

Single and Multi-trait Genomic Prediction for agronomic traits in

Euterpe edulis

Guilherme Bravim Canal^a, Cynthia Aparecida Valiati Barreto^b, Francine Alves Nogueira de Almeida^a, Iasmine Ramos Zaidan^a, Diego Pereira do Couto^a, Camila Ferreira Azevedo^b, Moysés Nascimento^b, Marcia Flores da Silva Ferreira^{a,*}, Adésio Ferreira^a.

^aDepartment of Agronomy, Federal University of Espírito Santo, Alegre, Espírito Santo, Brazil

^bDepartment of Statistics, Federal University of Viçosa, Viçosa, Minas Gerais, Brazil

Key message: In the first genomic selection study for *Euterpe edulis*, substantial gains for multiple traits of fruit production was reported. This is a key factor for the sustainable use of the species in the Atlantic Forest.

*Corresponding author

E-mail address: marcia.ferreira@ufes.br

Federal University of Espírito Santo, Department of Agronomy, Alegre, Espírito Santo, Brazil,

Zip Code: 29500-000.

Abstract

Popularly known as juçaizeiro, *Euterpe edulis* has been gaining prominence in the fruit growing sector and has demanded the development of superior genetic materials. As it is a native species and still little studied, the application of more sophisticated techniques can result in higher gains with less time. Until now, there are no studies that apply genomic prediction for this crop, especially in multi-trait analysis. In this sense, this study aimed to apply new methods and breeding techniques for the juçaizeiro, to optimize this breeding program through the application of genomic selection. This data consisted of 275 juçaizeiro genotypes from a population of Rio Novo do Sul-ES, Brazil. The genomic prediction was performed using the multi-trait (G-BLUP MT) and single-trait (G-BLUP ST) models and the selection of superior matrices was based on the selection index of Mulamba and Mock. Similar results for predictive ability were observed for both models. However, the G-BLUP ST model provided greater selection gains when compared to the G-BLUP MT. For this reason, the genomic estimated breeding values (GEBVs) from the G-BLUP ST were used to select the six superior genotypes (UFES.A.RN.390, UFES.A.RN.386, UFES.A.RN.080, UFES.A.RN.383, UFES.S.RN.098, and UFES.S.RN.093), to provide superior genetic materials for the development of seedlings and implantation of productive orchards, which will meet the demands of the productive, industrial and consumer market.

Key words: G-BLUP, juçara palm heart, breeding, wide genomic selection

1. Introduction

Native to the Atlantic Forest, *Euterpe edulis*, popularly known as juçaizeiro, has shown great economic potential (de Oliveira Maciel et al., 2019; Schulthess et al., 2016; Silva et al., 2018). Its fruits are used for the industrial production of processed pulp, which is similar to açaí (Silva et al., 2018). Given its productive potential and wild *status*, there is a demand for superior genetic materials to use the species as a new crop. However, due to a series of intrinsic characteristics, such as, the wild state of the species, slow development, difficulty in the practice of controlled pollination and propagation exclusively via seminal, classical breeding practices may not be sufficient to lead to satisfactory genetics gains in the short time currently required.

Within breeding programs, the genome-wide selection (GWS) proposed by Meuwissen et al. (2001), seeks through its techniques, such as G-BLUP (Genomic Best Linear Unbiased Prediction), the use of knowledge of genomic relationship to increase selective accuracy (Alkimim et al., 2020; de Resende et al., 2015). The application of the GWS in a natural population open-pollinated, as the base population of improvement of the juçaizeiro species, in which the degree of relationship between the individuals is unknown, is scarce in the literature. However, several studies were conducted on forest species from open-pollinated populations with knowledge of families (Bush and Thumma, 2013; El-Dien et al., 2018; El-Kassaby et al., 2011; Gamal El-Dien et al., 2016; Klápště et al., 2018) and support the assumption of increased accuracy with GWS application in a natural population without knowledge of relationship structure. This is due to the fact that, in these works, the use of the genomic relationship matrix (G) corrects the unrealistic priori that, in a given family, all individuals share the same genetic similarity with each other. In addition to the above, the accuracy of genomic prediction is generally higher, especially for breeding populations with superficial genealogy and disconnected families (Thavamanikumar et al., 2020).

Genomic prediction based on single traits has become a widely applied procedure

among breeders after reducing genotyping costs (Jarquín et al., 2019; Massman et al., 2013; Sousa et al., 2019), mainly using the G-BLUP method. However, this model may not satisfactorily reflect the interactive complexities between the analyzed traits, because it does not capitalize on the flows of information between traits through available information on genetic covariances (Gaire et al., 2022). These covariances are generated through shared genetic influence (pleiotropy) and/or by the non-random association between the alleles (linkage disequilibrium), which, consequently, are responsible for generating complex relationships between quantitative characters (Lynch and Walsh, 1998). Therefore, the multi-trait approach (G-BLUP MT) applied to GWS has been highlighted for being able to combine information and capture the effect of association between traits in order to predict genetic values more accurately (Bhatta et al., 2020; Gaire et al., 2022; Lado et al., 2018; Sapkota et al., 2020).

Currently, for the juçaizeiro, the techniques and approaches applied to its scarce existing breeding programs are simple. In general, the selection of the best genotypes are carried out in natural populations, based on data from a single year and with a mass phenotypic selection technique, due to the absence of experimental fields with adult representatives of the species, without pre-established delineation and lack of knowledge of relationship between the individuals. In this sense, the selective bias can compromise the success of the species as a crop, by indicating low productivity genetic materials. Therefore, the present work aims to apply new methods and breeding techniques for juçaizeiro, to optimize the breeding programs through the application of genomic selection. This study was carried out in a natural open-pollinated population containing 275 genotypes, with no knowledge of relationship between individuals in order to select the genotypes with the greatest potential to serve the productive and industrial sector; and to compare the efficiency between multi-trait (G-BLUP MT) and single-trait (G-BLUP ST) models.

2. Materials and Methods

2.1. Field experimental conduction

Experimental evaluations were carried out from 2018 to 2021 in the municipality of Rio Novo do Sul in the state of Espírito Santo, Brazil (Fig. 1). The 275 genotypes evaluated were selected for good phytosanitary and physiological conditions and full reproductive age. The genotypes are located in a commercial plantation belonging to two companies that process juçaizeiro fruit pulp, Açaí juçara® and Bonalotti®. The plantation was formed by a mixture of individuals that emerged by spontaneous development, through the action of dispersers, and the enrichment by the owners by the sowing of seeds in the area. Commercial planting does not have pre-defined spacing and basically no productive management treatment, only mowing at harvest time. Consequently, the field evaluations followed a blank design, with initial unknown relationship by descent between the genotypes.

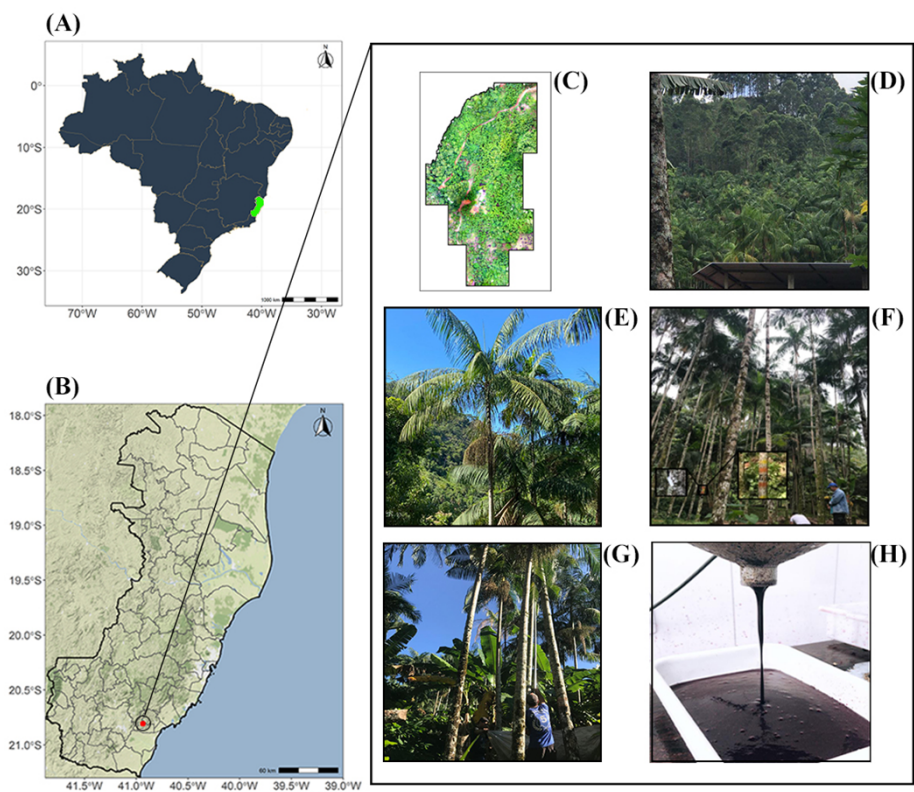


Fig. 1. Geographical location of the experimental field where the *Euterpe edulis* matrices are located. Map generated with R free environment *software*. (A) Delimitation of Brazil, in green, the geographic location of the state of Espírito Santo; (B) Graphic representation of the state of Espírito Santo, in red, the location of the municipality of Rio Novo do Sul; (C) Orthomosaic of the experimental area; (D) Image referring to a fragment of the experimental planting; (E) Juçaizeiro (*Euterpe edulis*) from experimental planting in productive phase; (F) Registration of genotypes identification in field; (G) Juçaizeiro (*Euterpe edulis*) fruits harvest; (H) Processed pulp of juçaizeiro (*Euterpe edulis*) fruits.

2.2. Phenotyping of plants

In the years 2018 to 2021, the number of bunches per plant (NB) was evaluated by visually counting the number of bunches with fruits of each matrix plant. In the years 2018, 2019 and 2021, the mass of fruits per bunch (MFB) (kg), rachis length (RL) (cm), equatorial diameter of fruits (EDF) (mm) and pulp yield (PY) (%) were evaluated. For the evaluation of MFB, in all evaluations, the same trained professional performed the determination of the harvest point of the bunches. The harvest of each matrix was carried out when the fruit maturation point reached the stage used for fruit processing in the industry. Each year, one bunch per plant was harvested, the fruits were separated from the rachillas, and they were weighed on a scale with an accuracy of 0.1 g. For RL, after separating the fruits, the length of the rachis of the inflorescence was measured with a tape measure.

A sample of fruits was taken from each matrix, packed in properly identified plastic bags and transported to the Plant Biometry laboratory at the Federal University of Espírito Santo, where morphometric evaluations of fruits and seeds were carried out in a completely randomized design. For EDF, data were measured in millimeters (mm) obtained using a 6" digital caliper (Zaasprecision®), performed on five fruits individually, as recommended by Marçal et al. (2015) who observed high repeatability values for these characteristics. The authors concluded that five measurements are necessary to carry out the measurement to reach

133 coefficients of determination of 95%, measurements above this amount increase costs and
134 evaluative time, bringing little additional information to the works.

135 The PY was estimated by the following relationship:

$$136 \quad PY = \frac{FFM - SFM}{MFF} * 100$$

137 where FFM is fruit fresh mass and SFM is seed fresh mass, measured by weighing four
138 replicates of 25 fruits and seeds, using an analytical balance (0.0001g).

139 2.3. Genomic DNA extraction

140 Genomic DNA was obtained from leaf samples of the genotypes under study. Extraction
141 was performed using the cetyltrimethylammonium bromide or CTAB method by Doyle (1990)
142 with modifications (Carvalho et al., 2020).

143 DNA concentrations and integrity were estimated using a Nanodrop™ 2000
144 spectrophotometer (Thermo Scientific). DNA quality was verified on 0.8% agarose gel. DNA
145 genotypes prepared for genotyping using the DArTseq™ methodology were sent to the
146 Service of Genetic Analysis for Agriculture (SAGA) in Mexico for high-throughput genotyping
147 using the DArTseq™ technology

148 The genome representation of the 275 genotypes was obtained from the reduction of
149 DNA complexity using two restriction enzymes, HpaII and MseI, and the ends of the cleaved
150 fragments were linked to a code adapter and a common adapter to identify each sample. The
151 fragments were amplified by PCR reaction; subsequently, equimolar amounts of amplification
152 products from each sample of the 96-well microtiter plate were pooled, purified and quantified,
153 then sequenced on the Illumina Novaseq 6000 System platform. The sequences were analyzed
154 using Dartsoft14, an automated genomics data analysis program and DArTdb, a laboratory

management system, developed and patented by DART Pvt. Ltd. (Australia), generating SNP marker data.

2.4. Quality control of molecular markers

The dataset of codominant markers of the SNP type was submitted to quality control analysis in the R (R Core Team, 2021). The quality parameters used were *Call Rate* (CR) of 90% and *Minor Allele Frequency* (MAF) of 1%. After quality control, the marker dataset reduced by 81.75% from 44,457 markers to 8,112.

2.5. Phenotypic data analysis

The phenotypic values were corrected for the effect of years in the R software according to the correction proposed by (Fanelli Carvalho et al., 2020). The linear model used was:

$$y_{ik} = \mu + a_k + g_i + e_{ik}$$

where y_{ik} is the phenotypic value for the genotype i and the year k ; μ is the population mean; a_k is the fixed effect of the k^{th} year; g_i is the fixed effect of the i^{th} genotype; and e_{ik} is the random effect of residual, with $e \sim N(0, I\sigma_e^2)$. The empirical best linear unbiased estimates (eBLUES) was calculated for each traits individually and their values were obtained by $\hat{y}_i = \mu + \hat{a}_k + g_i$.

2.6. Genomic prediction (GP)

Genomic prediction was performed using the multi-trait (G-BLUP MT) and single-trait

(G-BLUP ST) models. For this, was used the SOMMER package version 3.4 [26], implemented in the R software (R Core Team, 2021). The G-BLUP MT model for the prediction of the genomic estimated breeding values (GEBV) of the individuals used was:

$$y = Xb + Zu + e$$

Where, y is the vector of previously estimated eBLUES and structured as $y = [y_1 \ y_2 \ \dots \ y_n]'$; $y_1, y_2, \dots, y_n$ is the vector of observations for each characteristic; b is the vector of means of each characteristic structured as $b = [b_1 \ b_2 \ \dots \ b_n]'$ and with X incidence matrix; u is vector of individual additive genomic genetic values of each trait structured as $u = [u_1 \ u_2 \ \dots \ u_n]'$ with Z incidence matrix, with variance structure given by $u \sim N(0, G \otimes \Sigma_u)$, where G is the genomic relationship matrix between individuals for additive effects, Σ_u is the additive genetic covariance matrix and $\otimes$ denotes the Kronecker product; e is the random error vector with $e \sim N(0, I \otimes \Sigma_e)$ where Σ_e is the residual covariance matrix.

Covariance matrices can be written as:

$$\Sigma_u = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_{12}}^2 & \dots & \sigma_{u_{1n}}^2 \\ \sigma_{u_{21}}^2 & \sigma_{u_2}^2 & \dots & \sigma_{u_{2n}}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{u_{n1}}^2 & \sigma_{u_{n2}}^2 & \dots & \sigma_{u_n}^2 \end{bmatrix} \text{ and } \Sigma_e = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_{12}}^2 & \dots & \sigma_{e_{1n}}^2 \\ \sigma_{e_{21}}^2 & \sigma_{e_2}^2 & \dots & \sigma_{e_{2n}}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{e_{n1}}^2 & \sigma_{e_{n2}}^2 & \dots & \sigma_{e_n}^2 \end{bmatrix}$$

where $\sigma_{u_i}^2$ and $\sigma_{e_i}^2$ are, respectively, the additive and residual genetic variance associated with the i^{th} trait with $i = 1, \dots, n = 5$; $\sigma_{u_{ij}}^2$ and $\sigma_{e_{ij}}^2$ are, respectively, the additive and residual genetic covariance associated with the i^{th} and j^{th} traits with $j = 1, \dots, n = 5$ and $i \neq j$. The variance and covariance components were obtained via the restricted maximum likelihood method (REML). The additive genomic relationship matrix (G) was obtained as described by Van Raden [27] by the centralization of the matrix of markers:

$$G = \frac{WW^T}{\sum_{i=1}^n 2p_i(1 - p_i)}$$

where, W is the centered marker matrix, which specifies the marker genotypes for each individual as 0, 1 or 2; p_i is the frequency of the second allele at the *locus*.

With the genetic values of the individuals for the NB and MFB traits, the fruit production per plant (FPP) for each genotype was estimated.

The heritability in the strict sense of the i^{th} trait was estimated following the equation below:

$$h_u^2 = \frac{\sigma_{u_i}^2}{\sigma_{u_i}^2 + \sigma_{e_i}^2}$$

where $\sigma_{u_i}^2$ and $\sigma_{e_i}^2$ are, respectively, the additive genetic variance and the residual variance of the i^{th} trait. The genetic correlation between the i^{th} and j^{th} traits was obtained through the following equation:

$$r_{u_{ij}} = \frac{\sigma_{u_{ij}}^2}{\sqrt{\sigma_{u_i}^2 \times \sigma_{u_j}^2}}$$

where $\sigma_{u_i}^2$ and $\sigma_{u_j}^2$ are, respectively, the additive genetic variance of the i^{th} and j^{th} trait, and $\sigma_{u_{ij}}^2$ is the additive genetic covariance between the i^{th} and j^{th} traits.

210

2.7. Predictive ability and cross-validation

The predictive ability ($r_{y,\hat{u}}$) and the standard error of $r_{y,\hat{u}}$, were estimated through the cross-validation procedure, subdividing randomly the population into 5 folds. Thus, 220 genotypes were used for training set and 55 genotypes were used for validation set. For each

fold, the $r_{y,\hat{u}}$ was obtained by the correlation between the predicted GEBV's ($\hat{u}$) and the corrected phenotypic values (y).

2.8. Genetic selection based on selection index

With the GEBV'S predicted by the G-BLUP ST method, the selection of the best genotypes was carried out based on multiple variables. For this, the method of Mulamba and Mock (1978) was used, which is based on the sum of the individual ranks of each trait, creating a global rank. For the selection, the parameters of NB, MFB, PFP, RL and PY in the positive direction and EDF in the negative direction were considered. Two processes were performed before rank summing. The first was to transform the EDF values into classes, in order to create four classes. Class I for small fruits (values below the first quartile of the distribution of the genetic values); class II for small/medium size fruits (genotypes with values between the first and second quartile of the distribution of the genetic values); class III for medium fruits (genotypes with values between the second and third quartile of the genetic values), and class IV for large fruits (genotypes with values above the third quartile of the genetic values). The second change was the normalization of the number of ranks for each trait, to avoid traits with fewer ranks having a greater influence on the selection process. The standardization followed the expression below:

$$r_{ij} = \frac{nr_j - p_{ij}}{nr_j - 1}$$

where r_{ij} is the standardized rank value for genotype i and characteristic j ; nr_j is the number of ranks of trait j and p_{ij} is the rank of genotype i for the trait j .

In order to compare the selective efficiency between the G-BLUP ST and G-BLUP MT models, was estimated the expected selection gain for the population of the next selection cycle

and for the commercial seed donor population. For both methods and all characteristics, with the exception of PFP, estimates were obtained by:

$$SG_{ij} = h_{ij}^2 * SD_i$$

where, SG is the expected selection gain for characteristic i by model j ; h_{ij}^2 is the heritability of characteristic i estimated by model j and SD_i is the selection differential for characteristic i in model j estimated based on corrected phenotypic values.

The SG for PFP was estimated based on the product of the SG estimates of NB and MFB.

2.9. Comparison between methodologies

Cohen's Kappa coefficient (Cohen, 1960) was used to analyze the agreement between the best selected individuals between the G-BLUP ST e G-BLUP MT models, for the commercial seed donor population (formed by the six best ranked individuals), and the second selective cycle population (formed by the 50 best ranked individuals). Cohen's Kappa coefficient is given by:

$$kappa = \frac{NAO - NAEC}{NOA - NAEC}$$

where NAO are the number of observed agreements, $NAEC$ is the number of expected agreements by chance, and NOA is the number of analyzed observations (Resende et al., 2014).

3. Results

3.1. Data description

A summary of the descriptive statistics including mean, standard deviation, and maximum and minimum values for the five agronomic traits evaluated in this work is shown in Table 1. It is noted that in general, the population presented a small variation in the average phenotypic response between the years evaluated.

Table 1

Amplitudes, mean, standard deviation (SD) and coefficient of variation (CV) for rachis length (RL), fruit mass per bunch (MFB), number of bunches (NB), equatorial fruit diameter (EDF) and pulp yield (PY) for the years 2018, 2019, 2020 and 2021.

Var.	Year	Minimum	Maximum	Average	SD	CV (%)
RL (cm)	2018	32	99	60.6	14.35	23.68
	2019	33	104	62.72	14.15	22.56
	2021	34.3	90.3	59.31	10.39	17.52
MFB (kg)	2018	0.4	8.3	3.05	1.57	51.47
	2019	0.5	8.7	3.45	1.49	43.19
	2021	0.3	8.7	3.68	1.87	50.81
NB	2018	1	7	4.24	1.35	31.84
	2019	0	8	3.28	1.36	41.46
	2020	0	7	3.69	1.03	27.91
	2021	0	7	2.18	1.36	62.38
EDF (mm)	2018	11.15	18.9	14.03	1.15	8.20
	2019	10.36	21.3	13.24	1.09	8.23
	2021	10	19.94	13.31	1.25	9.39
PY (%)	2018	4.57	50.81	30.62	6.5	21.23
	2019	16.75	51.52	35.96	5.37	14.93
	2021	12.55	46.63	29.76	6.06	20.36

3.2. Genetic parameters, Genetic correlation and phenotypic correlation

The heritability and the environmental and additive genetics variance components were estimated for G-BLUP ST and G-BLUP MT models (Table 2). For both approaches, the higher residual variance compared to the variance due to genetic effects is emphasized, with the exception of EDF in the G-BLUP ST, where these estimates were equal (Table 2). For the traits evaluated, with RL exception, the G-BLUP ST model presented estimates of the genetic variance components (σ_u^2) higher than the G-BLUP MT (Table 2).

Table 2

Additive genetic variance (σ_u^2), residual variance (σ_e^2) and restricted heritability (h_u^2) considering single trait (ST) and multiple trait (MT) models in genomic data.

Model	Par.	RL	MFB	NB	EDF	PY
Single-trait	σ_u^2	59.55	1.04	0.32	0.81	19.71
	σ_e^2	73.53	1.15	0.72	0.41	4.96
	h_u^2	0.45	0.48	0.31	0.67	0.80
Multi-trait	σ_u^2	61.75	0.90	0.31	0.78	19.49
	σ_e^2	72.17	1.21	0.73	0.41	5.06
	h_u^2	0.46	0.43	0.29	0.65	0.79

Parameters (Par.); Rachis Length (RL); fruit mass per bunch (MFB); number of bunches (NB); equatorial diameter of fruit (EDF) and pulp yield (PY).

Comparing both models in relation to σ_u^2 , the G-BLUP ST provided estimates 15.56% (MFB), 3.23% (NB), 3.85% (EDF) e 1.13% (PY) higher than the estimated by G-BLUP MT. Due to these slight differences, and the similarity of the σ_e^2 observed between the G-BLUP MT and G-BLUP ST, in general, the h_u^2 were similar between the tested models, ranging from 0.01 to 0.05 magnitude. The most significant difference was observed for MFB, which obtained a 0.05 higher h_u^2 for the G-BLUP ST (Table 2).

The genetic and phenotypic correlations between the traits are shown in Fig. 2, which were calculated using the (co)variance estimates obtained by the G-BLUP MT model, so that the genetic correlation estimates are presented on the upper diagonal and, on the lower diagonal, estimates of phenotypic correlations are presented.

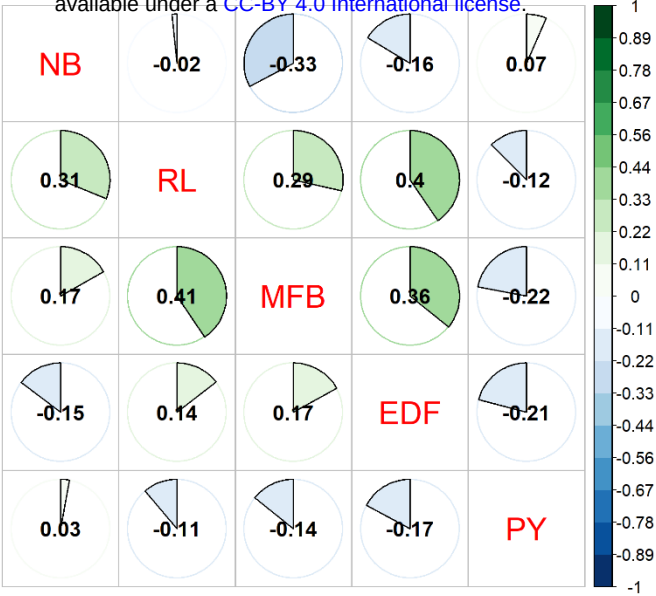


Fig. 2. Genetic (upper diagonal) and phenotypic (lower diagonal) correlation between the traits RL (Rachis Length), EDF (Equatorial Fruit Diameter), MFB (Fruit Mass per Bunch), NB (Bunch Number) and PY (Pulp yield).

Comparing the phenotypic and genotypic correlations between the traits in Fig. 2, it is observed that only two correlation estimates had changes in their directions (RL and NB; MFB and NB), with the change in correlation between MFB and NB being more pronounced. The results obtained for the phenotypic and genotypic correlations can be classified from low to moderate (Evans et al., 1996). For phenotypic correlation, the highest values observed were 0.41 (RL and MFB), in the positive direction, and -0.17 (EDF and PY), in the negative direction. While, for the genotypic correlations between traits, the highest values were 0.40 (RL and EDF) in the positive direction, and -0.37 (NB and MFB) in the negative direction (Fig. 2).

3.3. Predictive ability

The predictive ability ($r_{y,\hat{u}}$) for the G-BLUP ST and G-BLUP MT models is shown in Table 3. For the G-BLUP ST, the $r_{y,\hat{u}}$ ranged from 0.21 for NB to 0.49 for PY, and for the G-BLUP MT model, the $r_{y,\hat{u}}$ ranged from 0.18 for NB to 0.48 for PY. In general, the values of $r_{y,\hat{u}}$

observed for both methodologies were close, as well as NB and PY maintained their positions with the lowest and highest $r_{y,\hat{u}}$, respectively. The results show that both models are similar to perform the prediction of genomic genetic values.

Table 3

Predictive ability ($r_{y,\hat{u}}$) of the single-trait (G-BLUP ST), multi-trait (G-BLUP MT) models for RL (Rachis Length), EDF (Equatorial Fruit Diameter), MFB (Mass of Fruit per Bunch), NB (Number of Bunch) and PY (Pulp Yield).

Parameters	G-BLUP ST	G-BLUP MT
RL	0.29 (± 0.03)	0.22 (± 0.01)
MFB	0.27 (± 0.03)	0.26 (± 0.04)
NB	0.21 (± 0.06)	0.18 (± 0.05)
EDF	0.44 (± 0.02)	0.42 (± 0.01)
PY	0.49 (± 0.06)	0.48 (± 0.07)

3.4. Matrix selection and expected genetic advancement

The Cohen's Kappa coefficient was calculated for the top 20% of individuals (50) with the highest rank positions based on the Mulamba and Mock index (Mulamba and Mock, 1978), aimed to form the population of the second selection cycle. The agreement between the G-BLUP ST and G-BLUP MT models was 0.68 and is classified as substantial (0.60-0.80) (Sergio R. Munoz and Bangdiwala, 1997) (Fig. 3). However, when we reduce the selected population to six individuals, to build the donor population of commercial seeds, the coefficient is reduced to an agreement of 0.50, considered moderate.

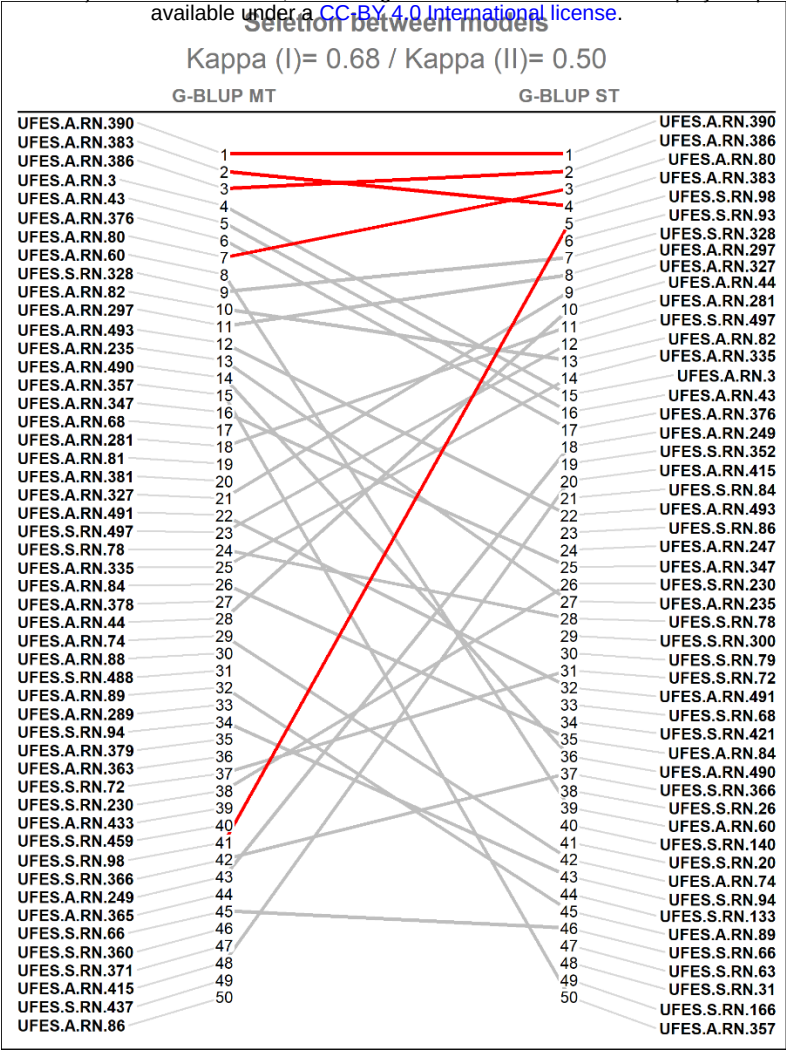


Fig. 3. The Cohen's Kappa coefficient showing the agreement between genotypes selected by the G-BLUP MT and G-BLUP ST models. Kappa (I) - 20% of the best genotypes (50 genotypes) and Kappa (II) - the six best genotypes selected.

Even though the agreements of the Cohen's Kappa coefficient were classified from moderate (0.41-60) to substantial (0.61-60) (Munoz and Bangdiwala, 1997), is observed in Fig. 3, a great divergence in the ranking of individuals between the methods used. To support the comparison of the efficiency between the methods, the selection gain (SG) estimate was obtained through the heritability and phenotypic means of the selected individuals, and the results are shown in Table 4. In general, higher SG resulting from the G-BLUP ST model is observed for both populations.

Table 4

Estimates of selection gains provided by the single-trait (G-BLUP ST) and multi-trait (G-BLUP MT) models for the selected genotypes (S.I.) for the population of the next selection cycle (50 genotypes), and the commercial seed donor population (six genotypes).

Trait	Population of the next selection cycle		Deed donor population	
	G-BLUP ST	G-BLUP MT	G-BLUP ST	G-BLUP MT
RL	31.25	31.64	30.58	32.58
MFB	2.06	1.62	2.40	1.76
NB	1.16	1.14	1.25	1.22
EDF	8.75	8.57	8.62	8.32
PY	27.60	27.10	30.15	29.86
PFP	2.39	1.85	3.00	2.15

RL: Rachis Length, EDF: Equatorial Fruit Diameter, MFB: Mass of Fruit per Bunch, NB: Number of Bunch, and PY: Yield of Pulp.

The population mean, the phenotypic and genotypic behavior of the selected individuals, based on the information from the G-BLUP ST model and the field phenotypic, is shown in Fig. 4. It is possible to observe that the phenotypic means of the six selected individuals, in the desired direction, have great performance in relation to the population average, being the improvement 31.13%, 51.05%, 98,40%, 11.56%, 15.57%, and 7.43% for NB, MFP, PFP, RL, PY and EDF, respectively.

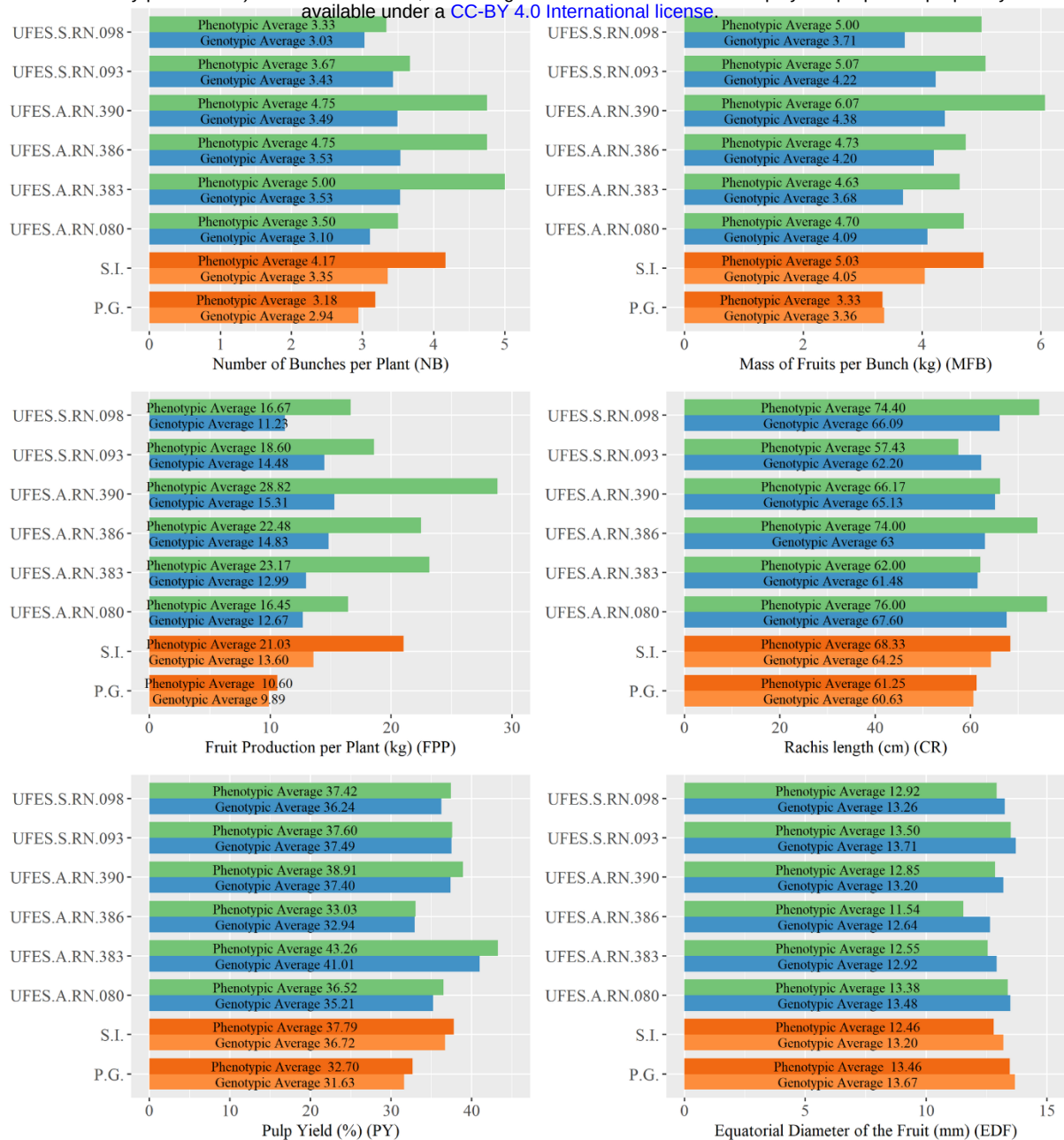


Fig. 4. Phenotypic means observed in the field and GEBV's + mean of selected genotypes and general population. S.I.: commercial seed donor population, P.G.: general population.

The selection of the six best genotypes provides a change in phenotypic responses equivalent to 0.99 bunches, 1.70 kg, 10.43 kg, 7.08 cm, 5.09%, and -1.00 mm for the traits NB, MFP, PFP, RL, PY and EDF, respectively (Fig. 4). With SG, is expected an improve on the averages in 0.41 bunches, -0.69 kg, 3.62 cm, 5.09%, and -0.47 mm, for NB, MFP, RL, PY, and

EDF, respectively. For PFP, the genomic values were estimated by the product of the genomic values of MFB and NB, and the difference between the means of the selected and the general population was 3.71 kg (Fig. 4).

4. Discussion

As there are few programs that are based on the improvement of juçaizeiro, also a few scientific works that are aimed at the application of selective techniques for individuals of this species. In this sense, the existing breeding programs are in the early stages of development, and thus, their techniques are basically based on classical methodologies with phenotypic information. In contrast to these processes, the present work innovates in being the first scientific study focused on genomic selection of *Euterpe edulis*, with the objective of increasing selective accuracy and selection gains within the breeding cycle.

4.1. Genetic parameters, genetic correlation and phenotypic correlation

The h_u^2 found in the present work for the traits under study indicate that the morphometric traits of fruits (EDF and PY) have a greater potential for heredity and genetic control, when compared to the other productive parameters related to clusters (CR, MFB and NB). This behavior was also observed in *Euterpe oleracea* (açazeiro), a species of the same genus as juçaizeiro (Farias Neto et al., 2008; Teixeira et al., 2012).

It is expected that traits with higher h_u^2 , consequently, have a higher predictive ability (Legarra et al., 2008). The same was found by Legarra et al. (2008), where the traits with the highest h_u^2 in the present study (EDF and PY) also had the higher predictive ability (Table 3). The higher values of h_u^2 for the traits of EDF and PY in relation to NB, MFB and RL can

demonstrate the dynamism of the behavior of the h_u as a function of the hereditary and environmental behavior. It is known that for quantitative traits, several environmental factors can affect phenotypic behavior. In this sense, it can be assumed that the genetic control of NB, MFB and RL is reduced due to the exposure of these traits for a longer period of time to environmental effects and more variable conditions that are not controlled, for example, direction of insertion of the bunch into the matrix plant, exposure to inclement weather, feeding the fauna, among others. While, the development of fruits in the bunch will occur, for example, in a fraction of the time of the infructescence development cycle, which will be located in the same region of the plant, with a similar influence on luminosity, temperature, and availability of nutrients. Therefore, genetic control can be expected to be greater.

Published scientific studies evaluating associations between *Euterpe edulis* traits are scarce in the literature (Marçal et al., 2015; Oliveira et al., 2015; Silva et al., 2018). However, these analyzes are fundamental for understanding the interactive behavior between traits, revealing linear cause and effect responses that are essential for the agricultural development of the crop and for breeding programs, as it allows the determination of different types of practical strategies.

The knowledge of correlations in the genetic scope is important in breeding programs because they are heritable and can be used in indirect selection (Cruz et al., 2012). However, the evaluation of phenotypic behavior is also essential, given the fact that traits genotypically and non-phenotypically associated end up not having practical applicability in the indirect selection process, since the phenotype is influenced both by genetic and environmental effects.

The (co)variances estimated by the G-BLUP MT, made it possible to estimate the Pearson correlation for genetic and phenotypic effects between the traits RL, EDF, MFB, NB and PY (Fig. 2), allowing a counterpoint between these estimates. In this sense, we noticed that

practically all the associations preserved their influence on the behavioral response, with the exception of the NB-RL and NB-MFB pairs, which had an inversion of the sense of their estimates. In this condition, indirect selection is compromised by the fact that selection by genotypic responses may not be expressed in the desired sense in the environment. With this, we can conclude that the analyzes of phenotypic and genotypic associations are complementary and make it possible to determine practical actions, using indirect processes to enable increased gains in breeding programs.

Evaluating biometric characteristics of juçaizeiro fruits, Marçal et al. (2015) reported a positive genetic association between EDF and fresh seed mass (0.75). Knowing that the juçaizeiro fruit is mostly composed of seed, and the pulp makes up a small portion of it, increasing EDF can reduce the percentage of pulp yield. This fact is confirmed by the association observed in Fig. 2, which shows inversely proportional associations between EDF and PY (-0.17 phenotypic; -0.21 genetic), that is, the increase in fruit size results in a reduction in the percentage of pulp per fruit, harmful to industrial processing.

The results observed by Oliveira et al. (2015) for the genetic correlation in *Euterpe edulis*, are close to those observed in the present work between the characteristics NB, PY and RL. In both situations, the magnitudes of the associations can be classified from very weak to weak (Evans et al., 1996). Oliveira et al. (2015) reported associations of 0.20, -0.10 and -0.02 for NB-PY, PY-RL and NB-RL, respectively, while the present work obtained estimates of 0.07, -0.12 and -0.02, respectively. Corroborating these results, Farias Neto et al. (2016) also found a weak correlation between NB-RL (-0.01) for *Euterpe oleracea*.

Regarding the correlation results observed in Fig. 2, we can determine that the use of indirect selection among the evaluated characteristics would be inefficient. Therefore, the effect

that the characteristics presented among themselves were mostly low, making the use of this practice unfeasible due to the small gains that could be obtained.

4.2. Predictive ability

In order to increase selection efficiency, choosing the best methodology for analyzing and predicting the genetic values of individuals is one of the fundamental aspects to be taken into account in a breeding program. In this study, we compare G-BLUP ST e G-BLUP MT models for the evaluation of genomic prediction in a base population of juçaizeiro. The existence of genetic correlations between the selection traits is the basis for the advantages presented by G-BLUP MT (Jia and Jannink, 2012) and the absence of correlation could lead to equivalent or even superior results when using G-BLUP ST (Jia and Jannink, 2012). Therefore, we can conclude that, depending on the analyzed traits and the existence of correlations, it is expected that the use of the GBLUP MT model will increase the accuracy of the evaluation predictions (Tsuruta et al., 2011).

However, the superiority of G-BLUP MT over G-BLUP ST for $r_{y,\hat{u}}$ was not observed in the present work. The results obtained by the G-BLUP MT model showed similar values of $r_{y,\hat{u}}$ when compared to the G-BLUP ST model, which may have been caused by the linear associations that varied from low to moderate intensity (≤ 0.42) (Fig. 2). Therefore, the present scenario of low association between the characteristics evaluated initially suggests the advantage that the use of genomic selection can provide to multi-trait models. As reported by Calus and Veerkamp (Calus and Veerkamp, 2011), who observed an improvement in $r_{y,\hat{u}}$ when marker information was added to the model for the traits that had associations less than 0.50. In this sense, it is expected that the flow of genetic and phenotypic information provides higher $r_{y,\hat{u}}$ when compared to G-BLUP ST.

The slightly lower performance of G-BLUP MT for predictive abilities can be explained

by the low genetic and residual correlations between traits, as reported by Runcie and Cheng (2019), who attribute such results to the fact that low correlations can cause imprecision in the estimates of genetic and residual covariance parameters may result in reduced model performance. However, as the $r_{y,\hat{u}}$ were close, there is no evidence of such negative effects on the covariance estimates, deducing that the flow of information between the characteristics was not sufficient to improve the $r_{y,\hat{u}}$ of the model G-BLUP MT in relation to the G-BLUP ST, considering that marker information was used by both models.

For the five characteristics evaluated, the biggest difference between the $r_{y,\hat{u}}$ for the methods tested was for RL, which presented values of 0.29 and 0.22, for the G-BLUP-ST and G-BLUP MT, respectively. The superiority of the G-BLUP-ST model diverges from patterns observed by other studies (Arojju et al., 2020; Bhatta et al., 2020; Guo et al., 2014; Jia and Jannink, 2012; Okeke et al., 2017; Rutkoski et al., 2012; Tsai et al., 2020; Tsuruta et al., 2011), in which superiority is observed for the G-BLUP MT model. As it was designed to benefit from the existence of genetic correlations between traits, it is expected that the G-BLUP MT model will present better results than the G-BLUP ST (Calus and Veerkamp, 2011; Schulthess et al., 2016). Due to the similarity of the results of the predictive ability of the models G-BLUP ST and G-BLUP MT, it is not possible, just for this parameter, to indicate the best model for the selection of juçaizeiro genotypes.

4.3. Matrix selection and expected genetic advancement

To build the second selective cycle breeding population, in order to maintain genetic variability, a selection intensity of approximately 20% was defined. In this condition, the Kappa agreement observed among the approaches evaluated was classified as substantial (0.68),

considering that its value was between 0.60 and 0.80 (Munoz and Bangdiwala, 1997). This similarity between the approaches can be explained by the fact that as the variables under study present low to moderate correlations (Evans et al., 1996), both methods ended up leading to similar results.

We should add that the high Kappa agreement between the methods is also associated to the conditions of the phenotypic data set used in the study, which has a low degree of imbalance. Thus, the predictive abilities of the approaches did not have a great impact on the selection of different genotypes between models; however, the influence on the ranking of the best individuals was strongly impacted, as shown in Fig. 3.

When considering the six best genotypes, the selection of the best genotypes had greater divergence between the models, in which the Kappa value was 0.50, evidencing the need to determine the best model to carry out the selection process, aiming at greater future gains. In this way, the models are chosen according to their ability to predict more accurately the GEBV's of the analyzed individuals. However, in the present study, $r_{y,\hat{u}}$ was not sufficient to determine the best method, due to the similar estimates observed.

Due to the similarities of the $r_{y,\hat{u}}$ and the great divergence in the ranking for the Kappa agreement, the choice of the best model for the selection of the juçaizeiro genotypes was based on the SG. In order to standardize the comparison between the models and eliminate the differences in the GEBVs associated with the form of prediction, the SG of each trait was estimated by the corrected phenotypic means and heritability from each model for the populations selected for the next selection cycle and seed donors. In this sense, the G-BLUP ST was chosen for the selection of juçaizeiro genotypes, because this model provided higher SG, in addition to being a simpler and more usual method for breeding programs.

The production of fruits per plant is a feature that serves both the productive sector and the industrial sector, as it provides greater economic return to producers and increases the amount of raw material for industries that currently suffer from a shortage of this base product for their operations. In this sense, observing the difference in selection by phenotypic values, the change in the average fruit production per plant has an increase of 9.67 kg, which indicates a great advance for the culture due to the possibility of achieving great gains in productivity, resulting in an incentive to install of culture.

Due to the juçaizeiro is a wild and perennial species, which presents a series of characteristics intrinsic to its development, the practical activities of conventional breeding techniques become more difficult. In general, when evaluating the structure of a breeding program aimed at juçaizeiro, it is expected that the gains obtained per unit of time will be reduced. In this sense, the present study stands out for applying genomic selection techniques in *Euterpe edulis*, a wild open-pollinated species, aiming to increase gains per unit of time.

Therefore, perhaps the biggest advantage brought by this methodology for a wild breeding population, as is the case of *Euterpe edulis*, is the use of the genomic matrix applied to the relationship correction. Therefore, the insertion of this information provides more accurate estimates and, consequently, greater reliability to the program. Furthermore, the use of genomic selection may allow a reduction in the selection cycle by being able to predict the behavior of non-evaluated genotypes.

All the traits evaluated in the study influence the economic potential of the species, covering the interests of the productive sector and the industrial sector. Additionally, in function of the alterations observed by the selection of the six superior genotypes, we can conclude that the genomic prediction using the G-BLUP ST was efficient to provide alterations of the means in

the desired directions. That is, obtaining an increase in the average phenotypic response of the selected genotypes for the characters of NB, MFB, PFP and PY, and a reduction for EDF.

5. Conclusion

Our results showed that the GBLUP-ST genomic prediction was more efficient in selecting the best genotypes, so that the selection provided substantial gains in the desired direction for the multiple traits. Thus, the six selected genotypes (UFES.A.RN.390, UFES.A.RN.386, UFES.A.RN.080, UFES.A.RN.383, UFES.S.RN.098 e UFES.S.RN.093) can be used as commercial seed donor matrices for the development of seedlings for the implantation of productive orchards, which will meet the demands of the productive sector and the consumer market.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the Conselho Nacional de Pesquisa (CNPq, Brazil) (Researcher productivity fellowship AF; MFSF #311840/2019-1), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) – Finance Code 001, and the Fundação de Amparo à Pesquisa do Espírito Santo (FAPES, Vitória – ES, Brazil) in partnership with VALE, for the financial support to this study. We thank the companies, Açaí Juçara and Bonaloti, for supporting research development and we are also grateful to Pedro and Vicente Bortoloti, the owners of the managed area. We would like to thank all colleagues at the Biometrics and Genetics and

Plant Improvement laboratories from the Universidade Federal do Espírito Santo, for helping to process and organize the samples.

CRediT authorship contribution statement

Guilherme Bravim Canal: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. **Cynthia Aparecida Valiati Barreto:** Methodology, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing. **Marcia Flores da Silva Ferreira:** Conceptualization, Resources, Writing - Review & Editing, Funding acquisition. **Camila Ferreira Azevedo:** Conceptualization, Methodology, Writing - Review & Editing. **Moysés Nascimento:** Conceptualization, Methodology, Writing - Review & Editing. **Francine Alves Nogueira de Almeida:** Data Curation, Methodology. **Diego Pereira do Couto:** Data Curation, Methodology. **Iasmine Ramos Zaidan:** Data Curation, Methodology. **Adésio Ferreira:** Conceptualization, Methodology, Resources, Formal analysis, Investigation, Writing - Review & Editing, Funding acquisition, Project administration.

Data availability statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Alkimim, E.R., Caixeta, E.T., Sousa, T.V., Resende, M.D.V., da Silva, F.L., Sakiyama, N.S., Zambolim, L., 2020. Selective efficiency of genome-wide selection in *Coffea canephora* breeding. *Tree Genet. Genomes* 16, 1–11.

Arojju, S.K., Cao, M., Trolove, M., Barrett, B.A., Inch, C., Eady, C., Stewart, A., Faville,

- M.J., 2020. Multi-Trait Genomic Prediction Improves Predictive Ability for Dry Matter Yield and Water-Soluble Carbohydrates in Perennial Ryegrass. *Front. Plant Sci.* 11. <https://doi.org/10.3389/fpls.2020.01197>
- 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, Genet.* 10, 1113–1124.
- Bush, D., Thumma, B., 2013. Characterising a *Eucalyptus cladocalyx* breeding population using SNP markers. *Tree Genet. Genomes* 9, 741–752. <https://doi.org/10.1007/s11295-012-0589-1>
- Calus, M.P.L., Veerkamp, R.F., 2011. Accuracy of multi-trait genomic selection using different methods. *Genet. Sel. Evol.* 43, 1–14.
- Carvalho, M.S., Ferreira, M.F. da S., Oliveira, W.B. dos S., Marçal, T. de S., Guilhen, J.H.S., Mengarda, L.H.G., Ferreira, A., 2020. Genetic diversity and population structure of *Euterpe edulis* by REML/BLUP analysis of fruit morphology and microsatellite markers. *Crop Breed. Appl. Biotechnol.* 20.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46.
- Covarrubias-Pazaran, G., 2016. Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One* 11, e0156744.
- Cruz, C.D., Regazzi, A.J., Carneiro, P.C.S., 2012. Modelos biométricos aplicados ao melhoramento. UFV, Viçosa.
- de Oliveira Maciel, L., de Moura, N.F., Leonardi, A., 2019. Cadeia produtiva do açaí juçara

na região do litoral norte do Rio Grande do Sul. Rev. Teor. e Evidência Econômica 25,
29–53.

de Resende, M.D. V, Ramalho, M.A.P., Guilherme, S.R., de FB Abreu, Â., 2015.

Multigeneration index in the within-progenies bulk method for breeding of
self-pollinated plants. Crop Sci. 55, 1202–1211.

Doyle, J.J., 1990. Isolation of plant DNA from fresh tissue. Focus (Madison). 12, 13–15.

El-Dien, O.G., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., El-Kassaby, Y.A., 2018.

Multienvironment genomic variance decomposition analysis of open-pollinated Interior
spruce (*Picea glauca x engelmannii*). Mol. Breed. 38, 26.

<https://doi.org/10.1007/s11032-018-0784-3>

El-Kassaby, Y.A., Cappa, E.P., Liewlaksaneeyanawin, C., Klápště, J., Lstibůrek, M., 2011.

Breeding without Breeding: Is a Complete Pedigree Necessary for Efficient Breeding?
PLoS One 6, e25737. <https://doi.org/10.1371/journal.pone.0025737>

Evans, G., Heath, A., Lalljee, M., 1996. Measuring left-right and libertarian-authoritarian
values in the British electorate. Br. J. Sociol. 93–112.

Fanelli Carvalho, H., Galli, G., Ventrone Ferrão, L.F., Vieira Almeida Nonato, J., Padilha, L.,

Perez Maluf, M., Ribeiro de Resende Jr, M.F., Guerreiro Filho, O., Fritsche-Neto, R.,
2020. The effect of bienniality on genomic prediction of yield in arabica coffee.

Euphytica 216, 101. <https://doi.org/10.1007/s10681-020-02641-7>

Farias Neto, J.T. de, Oliveira, M. do S.P. de, Yokomizo, G.K.I., 2016. Ganho esperado na

seleção de progênies de polinização aberta de *Euterpe oleracea* para produção de frutos.
Sci. For. 44. <https://doi.org/10.18671/scifor.v44n109.23>

Farias Neto, J.T. de, Resende, M.D.V. de, Oliveira, M. do S.P. de, Nogueira, O.L., Falcão,

P.N.B., Santos, N.S.A. dos, 2008. Estimativas de parâmetros genéticos e ganhos de

seleção em progênes de polinização aberta de açaizeiro. Rev. Bras. Frutic. 30, 1051–

1056.

Gaire, R., 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. Plant Genome.

<https://doi.org/10.1002/tpg2.20188>

Gamal El-Dien, O., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., El-Kassaby, Y.A., 2016.

Implementation of the realized genomic relationship matrix to open-pollinated white

spruce family testing for disentangling additive from nonadditive genetic effects. G3

Genes, Genomes, Genet. 6, 743–753.

Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L., Su, G., 2014. Comparison of single-trait and

multiple-trait genomic prediction models. BMC Genet. 15, 1–7.

Jarquín, D., Howard, R., Graef, G., Lorenz, A., 2019. Response Surface Analysis of Genomic

Prediction Accuracy Values Using Quality Control Covariates in Soybean. Evol.

Bioinforma. 15, 117693431983130. <https://doi.org/10.1177/1176934319831307>

Jia, Y., Jannink, J.-L., 2012. Multiple-Trait Genomic Selection Methods Increase Genetic

Value Prediction Accuracy. Genetics 192, 1513–1522.

<https://doi.org/10.1534/genetics.112.144246>

Klápště, J., Suontama, M., Dungey, H.S., Telfer, E.J., Graham, N.J., Low, C.B., Stovold,

G.T., 2018. Effect of Hidden Relatedness on Single-Step Genetic Evaluation in an

Advanced Open-Pollinated Breeding Program. J. Hered.

<https://doi.org/10.1093/jhered/esy051>

bioRxiv preprint doi: <https://doi.org/10.1101/2022.09.19.508517>; this version posted September 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

641 Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., Gutiérrez, L., 2018. Resource
642 allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum*
643 *aestivum* L.) baking quality. Theor. Appl. Genet. 131, 2719–2731.
644 <https://doi.org/10.1007/s00122-018-3186-3>

645 Legarra, A., Robert-Granié, C., Manfredi, E., Elsen, J.-M., 2008. Performance of Genomic
646 Selection in Mice. Genetics 180, 611–618. <https://doi.org/10.1534/genetics.108.088575>

647 Lynch, M., Walsh, B., 1998. Genetics and analysis of quantitative traits.

648 Marçal, T. de S., Ferreira, A., Oliveira, W.B. dos S., Guilhen, J.H.S., Ferreira, M.F. da S.,
649 2015. Correlações genéticas e análise de trilha para caracteres de fruto da palmeira
650 juçara. Rev. Bras. Frutic. 37, 692–698.

651 Massman, J.M., Gordillo, A., Lorenzana, R.E., Bernardo, R., 2013. Genomewide predictions
652 from maize single-cross data. Theor. Appl. Genet. 126, 13–22.

653 Meuwissen, T.H.E., Hayes, B.J., Goddard, M., 2001. Prediction of total genetic value using
654 genome-wide dense marker maps. Genetics 157, 1819–1829.

655 Mulamba, N.N., Mock, J.J., 1978. Improvement of yield potential of the ETO blanco maize
656 (*Zea mays* L.) population by breeding for plant traits [Mexico]. Egypt. J. Genet. Cytol.

657 Munoz, Sergio R., Bangdiwala, S.I., 1997. Interpretation of Kappa and B statistics measures
658 of agreement. J. Appl. Stat. 24, 105–112. <https://doi.org/10.1080/02664769723918>

659 Munoz, Sergio R, Bangdiwala, S.I., 1997. Interpretation of Kappa and B statistics measures of
660 agreement. J. Appl. Stat. 24, 105–112.

661 Okeke, U.G., Akdemir, D., Rabbi, I., Kulakow, P., Jannink, J.-L., 2017. Accuracies of
662 univariate and multivariate genomic prediction models in African cassava. Genet. Sel.
663 Evol. 49, 88. <https://doi.org/10.1186/s12711-017-0361-y>

bioRxiv preprint doi: <https://doi.org/10.1101/2022.09.19.508517>; this version posted September 19, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

664 Oliveira, W.B. dos S., Ferreira, A., Gullien, J.H.S., Marçal, I. de S., Ferreira, M.F. da S.,
665 Senra, J.F. de B., 2015. Path analysis and genetic diversity of *Euterpe edulis* Martius for
666 vegetative and fruit traits. Sci. For. 43, 303–311.

667 R CORE TEAM, 2021. R: A language and environment for statistical computing.

668 Resende, M.D. V, Silva, F.F., Azevedo, C.F., 2014. Estatística matemática, biométrica e
669 computacional: modelos mistos, multivariados, categóricos e generalizados
670 (REML/BLUP), inferência bayesiana, regressão aleatória, seleção genômica, QTL-
671 QWAS, estatística espacial e temporal, competição, sobrevivência. Suprema, Visconde
672 do Rio Branco 1.

673 Runcie, D., Cheng, H., 2019. Pitfalls and Remedies for Cross Validation with Multi-trait
674 Genomic Prediction Methods. G3 Genes|Genomes|Genetics 9, 3727–3741.
675 <https://doi.org/10.1534/g3.119.400598>

676 Rutkoski, J., Benson, J., Jia, Y., Brown-Guedira, G., Jannink, J.-L., Sorrells, M., 2012.
677 Evaluation of Genomic Prediction Methods for *Fusarium* Head Blight Resistance in
678 Wheat. Plant Genome 5, 51–61. <https://doi.org/10.3835/plantgenome2012.02.0001>

679 Sapkota, S., Boatwright, J.L., Jordan, K., Boyles, R., Kresovich, S., 2020. Multi-Trait
680 Regressor Stacking Increased Genomic Prediction Accuracy of Sorghum Grain
681 Composition. Agronomy 10, 1221. <https://doi.org/10.3390/agronomy10091221>

682 Schulthess, A.W., Wang, Y., Miedaner, T., Wilde, P., Reif, J.C., Zhao, Y., 2016. Multiple-
683 trait- and selection indices-genomic predictions for grain yield and protein content in rye
684 for feeding purposes. Theor. Appl. Genet. 129, 273–287. [https://doi.org/10.1007/s00122-](https://doi.org/10.1007/s00122-015-2626-6)
685 015-2626-6

686 Silva, J.Z. da, Reis, M.S. dos, Schulz, M., Borges, G. da S.C., Gonzaga, L.V., Costa, A.C.O.,

- Fett, R., Siqueira, A.P.S., Santos, K.F. dos, Barbosa, T.A., Freire, L. de A.S., Camêlo, Y.A., de Oliveira Maciel, L., de Moura, N.F., Leonardi, A., 2018. Fenologia Reprodutiva e Produção de Frutos em *Euterpe edulis* (Martius). *Ciência Florest.* 28, 295–309. <https://doi.org/10.5902/1980509831655>
- Sousa, T.V., Caixeta, E.T., Alkimim, E.R., Oliveira, A.C.B., Pereira, A.A., Sakiyama, N.S., Zambolim, L., Resende, M.D.V., 2019. Early Selection Enabled by the Implementation of Genomic Selection in *Coffea arabica* Breeding. *Front. Plant Sci.* 9. <https://doi.org/10.3389/fpls.2018.01934>
- Teixeira, D.H.L., Oliveira, M. do S.P. de, Gonçalves, F.M.A., Nunes, J.A.R., 2012. Correlações genéticas e análise de trilha para componentes da produção de frutos de açaizeiro. *Rev. Bras. Frutic.* 34, 1135–1142.
- 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, Genet.* 10, 3751–3763.
- Tsai, H.-Y., Cericola, F., Edriss, V., Andersen, J.R., Orabi, J., Jensen, J.D., Jahoor, A., Janss, L., Jensen, J., 2020. Use of multiple traits genomic prediction, genotype by environment interactions and spatial effect to improve prediction accuracy in yield data. *PLoS One* 15, e0232665. <https://doi.org/10.1371/journal.pone.0232665>
- Tsuruta, S., Misztal, I., Aguilar, I., Lawlor, T.J., 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *J. Dairy Sci.* 94, 4198–4204. <https://doi.org/10.3168/jds.2011-4256>
- VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423.

