



OPEN Uncovering morphometric, germination, and genetic divergence patterns in *Euterpe edulis* for breeding and conservation applications

Guilherme Bravim Canal¹, Marcello Zatta Péres², Francine Alves Nogueira de Almeida¹, Maurício dos Santos Araújo³, Jéssica Tetzner de Oliveira², Jônatas Gomes Santos¹, Gilza Barcelos de Souza¹, Marcia Flores da Silva Ferreira¹, Rodrigo Sobreira Alexandre² & Adésio Ferreira¹✉

Euterpe edulis exhibits significant ecological, social, and economic potential, comparable to *E. oleracea*, for both the fresh fruit pulp market and industrial applications as a natural pigment. The commercial management of *E. edulis* fruits is economically promising; however, further studies are required to support improvement, conservation, and propagation programs. In this study, we aimed to assess the genetic control of key traits by estimating variance components, genetic parameters, and the associative relationships between fruit morphometric traits, seed characteristics, and seedling emergence using Pearson correlation and path analysis. Additionally, we quantified the genetic divergence among *E. edulis* accessions. Thirteen phenotypic traits and eight microsatellite markers were analyzed in 72 genotypes from different origins. High heritability values (ranging from 0.75 to 0.99) were observed for fruit, seed, and emergence traits, whereas morphophysiological growth traits exhibited lower heritability estimates. Genetic correlations of up to 0.88 were detected among fruit morphometric traits. Seedling basal diameter and leaf area had direct positive effects on the seedling quality index, while seedling height exerted a direct negative effect. Genetic divergence analysis demonstrated the efficiency of sampling genotypes from different origins to preserve the species' genetic variability. These findings provide insights to guide the establishment of germplasm collections in the field, maximizing the potential for genetic recombination among divergent genotypes and offering a solid foundation for future studies employing expanded molecular resources to explore trait architecture more deeply.

The Atlantic Forest, the second-largest forest in South America, is one of the most threatened and species-rich biomes in the world¹ and is considered one of the 35 biodiversity hotspots on the planet². Due to its remarkable biodiversity, protecting this biome is essential for the conservation of threatened species and their interactions within the ecosystem³. Among its vast array of plant genetic resources, fruit-bearing palms play a crucial ecological role by providing food during periods of scarcity in the forest⁴, particularly those of the genus *Euterpe*. Notably, *E. edulis* Mart. stands out⁵. Commonly known as the juçara or juçaizeiro palm, this species holds significant ecological, cultural, and economic value⁶. However, human pressures on its natural resources, combined with the fragmentation of the biome, have led to its classification as a threatened species^{7,8}.

The juçara palm (*E. edulis*) exhibits significant economic potential through the sustainable management of its fruits, in addition to its crucial ecological role. The species has the capacity to support a production chain analogous to that of the açaí palm (*E. oleracea*)^{9,10}. The fruits of the juçara palm possess high nutraceutical value and substantial potential as a functional food, mainly associated with their high anthocyanin content¹¹. Furthermore, the pulp exhibits nutritional characteristics comparable to or superior to those of açaí pulp (*E.*

¹Department of Agronomy, Federal University of Espírito Santo, Alegre, Espírito Santo, Brazil. ²Department of Forest and Wood Sciences, Federal University of Espírito Santo, Jerônimo Monteiro, Espírito Santo, Brazil.

³Genetics Department, University of São Paulo, Genetic Diversity and Plant Breeding Laboratory, Piracicaba, SP, Brazil. ✉email: adesioferreira@gmail.com

oleracea)^{12–14}. However, the commercial exploitation of juçara requires the establishment of cultivated orchards to ensure economic viability. Industrial processors in the state of Espírito Santo report that the current supply of juçara fruits is insufficient to meet market demand. Moreover, the production potential is constrained by fruit availability, which is predominantly sourced through sustainable management in native forest fragments. To advance the commercial cultivation of the species, the development of superior genetic materials through breeding programs is essential. Such programs rely initially on a breeding foundation population with high genetic diversity and a thorough understanding of the genetic control of key traits, which are critical for defining effective strategies within the breeding program.

The estimation and prediction of genetic parameters and variance components in perennial species are typically conducted using the REML/BLUP method (restricted maximum likelihood/best linear unbiased prediction)¹⁵. In particular, for species of the genus *Euterpe*, this approach mitigates challenges associated with data collection, such as imbalance and the absence of a structured experimental design. By enabling the correction of observations, the REML/BLUP method enhances the accuracy of variance component estimates¹⁶, thereby improving the understanding of the genetic control of key traits.

Within breeding programs and seedling nursery practices, understanding the relationships between phenotypic traits and genetic diversity is fundamental for the effective conservation of genetic resources and the development of superior genotypes. Assessing genetic diversity is crucial to prevent the use of genetically similar materials, thereby reducing the risk of allele loss due to inbreeding. Consequently, analyses of both phenotypic and molecular diversity play a critical role in genotype characterization and selection, enhancing the efficient utilization of available genetic resources, ensuring the long-term sustainability of breeding programs, and contributing to the successful conservation of species' genetic resources. In this context, the objective of this study was to investigate genetic control through the estimation of variance components and genetic parameters, as well as the associative relationships among fruit and seed morphometric traits, seedling emergence, and early development. Additionally, this study aimed to quantify the genetic divergence within *E. edulis* germplasm, providing essential information to support the establishment of *ex situ* germplasm collections.

Material and methods

Plant material

Fruits from 72 *E. edulis* trees were sampled from fragments of the Atlantic Forest located on private properties across eight municipalities in the state of Espírito Santo, Brazil: Alegre, Domingos Martins, Dorcas do Rio Preto, Guaçuí, Ibitirama, Rio Novo do Sul, São José dos Calçados, and Venda Nova do Imigrante (Fig. 1). The selection criteria for the trees included a minimum distance of 150 meters between sampled individuals, good phytosanitary and physiological conditions, and superior fruit production potential compared to neighboring trees. The selected trees were located at altitudes ranging from 500 to 1500 meters. Voucher specimens of *E. edulis* were collected in Espírito Santo: VIES039447¹⁷, VIES038826¹⁸, and MBM00422915¹⁹. The specimens were deposited in the herbaria VIES (Universidade Federal do Espírito Santo) and MBM (Museu Botânico Municipal). Voucher information and high-resolution images are available through the Re flora Virtual Herbarium. All

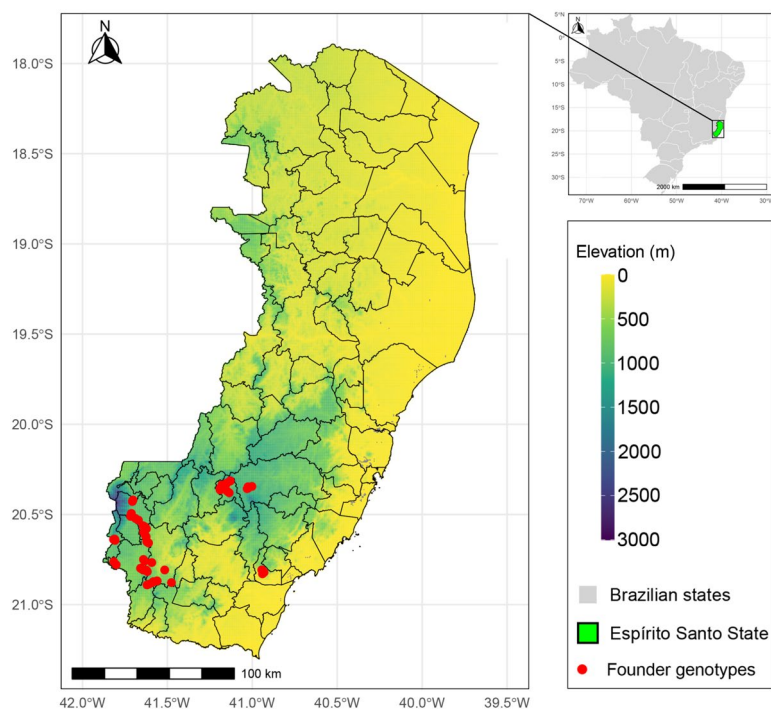


Fig. 1. Geographic location of sampled *E. edulis* founder genotypes, indicated by red dots, in the state of Espírito Santo, Brazil. Map generated using the R software.

collections were conducted with prior consent from landowners and duly authorized by the Brazilian Ministry of the Environment through the Chico Mendes Institute for Biodiversity Conservation (ICMBio/SISBIO) (Decisions n.o 59344-2 and 87764-3), following national regulations for scientific research.

Fruits were collected at full maturation, characterized by a blackish pulp color. Additionally, a fragment of the stipe cortex from each founder genotypes was collected, placed in Kraft paper bags containing silica gel, properly labeled, and transported to the Laboratory of Genetics and Plant Breeding (LGMV) at the Federal University of Espírito Santo (UFES), Alegre campus. Samples were stored at -80°C for 24 hours before being lyophilized for 72 hours.

Experimental conduction and phenotyping

The following morphometric traits were evaluated in fruits: Equatorial Diameter of the Fruit (EDF) in millimeters, measured using a digital caliper with a precision of 0.01 mm; Weight of 25 Fruits (WF) and Weight of 25 Seeds (WS) (g). The evaluations were conducted in a completely randomized design.

For seedling emergence analyses, the following traits were assessed: Emergence Speed Index (ESI), Mean Time to Emergence (MTE), and Emergence percentage (E) (%). Evaluations were conducted over 153 days in a greenhouse, using a completely randomized design with four replicates of 25 seeds per genotype. Seeds underwent dormancy breaking by soaking in water at 40°C for 20 minutes²⁰, and emergence was recorded every two days.

For seedling growth assessments, the following traits were measured: Shoot Dry Mass (SDM) (g), Total Dry Mass (TDM) (g), Dickson's Quality Index (DQI)²¹, Leaf Area (LA) (cm^2), Diameter of Seedlings at the Base (DSB) (mm), and Seedling Height (SH) (cm). Evaluations followed a completely randomized design with five assessment time points over 262 days, with measurements taken from 10 replicates per genotype per evaluation. Seedlings were transplanted 153 days after sowing in sand (considered time zero) and acclimatized in a greenhouse under intermittent irrigation for 30 days.

From time zero onwards, destructive analyses were performed on 10 plants per genotype every 50 days, totaling five destructive assessments. On the last evaluation date, an additional 12 days of growth were included to enable chlorophyll fluorescence analysis, resulting in a total growth period of 262 days after transplanting.

Chlorophyll fluorescence was measured using the potential quantum efficiency of photosystem II (Φ_{PSII}) F_v/F_m with a FluorPen FP110 portable fluorometer (Photon Systems Instruments). Measurements were taken from the adaxial part of the largest leaf, in the intermediate region, avoiding necrotic spots and the main vein. Leaves were adapted to darkness for 30 minutes using metal clips before measurement, performed at 8:00 AM and noon, on a total of 248 plants.

EDF was measured from 10 replicates per genotype; WF and WS were evaluated in four replicates of 25 fruits; ESI was calculated according to Maguire's equation²²; MTE was calculated according to Labouriau and Viladares²³; E was evaluated in four replicates of 25 seeds; SDM and TDM were assessed at five different time points with 10 replicates per genotype; DQI was calculated using the methodology of Dickson, Leaf, and Hosner²⁴; SH and DSB were evaluated six times, including time zero, with 10 replicates per genotype.

Genotyping by microsatellite markers

Genomic DNA from 59 each founder genotype was extracted from cortex samples using the protocol described by Carvalho²⁵. DNA quality, concentration, and integrity were verified using a NanoDrop Thermo Scientific® spectrophotometer with a 260/280 absorbance ratio between 1.8 and 2.0, and through 0.8% agarose gel electrophoresis stained with GelRed (Biotium).

Thirteen SSR loci developed by Gaiotto, Brondani, and Grattapaglia²⁶ for *E. edulis* were screened, out of which eight microsatellite markers generated reproducible banding pattern and scoring clarity to ensure the robustness of our dataset. Therefore, these eight markers (EE05, EE09, EE23, EE43, EE45, EE47, EE48 and EE52) were selected for final analysis. The polymerase chain reaction (PCR) was conducted in a final reaction volume of 20 μL following the manufacturer's recommendations. Each reaction contained 30 ng of genomic DNA, 1X commercial Gotaq buffer, 0.3 μM of each primer (forward and reverse), 1.5 mM MgCl_2 , 2.5 μg BSA, 0.25 mM dNTPs, and 1.25 U of Taq DNA polymerase (Promega). Amplifications were performed using a VeritiTM 96-Well Thermal Cycler (Thermo Fisher Scientific) with the following conditions: 2 cycles at 94°C for 4 minutes, 35 cycles at 94°C for 45 seconds, annealing at 55°C for 1 minute, and extension at 72°C for 1 minute, followed by a final extension at 72°C for 10 minutes. PCR products were analyzed via 1.5% agarose gel electrophoresis stained with GelRed (Biotium).

PCR products were diluted based on their concentrations observed in agarose gel electrophoresis. For fragment analysis, 1.0 μL of each sample, 0.25 μL of GeneScanTM 600 LIZTM Dye size standard (Thermo Fisher Scientific), and 8.75 μL of Hi-DiTM formamide (Thermo Fisher Scientific) were used. Samples were denatured at 95°C for 3 minutes in a VeritiTM 96-Well Thermal Cycler and analyzed using a SeqStudio Genetic Analyzer (Thermo Fisher Scientific). Allele fragment sizes were determined using GeneMapper V5.0 software (Softgenetics, State College, PA, USA), considering peaks with relative fluorescence intensity above 350 RFU as valid. A table was compiled containing the allele sizes (bp) detected for each genotype for subsequent genetic diversity analyses.

Data analysis

Estimates for quantifying genetic diversity via microsatellite molecular markers were obtained using the Software R²⁷. The diversity of parameters addressed in the present work were the number of observed alleles (N_A), observed heterozygosity (Ho) (Eq. 1), expected heterozygosity (He)²⁸ (Eq. 2), inbreeding coefficient (F) (Eq. 3) and the however polymorphic information measure (PIC)²⁹ (Eq. 4). The estimators of the mentioned parameters are presented below:

$$H_o = \frac{NH_o}{N} \quad (1)$$

where: H_o is the frequency of observed heterozygotes or observed heterozygosity; NH_o is the number of observed heterozygotes and N is the sample size.

The unbiased estimator of expected heterozygosity (H_e) for genetic diversity within a population used in the present study was:

$$H_e = \frac{N}{N-1} \left(1 - \sum_{i=1}^a p_i^2 \right) \quad (2)$$

where: H_e is the unbiased estimator of expected heterozygosity; p_i is the frequency of the i -th allele of the locus under study.

The inbreeding coefficient F can be estimated from the observed heterozygosity (H_o) and expected heterozygosity (H_e) using the following equation:

$$F = 1 - \left(\frac{H_o}{H_e} \right) \quad (3)$$

where: F is the inbreeding coefficient.

The Polymorphic Information Content (PIC) is a measure of the informativeness of a genetic marker. It is calculated as:

$$PIC = 1 - \sum_{i=1}^a p_i^2 - \sum_{i,j=1, i \neq j}^a p_i^2 p_j^2 \quad (4)$$

where: p_j is the frequency of the j -th allele of the locus under study.

The estimates of variance components, genetic parameters, and genetic value predictions for the genotypes under study were performed individually for each trait using the restricted maximum likelihood method (REML) and the best linear unbiased prediction (BLUP)^{30,31}. For all traits, the Likelihood Ratio Test (LRT) was conducted to evaluate the significance of the effects, ensuring that the final model included only significant effects (Table S1).

The adjusted model for biometric variables is given by:

$$y = 1\mu + X_1b + Z_1g + \varepsilon_1 \quad (5)$$

where y is the data vector ($n \times 1$), where n is the number of phenotypic observations; μ is a scalar corresponding to the model intercept; X_1 is an incidence matrix of dimension $n \times r$, where r is the number of replications; b represents the fixed effects vector with dimensions $r \times 1$; Z_1 is an incidence matrix of dimension $n \times i$ associated with the vector of random effects of genetic g with dimensions $i \times 1$, where i is the number of genotypes, $g \sim N(0, \sigma_g^2 I_i)$ where σ_g^2 is the genetic variance; and ε_1 is the residual error vector with dimensions $n \times 1$, $\varepsilon_1 \sim N(0, \sigma_\varepsilon^2 I_n)$, σ_ε^2 is the residual variance. I_x is an identity matrix with a dimension determined by its subscript [$x = i$ is the number of genotypes; n is the number of phenotypic observations]. The full model initially tested included the fixed effect of field fruit samplings and the random effect of sampling elevation. However, both effects were not significant based on the corrected AIC and BIC criteria for fixed effects³², as well as the AIC and BIC tests, respectively.

For emergence variables, the adjusted model includes an additional random effect:

$$y = 1\mu + X_2b + Z_1g + Z_2al + \varepsilon_2 \quad (6)$$

where X_2 is an incidence matrix of dimension $n \times m$, where m is the number of replications (r) added to the number of field fruit samplings (c); b represents the fixed effects vector with dimensions $m \times 1$; Z_2 is an incidence matrix of dimension $n \times l$ associated with the vector of random effects of sampling elevation al with dimensions $l \times 1$, where l is the number of sampling elevation, $al \sim N(0, \sigma_{al}^2 I_l)$ where σ_{al}^2 is the sampling elevation variance; $\varepsilon_2 \sim N(0, R \otimes I_n)$, σ_ε^2 is the residual variance and R is the variance-covariance (VCOV) residual matrix for the field fruit samplings with dimensions $c \times c$, where c is the number of field fruit samplings; $\otimes$ is the Kronecker product.

The model for seedling growth variables considers genotype-by-evaluation interaction:

$$y = 1\mu + X_3b + Z_1g + Z_3gm + \varepsilon_3 \quad (7)$$

where X_3 is an incidence matrix of dimension $n \times o$, where o is the number of replications (r) added to the number of different evaluations times (t); b represents the fixed effects vector with dimensions $o \times 1$; gm is the $it \times 1$ vector of random effects genotype-by-evaluations, with $gm \sim N(0, G_{gm} \otimes I_i)$; The matrix Z_3 is the incidence matrix for random effect gm ; $\varepsilon_3 \sim N(0, R \otimes I_n)$, σ_ε^2 is the residual variance. G_{gm} is the VCOV matrix for the effect of genotypes between evaluations with dimensions $t \times t$; R is the residual VCOV.

For the physiological variable group (F_v/F_m), the adjusted model simplifies to:

$$y = 1\mu + X_4b + Z_1g + \varepsilon_4 \quad (8)$$

where X_4 is an incidence matrix of dimension $n \times k$, where k is the number of replications (r) added to the number of different evaluations times of the F_v/F_m (h); b represents the fixed effects vector with dimensions $k \times 1$; $\varepsilon_4 \sim N(0, R \otimes I_n)$.

The significance of the effects tested in these models was assessed using LRT, and the corrected Bayesian Information Criterion (BICc)³³ was used to determine the inclusion of collection effects, evaluation times, and measurement times as fixed effects (Table S1).

To improve genetic value predictions, different assumptions were applied for $G(i \times i)$ and $R(n \times n)$, using various covariance structures^{34–37}. The selection of the best model was performed using the Akaike³⁸ and Bayesian³⁹ information criteria (Table S2).

The broad-sense heritability H^2 was estimated as⁴⁰:

$$H_g^2 = 1 - \frac{\overline{\Delta}_{BLUP}}{2\sigma_g^2} \quad (9)$$

where: $\overline{\Delta}_{BLUP}$ represents the average prediction error variance between genotype pairs, and σ_g^2 is the genotypic variance.

Prediction accuracy r was estimated as:

$$r = \sqrt{1 - \frac{PEV}{\sigma_g^2}} \quad (10)$$

where: PEV denotes the prediction error variance.

To evaluate the variability among genotypes, the LRT compared the reduced and full models, with significance assessed using the chi-square (χ^2) test⁴¹. Correlation estimates were computed as:

$$r_{y_1 y_2} = \frac{\text{Cov}(y_1, y_2)}{\sqrt{\sigma_{y_1}^2 \times \sigma_{y_2}^2}} \quad (11)$$

where $\text{Cov}(y_1, y_2)$ denotes the covariance between trait pairs, and $\sigma_{y_1}^2$ and $\sigma_{y_2}^2$ represent the variances of the respective traits.

Before conducting path analysis, the Variance Inflation Factor (VIF) method of Montgomery and Peck⁴² was applied to test for multicollinearity, and traits with $VIF > 5$ were excluded from the analysis. The analysis was based on the phenotypic correlation matrix, considering the main effect characteristic, DQI.

For genetic diversity analysis, only genotypes included in both phenotypic and molecular evaluations (59 founder genotypes) were considered. The analysis consisted of two stages: first, clusters formed by BLUP-predicted genotypic data and microsatellite molecular data were compared using standardized mean Euclidean distance (DEMP) for genotypic data and the unweighted index for molecular data⁴³. In the second stage, dissimilarity matrices were normalized to a scale from zero to one, and an average matrix of both normalized matrices was constructed. The UPGMA method was employed for clustering, and the number of groups was determined using Mojema's criterion⁴⁴ with $k = 1.25$. The association significance between matrices was tested using the Mantel test⁴⁵ with 5000 permutations. All statistical analyses were conducted in R⁴⁶.

Results

Phenotypic data

Substantial variation in phenotypic traits was observed between the founders genotypes, sampled to compose the breeding foundation population, as indicated by the descriptive analysis of the phenotypic data, graphical boxplot analysis, and data distribution, including means for the 13 evaluated traits (Fig. 2 and Fig. 3).

Variance components and genetic parameters

Genetic variability was detected between genotypes for all evaluated traits, as evidenced by the significance of the maximum likelihood ratio test (LRT) using the χ^2 test. These results, along with the estimates of variance components and genetic parameters, are presented in Table 1.

For biometric traits (EDF, WF, and WS) and emergence traits (ESI, MTE, and E), genetic effects had a greater influence on phenotypic expression, as indicated by higher estimates of σ_g^2 compared to σ_e^2 (Table 1). For seedling growth traits (SDM, TDM, DQI, LA, DSB, SH, and F_v/F_m), residual effects had a more pronounced impact on phenotypic expression. Consequently, heritability (H^2) values were higher for biometric and emergence traits (ranging from 0.75 for MTE to 0.99 for WF) compared to seedling growth traits (ranging from 0.22 for TDM to 0.32 for LA).

Genetic and phenotypic correlations

Most correlation estimates among the 13 evaluated traits were significant at 1%. In general, genetic correlations (r_g) were higher than phenotypic correlations (r_f) (Fig. 4A and Fig. 4B). Fifty-three significant phenotypic correlations were identified, with coefficients ranging from -0.64 to 0.96. Among these, 35 were significant at 0.1%, 7 at 1%, 12 at 5%, and 24 were not significant (Fig. 4A). Among genetic correlations, 55 were significant, with coefficients ranging from -0.64 to 0.96, including 37 at 0.1%, 5 at 1%, 14 at 5%, and 22 not significant (Fig. 4B).

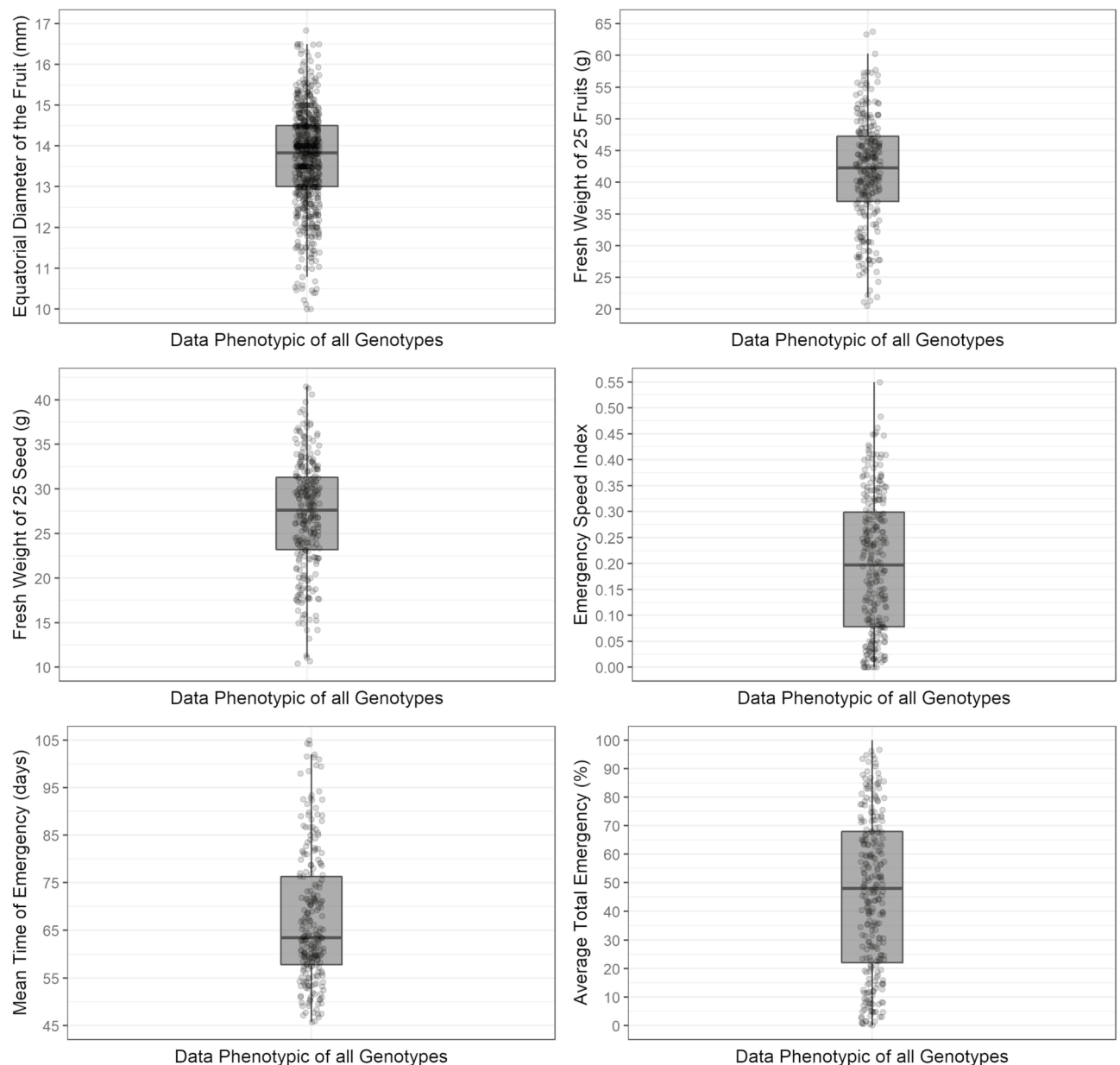


Fig. 2. Figures describing the phenotypic data of all measurements performed on all founders genotypes sampled for the traits equatorial diameter fruit (EDF) (mm); weight of 25 fruits (WF) (g); weight of 25 seeds (WS) (g); emergency speed index (IVE); mean time to emergence (MTE) (days) and emergency (E) (%).

The genetic correlations between fruit biometric traits (EDF, WF, and WS) were all significant and above 0.88, indicating a strong association. For emergence traits (ESI, MTE, and E), both positive and negative correlations were observed, with the highest correlation occurring between ESI and E (0.96) (Fig. 4B), suggesting a strong linear association between early-emerging seedlings and higher emergence rates. MTE exhibited several negative associations, particularly with E (-0.47). The genetic correlations between fruit biometric traits (EDF, WF, and WS) and dry mass seedling traits (SDM and TDM) exhibited moderate association values, ranging from 0.45 to 0.54 (Fig. 4). SDM and TDM showed stronger associations with DSB (0.86 and 0.87, respectively) and SH (0.70 and 0.73, respectively).

The mean genetic correlation (r_g) between the DQI and biometric traits of fruit and seed (EDF, WF, and WS) ranged from 0.44 to 0.46, indicating a moderate association between fruit/seed size and seedling quality. Among emergence traits, DQI was significantly correlated only with E (0.28), demonstrating a clear linear association. For growth traits, there was a strong positive correlation between DSB and SH (0.74), showing that seedling height growth follows diameter growth. The physiological parameter F_v/F_m showed no significant linear associations with most traits, except for E (-0.35).

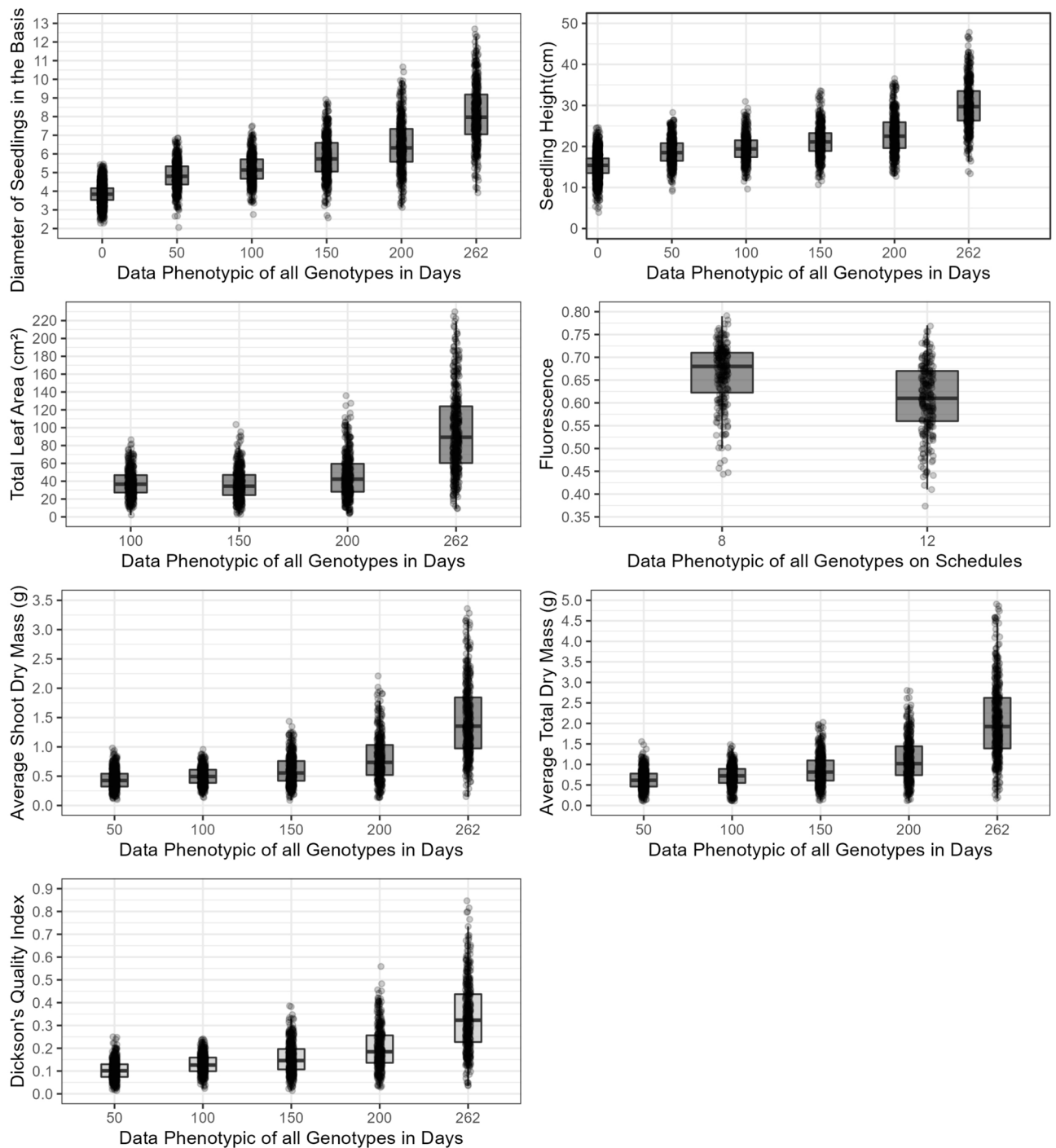


Fig. 3. Figures describing the phenotypic data of all measurements performed on all founders genotypes sampled for the traits diameter to collar height (DSB) (cm); seedling height (SH) (cm); leaf area (LA) (cm²) and chlorophyll fluorescence (F_v/F_m), shoot dry mass (SDM) (g); total dry mass (TDM) (g), Dickson quality index (DQI). The $\times$ axis is defined for measurement days (DSB, SH, LA, F_v/F_m , SDM, TDM, DQI) and measurement time (F_v/F_m).

Path analysis

Correlation analysis alone can lead to misinterpretations due to indirect effects among explanatory traits. Path analysis was performed to partition the estimates of phenotypic correlations into direct and indirect effects. The analysis is presented in Fig. 5, illustrating the direct effects of explanatory variables on DQI and their correlations (top).

The multicollinearity study was conducted, and traits WF, ESI, MTE, and E were removed to mitigate variance inflation factors. The coefficient of determination (R^2) was 80.57%, with a residual effect (EVR) of 0.44,

Group	Traits	H^2	r	AIC	BIC	LRT1	LRT2
Biometric	EDF	0.98	0.99	89.60	98.74	909.7***	–
	WF	0.99	0.99	727.41	734.72	847.39***	–
	WS	0.99	0.99	641.22	648.54	761.91***	–
Emergence	ESI	0.29	0.72	-1079.09	-1061.21	237.05***	0.12ns
	MET	0.76	0.87	1423.06	1437.23	173.91***	–
	E	0.69	0.83	1424.76	1442.47	224.66***	1.01×10^{-5} ns
Seedling Growth	DSB	0.87	0.93	326.23	409.40	383.41***	997.38***
	SH	0.90	0.95	13210.19	13267.65	561.34***	718.48***
	LA	0.85	0.92	12450.09	12499.53	160.71***	227.04***
	SDM	0.81	0.90	-7133.73	-7070.29	158.12***	1561.1***
	TDM	0.83	0.91	-3081.11	-3017.20	166.86***	261.73***
	DQI	0.84	0.91	-10636.38	-10573.81	190.34***	425.63***
Physiological	F_v/F_m	0.71	0.84	-2072.45	-2059.94	37.34***	-3.86×10^{-5} ns

Table 1. Estimates of variance components and genetic parameters in a breeding foundation population of *E. edulis* for: equatorial diameter fruit (EDF) (mm), weight of 25 fruits (WF) and seed (WS) (g), emergence speed index (ESI), mean emergence time (MET), percentage of emergence (E), shoot dry mass (SDM) (g), total dry mass (TDM) (g), Dickson's quality index (DQI), leaf area (LA) (cm²), diameter at stem height (DSB) (mm), seedling height (SH) (cm) and variable fluorescence/maximal fluorescence (F_v/F_m). LRT1: Maximum likelihood test for the variance component σ_{gat}^2 . LRT2: Maximum likelihood test for interaction effects. The tested effects were: Emergence traits: σ_{gat}^2 (variance of genotype $\times$ elevation interaction). Seedling growth traits: σ_{gm}^2 (variance of genotype $\times$ measurement interaction). Physiological traits: σ_{gh}^2 (variance of genotype $\times$ time interaction). *** Significant based on a χ^2 test with 1 degree of freedom at the 1% probability level.

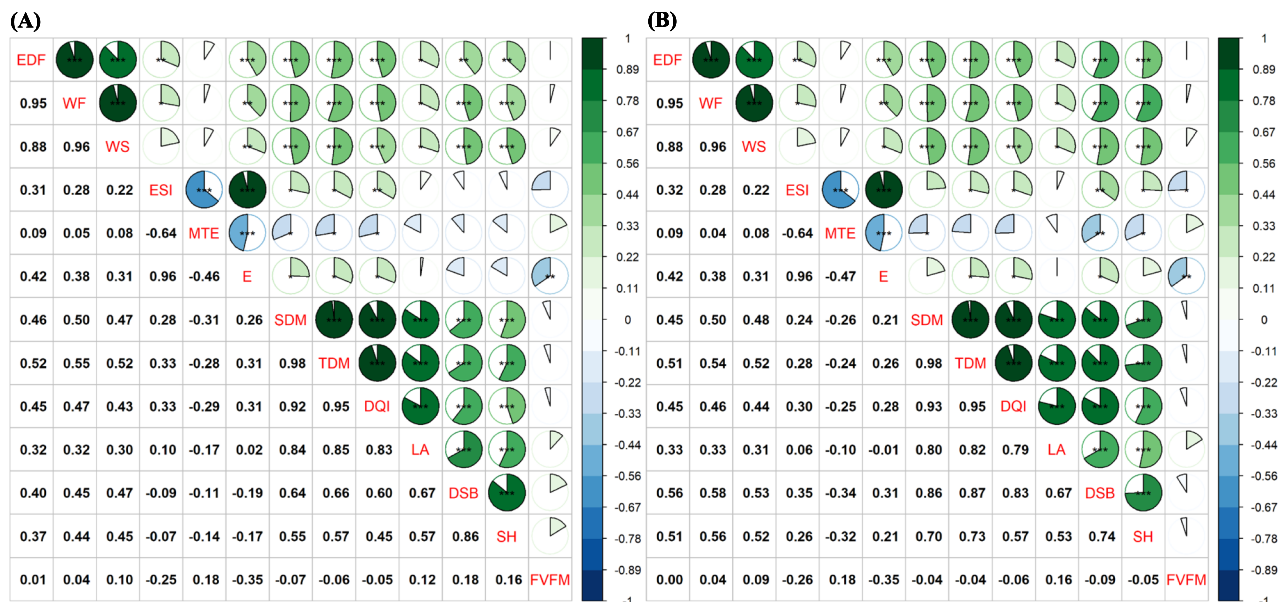


Fig. 4. Phenotypic (rf) (A) and genetic (rg) (B) correlation between traits. On the upper diagonal, the significance and correlation are presented graphically, while on the lower diagonal, the correlation is presented numerically for the characteristics, namely: equatorial diameter of the fruit (EDF) (mm), weight of 25 fruits (WF) and seed (WS) (g), emergence speed index (ESI), mean emergence time (MTE), percentage of emergence (E), shoot dry mass (SDM) (g), total dry mass (TDM) (g), Dickson quality index (DQI), total leaf area (LA) (cm²), diameter of seedling in the basis (DSB) (mm), seedlings height (SH) (cm) and variable fluorescence/maximum fluorescence (F_v/F_m).

indicating a good model fit for explaining DQI variations. Of the remaining explanatory traits, only F_v/F_m had a negative direct effect (-0.09) greater than the phenotypic correlation coefficient (-0.06) (Fig. 4B). SH, F_v/F_m , and MTE had direct negative effects (-0.29, -0.09, and -0.02, respectively) on DQI. LA (0.69) and DSB (0.33) (Fig. 4) had the most substantial direct influence on DQI, with LA standing out due to its high magnitude exceeding the residual effect (EVR = 0.44).

Microsatellite marker analysis

We initially tested a total of 13 microsatellite markers, only eight yielded consistent and reliable amplification across all individuals analyzed. In this eight evaluated loci, a total of 118 alleles were identified. The number of alleles per locus ranged from 8 (EE43) to 20 (EE05), with an average number of alleles per locus (N_m) of 14.75. All evaluated loci exhibited high polymorphic information content ($PIC \geq 0.5$). Observed heterozygosity (H_o) values ranged from 0.37 (EE09) to 0.98 (EE52), while expected heterozygosity (H_e) varied between 0.73 (EE43) and 0.93 (EE05). Inbreeding coefficient (F) estimates ranged from -0.13 (EE45) to 0.52 (EE09). Most loci showed negative F values, except for EE09 (0.52) and EE43 (0.04), which had positive values (Table 2). All other estimates should be considered null since their values were negative. As additional information for population diversity characterization, the mean PIC values were estimated, ranging from 0.68 (EE43) to 0.92 (EE05 and EE47), with an average of 0.84 (Table 2).

Diversity analysis

Genetic and molecular traits were used to assess genetic divergence. Genetic data were derived from BLUP estimates obtained via REML/BLUP. The greatest observed genetic distance (3.07) was between RNS09 and VNI14, while the smallest (0.02) was between IBI13 and RNS04, with a mean distance of 1.15. Seven groups were identified (Fig. 6A).

Microsatellite analysis showed genetic differences among genotypes from both geographically proximate (GUA.06 and AL.02) and distant populations (GUA.08 and PA.01). The smallest genetic distance (0.4285) was observed between RNS647 and GUA01. Nine groups were identified (Fig. 6B), with genetic distances ranging from 0.42 to 1.00 and a mean of 0.78.

A Mantel test revealed no correlation between the genetic and molecular distance matrices ($r = 0.098$, p -value = 0.022). The combined analysis resulted in a mean genetic distance of 0.49, with a minimum of 0.10 (VNI11 and VNI13) and a maximum of 0.89 (IBI06 and VNI02). Ten distinct groups were identified based on joint analysis (Fig. 7).

Discussion

Genetic analysis

The results of this study indicate genetic variability among genotypes for all evaluated traits, with biometric (EDF, WF, WS) and emergence traits (ESI, MTE, E) showing strong genetic control, as demonstrated by their high

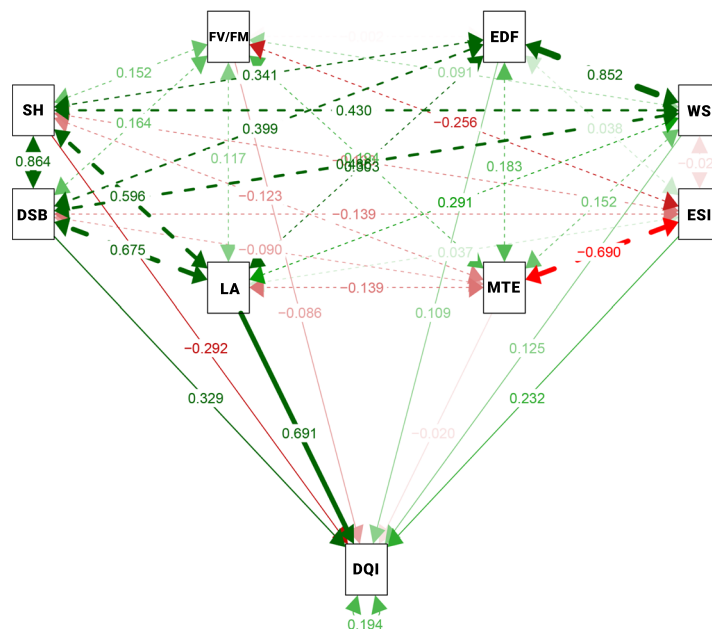


Fig. 5. Path analysis between the characteristics equatorial diameter of the fruits (EDF) (mm), eight of 25 seeds (WS), emergence speed index (ESI), mean emergence time (MTE), Dickson's quality index (DQI), total leaf area (LA) (cm^2), diameter of seedlings in the basis (DSB) (mm), seedlings height (SH) (cm) and variable fluorescence/maximal fluorescence (F_v/F_m). Coefficient of determination (R^2) = 80.57%; residual variable effect (EVR) = 0.4408.

Locus	N	N _A	H _o	H _e	F	PIC
EE05	58	20	0.95	0.93	−0.02	0.92
EE09	49	9	0.37	0.77	0.52	0.74
EE23	55	14	0.93	0.87	−0.07	0.85
EE43	53	8	0.70	0.73	0.04	0.68
EE45	59	14	0.93	0.83	−0.13	0.81
EE47	58	19	0.95	0.92	−0.03	0.92
EE48	58	17	0.95	0.88	−0.07	0.87
EE52	53	17	0.98	0.92	−0.07	0.91

Table 2. Genetic diversity indices derived from the analysis of eight microsatellite loci in founder genotypes sampled in the state of Espírito Santo, Brazil, for the establishment of the breeding foundation population and conservation of *E. edulis*. N: sample size; N_A: number of observed alleles; H_o: observed heterozygosity; H_e: expected heterozygosity; F: inbreeding coefficient.

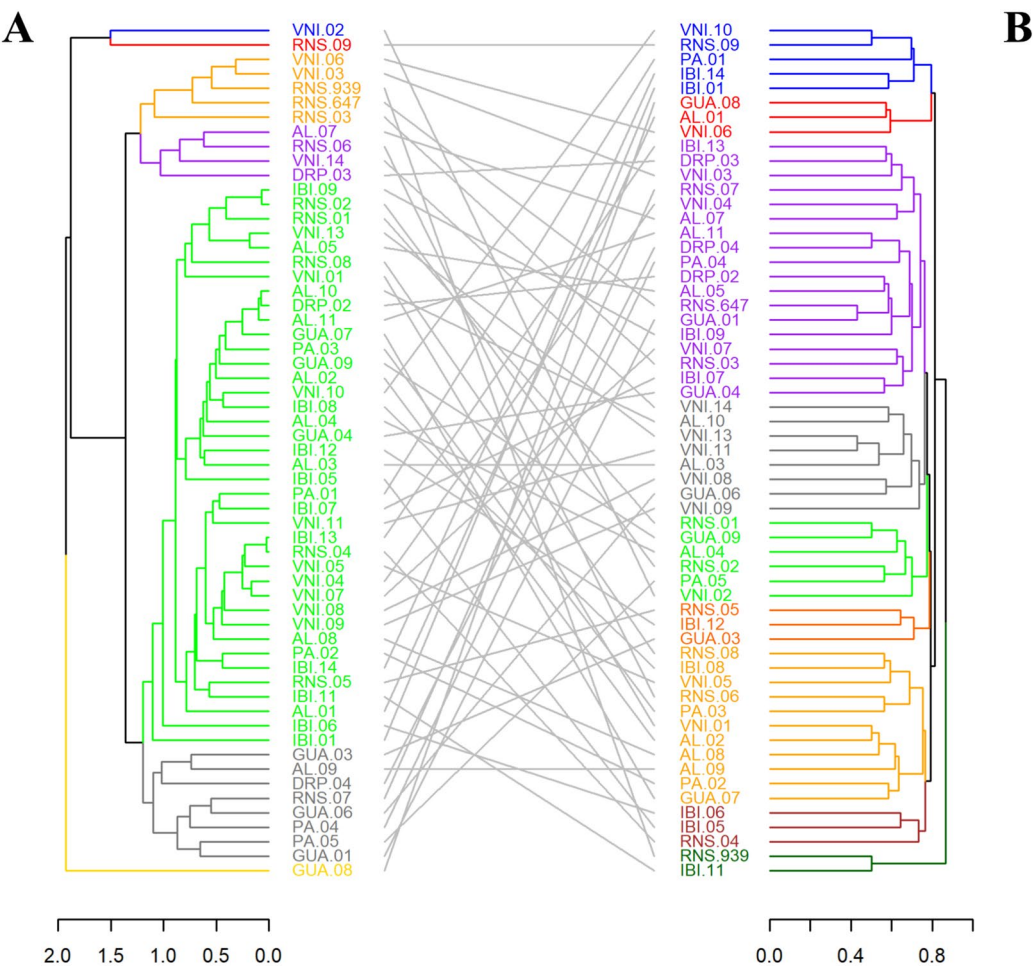


Fig. 6. Dendrograms generated based on dissimilarity values, by the predicted BLUPs for each founders genotypes (genetic value) (A) and by molecular markers (B), obtained by the Standardized Mean Euclidean Distance (DEMP) and by the unweighted index (INP), respectively. Both clusters are generated by the method of average linkage between groups (UPGMA) and cut-off point with k=1.25. The different colors represent each considered group. The initial acronyms refer to the municipalities where the founders genotypes was sampled, being: Alegre (AL), Domingos Martins (PA), Dorés do Rio Preto (DRP), Guaçuí (GUA), Ibitirama (IBI), Rio Novo do Sul (RNS) and Venda Nova do Imigrante (VNI) Groups formed in A: Group I (blue), Group II (red), Group III (orange), Group IV (purple), Group V (green), Group VI (gray) and Group VII (yellow). Groups formed in B: Group I (blue), Group II (red), Group III (purple), Group IV (grey), Group V (green), Group VI (orange), Group VII (yellow), Group VIII (brown) and Group IX (dark green).

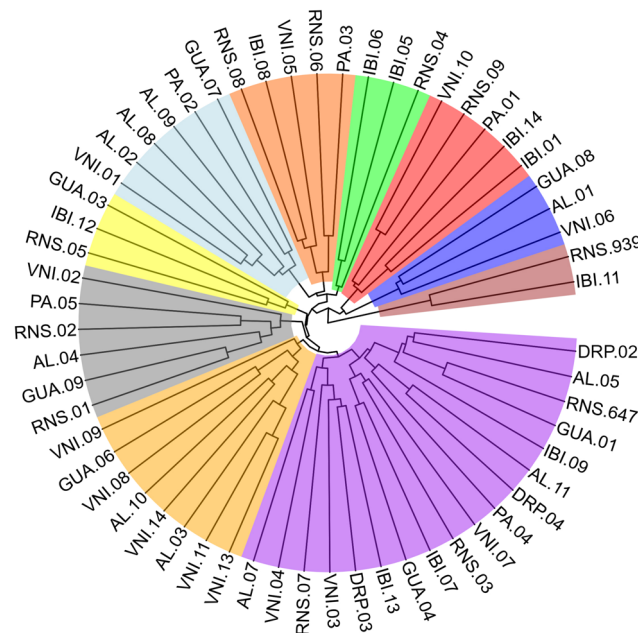


Fig. 7. Dendrograms generated based on the average dissimilarity values of the individual genotypic and molecular distance matrices, obtained by the Standardized Mean Euclidean Distance method (DEMP) and by the unweighted index (INP), respectively. The clustering was generated by the method of mean linkage between groups (UPGMA) and the cut-off point determined according to Mojema⁴⁴ with $k=1.25$. The different colors represent each considered group. The initial acronyms refer to the municipalities where the matrix was sampled, being: Alegre (AL), Domingos Martins (PA), Doros do Rio Preto (DRP), Guaçuí (GUA), Ibitirama (IBI), Rio Novo do Sul (RNS) and Venda Nova do Imigrante (VNI). Groups formed: Group I (purple), Group II (orange), Group III (gray), Group IV (yellow), Group V (light blue), Group VI (dark orange), Group VII (green), Group VIII (red), Group IX (dark blue) and Group X (brown).

heritability values (Table 1). This is particularly relevant for breeding programs, as high heritability facilitates the selection process and enhances the potential for genetic gain across generations. In addition, traits such as fruit and seed size (WF and WS) and emergence performance (e.g., ESI) not only exhibited high heritability but also showed positive genetic correlations with seedling quality (e.g., DQI). This suggests that selecting genotypes with desirable fruit and seed traits may indirectly enhance seedling vigor - a key objective in both conservation and domestication efforts for *E. edulis*. On the other hand, traits related to seedling growth appear to be more influenced by environmental factors, emphasizing the need for robust experimental designs and evaluations under diverse conditions to increase selection accuracy⁴⁷. In this context, the observed results are crucial for guiding the development of a genetic improvement program for *E. edulis*. They enhance the understanding of the genetic control of the evaluated traits and their associations, enabling breeders to design more effective strategies across experimental, analytical, and selection processes, ultimately aiming to maximize genetic gains.

Morphophysiological growth traits-such as seedling height (SH), leaf area (LA), shoot and total dry mass (SDM, TDM), and chlorophyll fluorescence (F_v/F_m) - generally exhibited lower heritability estimates in our study due to a greater influence of environmental variation and genotype $\times$ environment interactions over time. These traits are measured during the seedling developmental phase and are highly plastic, being affected by subtle variations in microenvironmental conditions (e.g., light, temperature, humidity, soil heterogeneity); additionally, these traits often involve complex physiological processes that are controlled by multiple genes with small effects, making the genetic contribution more diffuse. As a result, the residual variance (σ_e^2) tends to be higher relative to the genetic variance (σ_g^2), leading to lower heritability estimates (H^2). In contrast, fruit and seed traits are typically less plastic and more developmentally stable, especially when measured at a single time point, which contributes to higher heritability values. These results for fruit-related traits are consistent with previous studies^{11,48,49}, which have also reported elevated values in multi-year evaluations⁵⁰. These findings reinforce the strong genetic control of these morpho-agronomics traits, which holds considerable relevance for the fruit processing industry. In summary, the lower heritability estimates for morphophysiological traits reflect their greater environmental sensitive.

Given their relevance as variable, informative, and robust markers⁵¹, capable of detecting high levels of polymorphism⁵² and effectively identifying genetic divergence among genotypes^{53,54}, we deliberately employed Simple Sequence Repeat (SSR) molecular markers in this study. This choice aimed to assess the molecular genetic diversity within the sampled population for the construction of its breeding foundation population. This strategic approach was adopted to validate the efficiency of the sampling process, providing a comprehensive analysis of genetic diversity, as well as allele identification and distribution. These findings are crucial to ensuring the success of the *E. edulis* conservation and breeding program, offering fundamental insights to enhance the

efficiency and sustainability of the initiative. Although the number of markers used was limited (eight), they were sufficiently informative and successfully discriminated against genotypes into distinct genetic groups. These markers used here have been validated in previous studies. For example, Pereira et al.⁵⁵ employed nine microsatellite markers, including seven of those used in our study, to investigate *E. edulis* populations from different regions of Brazil. Their findings corroborate our results, reinforcing the patterns of genetic structure and diversity we report here. Additionally, previous analyses comparing high-throughput SNP markers and eight SSR markers have confirmed similar patterns of genetic differentiation and diversity⁵⁴, supporting the conclusions derived from our microsatellite data.

The results presented in Table 2 indicate that all analyzed loci can be classified as highly polymorphic and effective for analysis since the observed heterozygosity (H_O) values were above 0.70. The term polymorphism refers to the presence of multiple variations of a specific gene⁵⁶. In genetic diversity studies, highly polymorphic loci are crucial as they reflect strong genetic variation that can be exploited for precise differentiation of genetic materials^{57–59}.

Similar to H_O , the polymorphic information content (PIC) is an informativity estimate used to infer the degree of polymorphism of a given marker⁵⁸. PIC estimates are classified as poorly informative when below 0.25, moderately informative between 0.25 and 0.50, and highly informative when exceeding 0.50²⁹. Accordingly, all markers used in this study can be classified as highly informative, as the lowest estimated PIC value was 0.68 for locus EE43.

The observed F values for most loci were negative, indicating heterozygote excess and confirming allele exchange among genotypes within populations. Positive F values were found for loci EE09 and EE43, suggesting an excess of homozygotes. Since *E. edulis* is predominantly allogamous, an acceptable self-fertilization rate (5%) may explain the increased homozygosity at these loci.

The results of this study support the effectiveness of sampling broad genetic variability to establish a breeding fundation population or a germplasm collection for species conservation. This effectiveness is particularly evident when compared to previous studies, which, although indicating high diversity, presented results similar to or lower than those observed in this study Carvalho et al.⁶⁰ (H_e : 0.78–0.86); Coelho et al.⁶¹ (N: 250; H_e : 0.61–0.75; F : 0.05–0.36); Conte et al.⁶² (N: 600; H_e : 0.781–0.785); Carvalho et al.⁶³ (N: 463; H_e : 0.72–0.86; F : 0.10–0.29); Gaiotto et al.⁶⁴ (N: 583; H_e : 0.73–0.79; F : low or null); Novello et al.⁶⁵ (N: 361; H_e : 0.76–0.83; F : 0.10–0.34); Pereira et al.⁵⁵ (N: 527; H_e : 0.54–0.81; F : 0.03–0.28).

Genetic diversity is a key factor in the construction of germplasm banks and the development of founder populations, not only for species conservation but also for providing a diverse set of characteristics that can be explored in breeding programs. The markers used identified a total of 118 alleles in 59 genotypes. This extensive allelic sampling becomes even more evident when compared to other studies, such as the genetic diversity characterization of *E. edulis* conducted by Pereira et al.⁵⁵. In their study, the authors used nine SSR markers to evaluate 527 individuals from 26 locations along the species' natural distribution in Brazil, identifying a total of 178 alleles.

Moreover, this study effectively demonstrated the high genetic variability of *E. edulis* in Espírito Santo, surpassing previous studies with lower diversity estimates. Specifically, Carvalho et al.⁶⁶ assessed 160 genotypes in four populations using seven SSR markers, reporting H_O : 0.36–0.68, H_e : 0.60–0.79, N_A : 43, and F : 0.14–0.34. Additionally, Pereira et al.⁵⁵ reported values for the same region: H_O : 0.40–0.60, H_e : 0.80, and F : 0.10–0.30. Thus, this study significantly contributes to the understanding and conservation of the species' genetic diversity in the region.

Previous studies indicate that most of the genetic variation in *E. edulis* is concentrated within populations^{55,62,66}. While confirming this diversity pattern, other researchers emphasize that population differentiation, although low, is statistically significant^{64,65}. Interpreting these findings with caution is essential, as misinterpretation could lead to inadequate decision-making, particularly in promoting genetic material collection concentrated in only a few locations. This can be illustrated by drawing a parallel with the study conducted by Carvalho et al.⁶⁶, which, although it provided a good estimate of genetic diversity (with an average of approximately 6 alleles per marker), may underestimate the genetic variation among populations. In contrast, the approach adopted in the present study, based on a broader sampling strategy that prioritizes the inclusion of genotypes from a larger number of populations across diverse geographic locations, proved to be more effective. This strategy resulted in a higher average number of alleles per marker (approximately 15), offering greater efficiency for conservation programs and a broader genetic base to be explored in breeding initiatives. This is particularly advantageous because the number of sampled alleles directly reflects the genetic diversity of a germplasm collection or breeding base population, which is the driving force behind selection and population evolution, for favoring resistance to pests and diseases, enhances adaptation to a wide range of biotic and abiotic stresses, and supports improvements in quality and yield of the cultures⁶⁷.

This study underscores the importance of a broader genetic sampling approach for *E. edulis*, covering a large number of geographically distant populations. This strategy has proven effective in maximizing the species' genetic diversity, aligning with conservation objectives. Accordingly, a sampling approach that maximizes the number of collection sites with a reduced number of genotypes per population (considering the effective population size, forest fragment size, and a minimum distance of 150 meters between genotypes) may be an efficient strategy for constructing a high-quality germplasm collection and breeding foundation population for *E. edulis*.

Path analysis

Only the analysis of correlation coefficients can provide misinterpretations between the character associations⁶⁸, since the estimates have indirect effects of other variables, so that the correlation estimates do not make it possible to verify the causes of direct effects⁶⁹. Thus, it is necessary to carry out path analysis, as it allows the

decomposition of correlation estimates into their direct and indirect effects, allowing the knowledge of the influences of the explanatory characters on the one of greatest interest, complementing the results obtained by the correlations and making it possible to verify the real causes and effects between the characteristics⁶⁹.

With the removal of traits exhibiting VIF values greater than five, the path analysis yielded an R^2 estimate exceeding 0.70 and an EVR of 0.4408, indicating a good model fit for explaining variations in IQD. The presence of at least one trait with a VIF greater than five is sufficient for the associated regression coefficients to be highly influenced by multicollinearity⁴⁴, highlighting the importance of excluding such traits from the analysis. In cases of high correlation but low direct effect, the most effective strategy for achieving satisfactory improvements in the primary variable is the simultaneous selection of traits, with emphasis on those exhibiting significant indirect effects⁴³.

Among the evaluated traits, seedling height (SH) was the most notable due to the magnitude of its alteration and the shift in sign between the associations observed in the phenotypic correlation analysis (0.45) and in the path analysis (-0.29). The magnitude of this change, along with the shift in the direction of the association, underscores the need for caution when selecting seedlings based solely on shoot height (SH). By decomposing the direct effect of SH on DQI through path analysis (Fig. 5), it was observed that the associations identified in Fig. 4 were strongly influenced by the indirect effects of DSB and LA, since plants with higher expression of these traits tend to exhibit greater SH. Therefore, when isolating the direct effect of SH on DQI, path analysis may have revealed potential effects such as seedling etiolation responses, which can lead to lower-quality genotypes, i.e., reduced DQI. Depending on the light intensity, seedlings may exhibit an etiolated growth pattern⁷⁰, which is finely regulated by process controlled by hormones, primarily auxins (IAA) and gibberellins (GAs), which promote cell elongation⁷¹. Excessive stem elongation has been widely recognized since the discovery of the first gibberellin (GA), which was found to cause lodging and yield loss in rice⁷². The genes GA 20-oxidase (GA20ox) and GA 2-oxidase (GA2ox) encode, respectively, enzymes that catalyze the biosynthesis and inactivation of gibberellins, regulated the accumulation of active GAs⁷¹, crucial for normal stem development and seedling quality.

However, in *E. edulis*, a climax species⁷³ that requires shading during its early developmental stages^{74,75}-a condition provided in the present experiment through 50% shade-we hypothesize that reduced light availability may have led to lower levels of the far-red absorbing form of phytochrome (Pfr). This, in turn, could increase tissue sensitivity to gibberellins (GAs) and consequently upregulate the expression of GA20ox, as previously observed in pea (*Pisum sativum* L.)⁷⁶. This mechanism may have contributed to the observed elongation (etiolation) of *E. edulis* seedlings. Nevertheless, under such environmental conditions, apical meristematic cells may be deprived of metabolic sugars that directly regulate the expression of *CYCB1;1/CDKB1;1*, along with a decline in *CYCD3;1* activity, which is essential for the activation of cell division in meristematic regions⁷⁷.

Thus, the isolated use of SH for selecting higher-quality seedlings should be avoided, as its direct effect is inversely proportional to DQI, potentially leading to undesirable outcomes. However, when considering SH for selecting seedlings with higher DQI, leaf area (LA) and dry stem biomass (DSB) should be used simultaneously, as they were the primary traits explaining variations in DQI, and presents high indirect effect level on SH. In a study on *Syagrus romanzoffiana* (Cham.), a positive correlation between DSB and DQI (0.75) was also observed, confirming the effectiveness of DSB assessment in indicating seedling quality⁷⁸. These findings indicate that, for identifying seedling quality in visual assessments or future analyses, DSB and LA stand out as the most promising traits due to their ease of measurement and strong direct effect on DQI. Therefore, Pearson correlation was valuable for identifying associations among fruit, seed, and seedling traits, suggesting that larger fruits and seeds are generally linked to more vigorous seedlings. However, to clarify whether these associations were direct or indirect by other traits, we conducted path analysis using DQI as the response variable. This approach revealed that the influence of fruit and seed traits on seedling quality is mostly indirect, acting through traits such as leaf area and stem diameter.

Diversity analysis

The observed genotypic divergence values exhibited substantial variation, which may be related to the degree of divergence between the genotypes, with RNS.09 and VNI.02 being the most divergent. In this context, genotypes with high genotypic divergence can be considered better options for parental selection in breeding programs. This is because genotypes that exhibit greater dissimilarity can produce offspring with increased genetic variability and a higher heterotic effect⁷⁹, leading to a phenotypic response in productivity above the population average and, consequently, greater selective gains. A comprehensive understanding of the degree of divergence among accessions requires genotype analyses based on morpho-agronomic and molecular data⁸⁰. This allows conclusions about the genetic divergence among the analyzed genotypes and supports decision-making regarding controlled crossing strategies, field organization, and conservation population management.

The Mantel test, performed between the distance matrices calculated from genotypic and molecular data, revealed an absence of correlation between them. A similar pattern was observed in the characterization of *E. oleracea* accessions based on morpho-agronomic and molecular traits⁸¹. According to the authors, the lack of correlation is due to the fact that morpho-agronomic traits are controlled by multiple genes and are influenced by environmental factors, unlike genetic markers, which are associated with and distributed throughout the genome. In the present study, data correction and the use of genetic values for diversity analysis may explain the observed differences, as the high number of genes controlling quantitative traits enhances the capture of diverse genetic effects, leading to results distinct from those obtained with a limited number of microsatellite markers.

The different classifications and lack of association between the distance matrices highlight the necessity of integrating morphometric and molecular data to enhance the discriminatory power of genotypes, thereby generating more accurate differentiation results. To achieve this, dissimilarity indices can be used in combination

with genotype dissimilarity analyses⁸², as applied in the present study and illustrated in Fig. 6, with the aim of improving the efficiency of genotype characterization and differentiation⁸².

Regarding cluster analysis based on both data types, no grouping by geographic distribution was observed. Genotypes sampled from distant locations appeared within the same branches (Fig. 6). Marçal et al.⁸³ reported similar results when evaluating the genetic diversity of *E. edulis* in forest fragments in Espírito Santo. According to the authors, the random distribution of origins within groups suggests the presence of genetic diversity among the accessions. The mixture observed in this study confirms the efficiency of the sampling strategy in capturing genetic diversity within and between municipalities, indicating the sampled matrices' capacity to support the conservation of the species' genetic resources in an *ex-situ* germplasm bank.

Although the groups were not defined by geographic proximity, the purple and orange clusters (Fig. 6B) were the largest and stood out due to their constituent genotypes. In the purple cluster, genotypes originated from different regions, whereas the orange cluster predominantly grouped genotypes from Venda Nova do Imigrante. This finding may serve as preliminary evidence of limited genetic variability in the forest fragments of this region, leading to the hypothesis of a high inbreeding coefficient or a founder effect resulting from a small sample size. The differentiation of this group from others underscores the distinctiveness of this locality, emphasizing the importance of sampling VNI genotypes to maximize genetic diversity in the breeding founders populations and germplasm collection.

The genetic and molecular diversity analyses revealed substantial divergence among genotypes from different collection sites, supporting the recommendation to prioritize sampling across multiple geographic regions. By combining genotypes from distinct genetic groups in germplasm collections, aligning their spatial arrangement in the field based on their genetic distance estimates, and placing genetically more distant genotypes in closer proximity, it is possible to enhance the potential for genetic recombination and broaden the genetic base of breeding populations. This approach not only fosters the development of improved genotypes but also contributes to the long-term conservation of genetic resources in *E. edulis*. The intrapopulational diversity of *E. edulis* has been documented in studies of natural populations, where it has been linked to factors such as reproductive system function, pollen and seed dispersal patterns, forest fragmentation and conservation status, and geographic origin⁵⁵. Genetic diversity across different geographic regions implies varying adaptive capacities among genotypes, which is crucial for breeding programs⁸³ and the conservation of *E. edulis*.

The results obtained in this study can serve as a foundation for breeding and conservation programs. By selecting more divergent genotypes for crossing, it is possible to maximize the genetic diversity of future progenies while minimizing inbreeding depression. Consequently, these breeding foundation population will play a crucial role in supporting *E. edulis* conservation, preservation and breeding programs by providing seeds with high genetic variability to aid in the enrichment of forest fragments. Moreover, they will serve as genetic reservoirs that can be leveraged for breeding programs. The results of this study provide valuable insights into genetic control and trait associations, while also offering relevant information on the diversity of the sampled population. These findings are crucial for supporting the structuring and optimization of breeding and conservation programs for the species.

Conclusions

Emergence and morphometric traits of fruits and seeds exhibited high H^2 , whereas seedling traits had relatively smaller H^2 estimates. The significant positive r_g between the morphometric characteristics of fruits and seeds and those related to growth indicates the potential for using these descriptors for early indirect selection to develop higher-quality seedlings. For seedling selection targeting quality (DQI), indirect selection can be primarily based on LA and DBS. Path analysis confirmed that LA and DSB exert the strongest direct effects on DQI, reinforcing their role as reliable early indicators of seedling quality. The detection of significant genetic variability for all traits, especially fruit and seed biometric characteristics, highlights the existence of broad genetic diversity among the *E. edulis* genotypes evaluated. Moreover, the low significance of genotype $\times$ time and genotype $\times$ environment interactions suggests phenotypic stability under the conditions tested, which is advantageous for selection.

The high magnitudes of the genetic diversity index estimates from the molecular markers may have been a direct response to the sampling method employed. This confirms the effectiveness of the sampling strategy used in this study in maximizing the capture of genetic variability for the establishment of a robust breeding founding population for genetic breeding programs and construction of germplasm in genetic conservation programs. Additionally, the high polymorphism, heterozygosity, and PIC values across loci reinforce the usefulness of these markers for characterizing genetic resources in *E. edulis*. The absence of correlation between genetic and molecular distance matrices (Mantel test) suggests that neutral markers alone may not fully capture functional genetic variation relevant to phenotypic performance, underscoring the importance of integrative approaches. The diversity analysis, using molecular and genotypic information, provided insights-via the dendrogram-on how to guide the establishment of breeding orchards of an *ex-situ* to maximize crosses between the most divergent genotypes. The grouping patterns identified offer a practical framework for crossing strategies aimed at maximizing heterosis and genetic gain.

Data availability

The phenotypic data generated and analyzed during the current study are available from the corresponding author upon reasonable request. The microsatellite marker datasets are publicly available at: <https://github.com/GuilhermeBravim/-Divergence-Patterns-in-Euterpe-edulis-for-Breeding-and-Conservation-Applications>.

Received: 19 February 2025; Accepted: 14 May 2025

References

- Marques, M. et al. The atlantic forest: History, biodiversity, threats and opportunities of the mega-diverse forest. *Biotropica* **53**, 279–287 (2021).
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- Galetti, M. et al. Functional extinction of birds drives rapid evolutionary changes in seed size. *Science* **340**, 1086–1090 (2013).
- Tres, A., Tetto, A., de Freitas Milani, J. et al. Reproductive phenology of *Euterpe edulis* Mart. in two altitudinal classes in the Brazilian atlantic forest. *Rev Ibero-Am Cienc Ambient.* **11**, 23–35 (2020).
- Salomão, A. & Santos, I. *Criopreservação de germoplasma de espécies frutíferas nativas* (2018).
- Guimarães, L. d. O. & de Souza, R. *Palmeira juçara: patrimônio natural da Mata Atlântica no Espírito Santo* (2017).
- Martinelli, G. & Moraes, M. *Livro vermelho da flora do Brasil* (Andrea Jakobsson: Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, 2013).
- Joly, C., Metzger, J. & Tabarelli, M. Experiences from the Brazilian atlantic forest: ecological findings and conservation initiatives. *New Phytol.* **204**, 459–473 (2014).
- Maciel, L. d. O., Moura, N. d. & Leonardi, A. Cadeia produtiva do açaí juçara na região do litoral norte do Rio Grande do Sul. *Rev Teor Evid Econ.* **25**, 29–53 (2019).
- Tedesco, G., Campos, F., de Azevedo, B. & Mathias, R. *Análise da cadeia produtiva do açaí catarinense com ênfase nos diferentes atores e atividades produtivas* (Editora, 2021).
- Canal, G. B. et al. Selection of accessions of *Euterpe edulis* Mart. based on fruit and pulp characteristics. *Tree Genetics & Genomes* **21**, 1–11 (2025).
- Borges, G. D. S. C. et al. Chemical characterization, bioactive compounds, and antioxidant capacity of jussara (*Euterpe edulis*) fruit from the atlantic forest in southern Brazil. *Food Res. Int.* **44**, 2128–2133 (2011).
- Silva, P. P. M. d. et al. Physical, chemical, and lipid composition of juçara (*Euterpe edulis* Mart.) pulp. *Brazilian Journal of Food and Nutrition* **24**, 7–13 (2013).
- Schulz, M. et al. Bioaccessibility of bioactive compounds and antioxidant potential of juçara fruits (*Euterpe edulis* Martius) subjected to in vitro gastrointestinal digestion. *Food Chem.* **228**, 447–454 (2017).
- Dourado, C. et al. Selection strategies for growth characters and rubber yield in two populations of rubber trees in Brazil. *Ind. Crops. Prod.* **118**, 118–124 (2018).
- Rodrigues, H., Cruz, C., Macêdo, J. d. & et al. Genetic variability and progeny selection of peach palm via mixed models (REML/BLUP). *Acta Scientiarum Agronomy* **39**, 165–173 (2017).
- Reflora - Herbário Virtual: *Euterpe edulis* from Domingos Martins (VIES039447). <https://reflora.jbrj.gov.br/reflora/herbarioVirtual/ConsultaPublicoHVUC/ConsultaPublicoHVUC.do?idTestemunho=4569793> (2025). Accessed on April 17, 2025.
- Reflora - Herbário Virtual: *Euterpe edulis* from Rio Novo do Sul (VIES038826). <https://reflora.jbrj.gov.br/reflora/herbarioVirtual/ConsultaPublicoHVUC/ConsultaPublicoHVUC.do?idTestemunho=4950057> (2025). Accessed on April 17, 2025.
- Reflora - Herbário Virtual: *Euterpe edulis* from Venda Nova do Imigrante (MBM00422915). <https://reflora.jbrj.gov.br/reflora/herbarioVirtual/ConsultaPublicoHVUC/ConsultaPublicoHVUC.do?idTestemunho=5882160> (2025). Accessed on April 17, 2025.
- Cursi, P. R. & Cicero, S. M. Fruit processing and the physiological quality of *Euterpe edulis* Martius seeds. *J. Seed Sci.* **36**, 134–142 (2014).
- Dickson, A., Leaf, A. L. & Hosner, J. F. Quality appraisal of white spruce and white pine seedling stock in nurseries. *The For. Chron.* **36**, 10–13 (1960).
- Maguire, J. Speed of germination-aid in selection and evaluation for seedling emergence and vigor. *Crop. Sci.* **2**, 176–177 (1962).
- Labouriau, L. & Viladares, M. On the germination of seeds of *Calotropis procera* (Ait.). *Anais da Sociedade Botânica do Brasil* (1976).
- Dickson, A., Leaf, A. & Hosner, J. Quality appraisal of white spruce and white pine seedling stock in nurseries. *For. Chron.* **36**, 10–13 (1960).
- Carvalho, M., Noia, L., Ferreira, M. D. S. & Ferreira, A. DNA de alta qualidade isolado a partir do córtex de *Euterpe edulis* Mart. (Arecaceae). *Cienc Florest* **29**, 396–402 (2019).
- Gaiotto, F., Brondani, R. V. & Grattapaglia, D. Microsatellite markers for heart of palm—*Euterpe edulis* and *E. oleracea* Mart. (Arecaceae). *Mol. Ecol. Notes* **1**, 86–88 (2001).
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2023).
- Liu, B. H. *Statistical genomics: linkage, mapping, and QTL analysis* (CRC Press, 2017).
- Botstein, D., White, R., Skolnick, M. & Davis, R. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314 (1980).
- Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).
- Henderson, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423–447 (1975).
- Verbyla, A., Faveri, J., Deery, D. & Rebetzke, G. Modelling temporal genetic and spatio-temporal residual effects for high-throughput phenotyping data. *Aust. & New Zealand J. Stat.* **63**, 284–308 (2021).
- Verbyla, A. P. A note on model selection using information criteria for general linear models estimated using reml. *Aust. N. Z. J. Stat.* **61**, 39–50 (2019).
- Faveri, J. et al. Statistical methods for analysis of multi-harvest data from perennial pasture variety selection trials. *Crop. Pasture Sci.* **66**, 947–962 (2015).
- Souza, V. d., Ribeiro, P. d. O., Vieira Junior, I. et al. Exploring genotype × environment interaction in sweet sorghum under tropical environments. *Agron. J.* **113**, 3005–3018 (2021).
- Chaves, S. et al. Application of linear mixed models for multiple harvest/site trial analyses in perennial plant breeding. *Tree Genet. & Genomes* **18**, 44 (2022).
- Araújo, M. S., Chaves, S. F. S., Damasceno-Silva, K. J., Dias, L. A. S. & Rocha, M. M. Modeling covariance structures for genetic and non-genetic effects in cowpea multi-environment trials. *Agron. J.* **115**, 1248–1256 (2023).
- Akaike, H. A new look at the statistical model identification. *IEEE Trans Automat Contr.* **19**, 716–723 (1974).
- Schwarz, G. Estimating the dimension of a model. *The Annals Stat.* **6**, 461–464 (1978).
- Cullis, B. R., Smith, A. B. & Coombes, N. E. On the design of early generation variety trials with correlated data. *J. Agric. Biol. Environ. Stat.* **11**, 381–393 (2006).
- Resende, M., Silva, F. & Azevedo, C. Estatística matemática, biométrica e computacional. *Suprema, Visconde do Rio Branco*, 881p (2014).
- Montgomery, D. C., Peck, E. A. & Vining, G. G. *Introduction to linear regression analysis* (John Wiley & Sons, 2021).
- Cruz, C., Ferreira, F. & Personi, L. Biometria aplicada ao estudo da diversidade genética. *Visconde do Rio Branco Suprema* **620** (2011).
- Mojena, R. Hierarchical grouping methods and stopping rules: An evaluation. *The Comput. J.* **20**, 359–363 (1977).

45. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
46. Team, R. C. R language definition. *Vienna, Austria: R foundation for statistical computing* **3**, 116 (2024).
47. Rasheed, A. et al. Study of genetic variability, heritability, and genetic advance for yield-related traits in tomato (*Solanum lycopersicon* Mill.). *Front. Genet.* **13**, 1030309 (2023).
48. Canal, G. B. et al. Single and multi-trait genomic prediction for agronomic traits in *Euterpe edulis*. *PLoS One* **18**, e0275407 (2023).
49. Marçal, T. et al. Repeatability of biometric characteristics of juçara palm fruit. *Biosci. J.* **53**, 890–898 (2016).
50. Canal, G. B. et al. Genomic studies of the additive and dominant genetic control on production traits of *Euterpe edulis* fruits. *Sci. Rep.* **13**, 9795 (2023).
51. Sagar, T., Kapoor, N. & Mahajan, R. Microsatellites as potential molecular markers for genetic diversity analysis in plants. In *Molecular Marker Techniques: A Potential Approach of Crop Improvement*, 81–101 (Springer, 2023).
52. Padmakar, B., Sailaja, D. & Aswath, C. Molecular exploration of guava (*Psidium guajava* L.) genome using ssr and rapd markers: a step towards establishing linkage map. *J. Hort. Sci.* **10**, 130–135 (2015).
53. Senan, S., Kizhakayil, D., Sasikumar, B. & Sheeja, T. E. Methods for development of microsatellite markers: an overview. *Notulae Sci. Biol.* **6**, 1–13 (2014).
54. Almeida, F. A. N. et al. Genetic diversity analysis of *Euterpe edulis* based on different molecular markers. *Tree Genet. & Genomes* **20**, 31 (2024).
55. Pereira, A. G. et al. Patterns of genetic diversity and structure of a threatened palm species (*Euterpe edulis* Arecaceae) from the Brazilian Atlantic forest. *Heredity* **129**, 161–168 (2022).
56. Bahar, A. & Esra, I. Molecular marker technologies in food plant genetic diversity studies: An overview. *Foods Raw Mater* **11**, 282–292 (2023).
57. Maske, J. M., Keshavrao, Z. R. & Amarsing, R. C. Analysis of genetic diversity of commercial tomato varieties using molecular marker viz. rapd. *Int. J. Curr. Microbiol. Appl. Sci.* **7**, 3559–3565 (2018).
58. Baldaniya, V. G., Narwade, A. V., Patel, P. B. & Karmakar, N. Molecular diversity analysis of rice (*Oryza sativa* L.) genotypes using rapid and ssr marker. *Emergent Life Sci. Res.* **8**, 113–123 (2022).
59. György, Z., Alam, S., Priyanka, P. & Zámbořině Németh, E. Genetic diversity and relationships of opium poppy accessions based on ssr markers. *Agriculture* **12**, 1343 (2022).
60. Carvalho, C. S. et al. Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic forest of Brazil. *Conserv. Genet.* **18**, 467–478 (2017).
61. Coelho, G. M. et al. Genetic structure among morphotypes of the endangered Brazilian palm *Euterpe edulis* Mart (Arecaceae). *Ecol. Evolution* **10**, 6039–6048 (2020).
62. Conte, R. & Reis, R., MS acknd Vencovsky. Effects of management on the genetic structure of *Euterpe edulis* Mart. populations based on microsatellites. *Sci. For.* **72**, 81–88 (2006).
63. Silva, Carvalho C., Ribeiro, M., Côrtes, M., Galetti, M. & Collevatti, R. Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. *Heredity* **115**, 216–224 (2015).
64. Gaiotto, F. A., Grattapaglia, D. & Vencovsky, R. Genetic structure, mating system, and long-distance gene flow in heart of palm (*Euterpe edulis* Mart.). *J. Hered.* **94**, 399–406 (2003).
65. Novello, M. et al. Genetic conservation of a threatened neotropical palm through community-management of fruits in agroforests and second-growth forests. *For Ecol Manage.* **407**, 200–209 (2018).
66. Carvalho, M. S. et al. Genetic diversity and population structure of *Euterpe edulis* by reml/blup analysis of fruit morphology and microsatellite markers. *Crop Breed Appl. Biotechnol.* **20**, 1–9 (2020).
67. Salgotra, R. K. & Chauhan, B. S. Genetic diversity, conservation, and utilization of plant genetic resources. *Genes* **14**, 174 (2023).
68. Khan, A. M. R., Eyasmin, R., Rashid, M. H., Ishtiaque, S. & Chaki, A. K. Variability, heritability, character association, path analysis and morphological diversity in snake gourd. *Agric. Natural Resources* **50**, 483–489 (2016).
69. Kang, M. S., Miller, J. D. & Tai, P. Y. P. Genetic and phenotypic path analyses and heritability in sugarcane. *Crop. Sci.* **23**, 643–647 (1983).
70. Razzaq, K. & Du, J. Phytohormonal regulation of plant development in response to fluctuating light conditions. *Journal of Plant Growth Regulation* 1–34 (2024).
71. Wang, B. et al. Comparative transcriptome analyses provide novel insights into etiolated shoot development of walnut (*Juglans regia* L.). *Planta* **252**, 1–18 (2020).
72. Peng, J. et al. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* **400**, 256–261 (1999).
73. Silva, JZd., Lauterjung, M. B. & Reis, MSd. Influence of reproduction and basal area on the increment of *Euterpe edulis*. *Floresta Ambient.* **27**, e20180058 (2020).
74. Santos, M. et al. Low light availability affects leaf gas exchange, growth and survival of *Euterpe edulis* seedlings transplanted into the understory of an anthropic tropical rainforest. *South For.* **74**, 167–174 (2012).
75. Nakazono, E., Costa, M. C. D., Futatsugi, K. & Paulilo, M. T. S. Crescimento inicial de *Euterpe edulis* Mart. em diferentes regimes de luz. *Rev. Bras. Bot.* **24**, 173–179 (2001).
76. Ait-Ali, T. et al. Regulation of gibberellin 20-oxidase and gibberellin 3 β -hydroxylase transcript accumulation during de-etiolation of pea seedlings. *Plant Physiol.* **121**, 783–791 (1999).
77. Rawat, S. S. & Laxmi, A. Sugar signals pedal the cell cycle!. *Front. Plant Sci.* **15**, 1354561 (2024).
78. Souza, A. M. B. et al. Crescimento inicial de mudas de *Syagrus romanzoffiana* em substrato à base de biossólido. *Pesq. Agropec. Trop.* **52**, e70577 (2022).
79. Galate, R. S., Mota, M. C., Gaia, J. M. D. & Costa, M. d. S. Phenotypic distance among assai palm's mother plants (*Euterpe oleracea* Mart.) from Eastern Amazon. *Semina* **35**, 1667–1682 (2014).
80. Vieira, E. A., et al. Association between genetic distances in wheat (*Triticum aestivum* L.) as estimated by aifp and morphological markers. *Genet. Mol. Biol.* **30**, 392–399 (2007).
81. Mota Rios, R., Mochiutti, S., Borges, W. L. & Dias, L. A. V. Morphoagronomic and molecular characterization of *Euterpe oleracea* accessions from eastern Brazilian Amazon. *Acta. Sci. Biol. Sci.* **43**, e58099–e58099 (2021).
82. Vieira, E. A. et al. Caracterização fenotípica e molecular de acessos de mandioca de indústria com potencial de adaptação às condições do cerrado do Brasil Central. *Semina: Ciências Agrárias* **34**, 567–581 (2013).
83. Marçal, T. S., Bernardes, C. O., Oliveira, W. B. S. et al. Genetic diversity of *Euterpe edulis* Martius based on fruit traits. *Biosci. J.* **1549**–1556 (2020).

Acknowledgements

We would like to thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES), São Paulo Research Foundation (FAPESP), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) by granting scholarships to authors. We also would like to thank all property owners for kindly allowing us to sample the plants we thought, especially Pedro Bortolotti Menegardo, his wife Mercedes Sartório Menegardo and family; and Vicente de Paulo Menegardo Bortolotti and his wife Maria de Lourdes Bonadiman Bortolotti and family. Maurício dos Santos Araújo was supported by FAPESP (São Paulo Research Foundation, Grant 2024/01868).

Author contributions

GBC: Investigation, writing - original draft, Writing - review & editing. MZP: Investigation, Writing - original draft. JGS, JTO, MSA, GBS, and FANA: Investigation. RSA: Supervision, Conceptualization, Writing - original draft, Writing - review & editing. AF: Supervision, Conceptualization, and Statistical Analysis. MFSF: Supervision, Conceptualization, Writing - original draft, Writing - review & editing. All authors read and approved the final manuscript.

Funding

This work was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES, Finance Code 001), Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES), São Paulo Research Foundation (FAPESP, Grant 2024/01868), and National Council for Scientific and Technological Development (CNPq).

Declarations

Competing interests

The authors declare no competing interests.

Ethical statement

We confirm that we have complied with all the necessary regulations for this type of research.

Permission to collect

The authors declare that all collections of *E. edulis* were conducted on private properties with prior authorization from the respective landowners. Furthermore, all sampling activities were formally authorized by the Brazilian Ministry of the Environment through the Chico Mendes Institute for Biodiversity Conservation (ICMBio), under the SISBIO system. It is important to emphasize that these activities were carried out exclusively for scientific purposes.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-02606-7>.

Correspondence and requests for materials should be addressed to A.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025