J. F. R., Silva, F. F., Resende, M. D. V., Muñoz, P., Kirst, M., et al. (2019). Genomic prediction of additive and non-additive effects using genetic markers and pedigrees. *G3 (Bethesda)* 8; 9 (8), 2739–2748. doi: 10.1534/g3.119.201004
- Azevedo, C. F., Resende, M. D. V., Silva, F. F., Viana, J. M. S., Valente, M. S. F., Resende, C.OMMA Jr., M. F. R., et al. (2015). Ridge, lasso and Bayesian additive-dominance genomic models. *BMC Genet.* 16 (105), 1–13. doi: 10.1186/s12863-015-0264-2
- Bakare, M. A., Kayondo, S. I., Aghogho, C. I., Wolfe, M. D., Parkes, E. Y., Kulakow, P., et al. (2022). Exploring genotype by environment interaction on cassava yield and yield related traits using classical statistical methods. *PLoS One* 17 (7), e0268189. doi: 10.1371/journal.pone.0268189
- Barandica, O. J., Pérez, J. C., Lenis, J. I., Calle, F., Morante, N., Pino, L., et al. (2016). Cassava breeding II: phenotypic correlations through the different stages of selection. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01649
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Software* 67 (1), 1–48. doi: 10.18637/jss.v067.i01
- Bouvet, J. M., Makouanzi, G., Cros, D., and Vigneron, P. H. (2016). Modelling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. *Heredity* 116, 146–157. doi: 10.1038/hdy.2015.78
- Browning, B. L., and Browning, S. R. (2016). Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* 98 (1), 116–126. doi: 10.1016/j.ajhg.2015.11.020
- Calle, F., Perez, J. C., Gaitán, W., Morante, N., Ceballos, H., Llano, G., et al. (2005). Diallel inheritance of relevant traits in cassava (*Manihot esculenta* crantz) adapted to acid-soil savannas. *Euphytica* 144, 177–186. doi: 10.1007/s10681-005-5810-y
- Ceballos, H., Iglesias, C. A., Pérez, J. C., and Dixon, A. G. (2004). Cassava breeding: opportunities and challenges. *Plant Mol. Biol.* 56 (4), 503–516. doi: 10.1007/s11103-004-5010-5
- Ceballos, H., Jaramillo, J. J., Salazar, S., Pineda, L. M., Calle, F., Setter, T., et al. (2017). Induction of flowering in cassava through grafting. *J. Plant Breed Crop Sci.* 9, 19–29. doi: 10.5897/JPBBS2016.0617
- Ceballos, H., Kulakow, P., and Hershey, C. (2012). Cassava breeding: current status, bottlenecks and the potential of biotechnology tools. *Trop. Plant Biol.* 5, 73–87. doi: 10.1007/s12042-012-9094-9
- Ceballos, H., López-Lavalle, L. A. B., Calle, F., Morante, N., Ovalle, T. M., and Hershey, C. (2016b). Genetic distance and specific combining ability in cassava. *Euphytica* 210 (1), 79–92. doi: 10.1007/s10681-016-1701-7
- Ceballos, H., Pérez, J. C., Barandica, O. J., Lenis, J. I., Morante, N., Calle, F., et al. (2016a). Cassava breeding I: the value of breeding value. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01227
- Cohen, J. A. (1960). Coefficient of agreement for nominal scales. *Educ. Psychol. Meas* 20 (1), 37–46. doi: 10.1177/001316446002000104
- Covarrubias-Pazarán, G. (2016). Genome-assisted prediction of quantitative traits using the r package sommer. *PLoS One* 11 (6), 1–15. doi: 10.1371/journal.pone.0156744
- Crossa, J., De Los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186 (2), 713–724. doi: 10.1534/genetics.110.118521
- Crossa, J., Pérez, P., Hickey, J., Burgueño, J., Ornella, L., Cerón-Rojas, J., et al. (2013). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112, 48–60. doi: 10.1038/hdy.2013.16
- Denis, M., and Bouvet, J. M. (2013). Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. *Tree Genet. Genomes* 9 (1), 37–51. doi: 10.1007/s11295-012-0528-1
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bull.* 19, 11–15.
- Esfandyari, H., Bijma, P., Henryon, M., Christensen, O. F., and Sørensen, A. C. (2016). Genomic prediction of crossbred performance based on purebred landrace and Yorkshire data using a dominance model. *Genet. Sel. Evol.* 48 (40), 1–9. doi: 10.1186/s12711-016-0220-2
- Ferguson, M., Rabbi, I., Kim, D. J., Gedil, M., Lopez-Lavalle, L. A. B., and Okogbenin, E. (2012). Molecular markers and their application to cassava breeding: past, present and future. *Trop. Plant Biol.* 5 (1), 95–109. doi: 10.1007/s12042-011-9087-0
- Food and Agriculture Organization of the United Nations (2022). *FAOSTAT statistical database* (Rome: FAO).
- Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41, 55. doi: 10.1186/1297-9686-41-55
- Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173 (3), 1761–1776. doi: 10.1534/genetics.105.049510
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinf.* 12 (186), 1–12. doi: 10.1186/1471-2105-12-186
- Hamblin, M. T., and Rabbi, I. Y. (2014). The effects of restriction-enzyme choice on properties of genotyping-by-sequencing libraries: a study in cassava. *Crop Sci.* 54 (6), 2603–2608. doi: 10.2135/cropsci2014.02.0160
- Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91 (1), 47–60. doi: 10.1017/S0016672308009981
- Heidaritabar, M., Wolc, A., Arango, J., Zeng, J., Settar, P., Fulton, J., et al. (2016). Impact of fitting dominance and additive effects on accuracy of genomic prediction of breeding values in layers. *J. Anim. Breed. Genet.* 133, 334–346. doi: 10.1111/jbg.12225
- Jaramillo, G., Morante, N., Pérez, J. C., Calle, F., Ceballos, H., Arias, B., et al. (2005). Diallel analysis in cassava adapted to the midaltitude valleys environment. *Crop Sci.* 45 (3), 1058–1063. doi: 10.2135/cropsci2004.0314
- Kawano, K., Fukuda, W. M. G., and Cenpukdee, U. (1987). Genetic and environmental effects on dry matter content of cassava root. *Crop Sci.* 27 (1), 69–74. doi: 10.2135/cropsci1987.0011183X002700010018x
- Kawano, K., Narintaraporn, K., Narintaraporn, P., Sarakarn, S., Limsila, A., Limsila, J., et al. (1998). Yield improvement in a multistage breeding program for cassava. *Crop Sci.* 38 (2), 325–332. doi: 10.2135/cropsci1998.0011183X003800020007x
- Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F., and Fritz, S. (2011). Improved lasso for genomic selection. *Genet. Res.* 93 (1), 77–87. doi: 10.1017/S0016672310000534
- Lyra, D. H., Galli, G., Alves, F. C., Granato, I. S. C., Vidoti, M. S., Sousa, M. B., et al. (2019). Modeling copy number variation in the genomic prediction of maize hybrids. *Theor. Appl. Genet.* 132 (1), 273–288. doi: 10.1007/s00122-018-3215-2
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 (4), 1819–1829. doi: 10.1093/genetics/157.4.1819
- Muñoz, P. R., Resende, M. F. R., Gezan, S. A., Resende, M. D. V., De Los Campos, G., Kirst, M., et al. (2014). Unraveling additive from nonadditive effects using genomic kinship matrices. *Genetics* 198 (4), 1759–1768. doi: 10.1534/genetics.114.171322
- Oliveira, E. J., Aud, F. F., Morales, C. F. G., Oliveira, S. A. S., and Santos, V. S. (2016). Non-hierarchical clustering of *Manihot esculenta* crantz germplasm based on quantitative traits. *Cienc. Agron.* 47 (3), 548–555. doi: 10.5935/1806-6690.20160066
- Oliveira, E. J., Filho, O. S., and Santos, V. S. (2015b). Classification of cassava genotypes based on qualitative and quantitative data. *Genet. Mol. Res.* 14 (1), 906–924. doi: 10.4238/2015
- Oliveira, E. J., Resende, M. D. V., Santos, V. S., Ferreira, C. F., Oliveira, G. A. F., Silva, M. S., et al. (2012). Genome-wide selection in cassava. *Euphytica* 187 (2), 263–276. doi: 10.1007/s10681-012-0722-0
- Oliveira, E. J., Santana, F. A., Oliveira, L. A., and Santos, V. S. (2015a). Genotypic variation of traits related to quality of cassava roots using affinity propagation algorithm. *Sci. Agric.* 72 (1), 53–61. doi: 10.1590/0103-9016-2014-0043
- Park, T., and Casella, G. (2008). The Bayesian lasso. *J. Am. Stat. Assoc.* 103 (482), 681–686. doi: 10.1198/016214508000000337
- Parkes, E., Aina, O., Kingsley, A., Iluebbey, P., Bakare, M., Agbona, A., et al. (2020). Combining ability and genetic components of yield characteristics, dry matter content, and total carotenoids in provitamin a cassava F1 cross-progeny. *Agronomy* 10, 1850. doi: 10.3390/agronomy10121850
- Parkes, E. Y., Fregene, M., Dixon, A., Boakye-Peprah, B., and Labuschagne, M. T. (2013). Combining ability of cassava genotypes for cassava mosaic disease and cassava bacterial blight, yield and its related components in two ecological zones in Ghana. *Euphytica* 194 (1), 13–24. doi: 10.1007/s10681-013-0936-9
- Perez, P., and De Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198 (2), 483–495. doi: 10.1534/genetics.114.164442

- Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161 (1), 209–228. doi: 10.1007/s10681-007-9449-8
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News* 6 (1), 7–11.
- Raftery, A. E., and Lewis, S. M. (1992). Practical Markov chain Monte Carlo: Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Statist. Sci.* 7 (4), 493–497. doi: 10.1214/ss/1177011143
- R Core Team (2018). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: <https://www.R-project.org/>.
- Russell, L. (2018) *Emmeans: estimated marginal means, aka least-squares means*. Available at: <https://CRAN.R-project.org/package=emmeans>.
- Shull, G. H. (1908). The composition of a field of maize. *J. Hered.* 296–301. doi: 10.1093/jhered/os-4.1.296
- Souza, L. S., Farias, A. R., Mattos, P. L. P., and Fukuda, W. M. G. (2006). Aspectos socioeconômicos e agrônômicos da mandioca. Embrapa Mandioca e Fruticultura Tropical: Cruz das Almas (BA).
- Su, G., Christensen, O. F., Ostensen, T., Henryon, M., and Lund, M. S. (2012). Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS One* 7 (9), e45293. doi: 10.1371/journal.pone.0045293
- Tan, B., Grattapaglia, D., Wu, H. X., and Ingvarsson, P. K. (2018). Genomic kinships reveal significant dominance effects for growth in hybrid *Eucalyptus*. *Plant Sci.* 267, 84–93. doi: 10.1016/j.plantsci.2017.11.011
- Tumuhimbise, R., Melis, R., and Shanahan, P. (2014). Diallel analysis of early storage root yield and disease resistance traits in cassava (*Manihot esculenta* crantz). *Field Crops Res.* 167, 86–93. doi: 10.1016/j.fcr.2014.07.006
- Vanraden, P. M. (2007). Genomic measures of kinship and inbreeding. *Interbull Anal. Meet. Proc.* 37, 33–36. doi: 10.3168/jds.2007-0980
- Vitezica, Z. G., Varona, L., and Legarra, A. (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* 195 (4), 1223–1230. doi: 10.1534/genetics.113.155176
- Wolfe, M. D., Chan, A. W., Kulakow, P., Rabbi, I., and Jannink, J. L. (2021). Genomic mating in outbred species: predicting cross usefulness with additive and total genetic covariance matrices. *Genetics* 219 (3), iyab122. doi: 10.1093/genetics/iyab122
- Wolfe, M. D., Kulakow, P., Rabbi, I. Y., and Jannink, J. L. (2016a). Marker-based estimates reveal significant non-additive effects in clonally propagated cassava (*Manihot esculenta*): implications for the prediction of total genetic value and the selection of varieties. *G3* 6 (11), 3497–3506. doi: 10.1534/g3.116.033332
- Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., et al. (2016b). Genome-wide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. *Plant Genome* 9 (2), 1–13. doi: 10.3835/plantgenome2015.11.0118
- Zacarias, A. M., and Labuschagne, M. T. (2010). Diallel analysis of cassava brown streak disease, yield and yield related characteristics in Mozambique. *Euphytica* 176 (3), 309–320. doi: 10.1007/s10681-010-0203-2
- Zhang, Z., Erbe, M., He, J., Ober, U., Gao, N., Zhang, H., et al. (2015). Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. *G3* 5 (4), 615–627. doi: 10.1534/g3.114.016261
- Zuk, O., Hechter, E., Sunyaev, S. R., and Lander, E. S. (2012). The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U.S.A.* 109 (4), 1193–1198. doi: 10.1073/pnas.1119675109