AGRONOMY (AGRONOMIA)

Revista Brasileira de Ciências Agrárias
ISSN (on line) 1981-0997
v.17, n.4, e1592, 2022
Recife, PE, UFRPE. www.agraria.pro.br
DOI: 10.5039/agraria.v17i4a1592 - Protocol 1592
Subm. 06/11/2021 • App. 15/11/2022 • Pub. 22/12/2022

Characterization of the genetic structure of a cupuassu tree population collected in primary forest

Thalita Gomes dos Santos[1]*, Jack Loureiro Pedroza Neto[1], Saulo Fabrício da Silva Chaves[2], Rafael Moysés Alves[3], Ana Beatriz Machado da Silva[1], Andreza Rafaely Martins Jose[4]

[1] Universidade Federal do Ceará, Fortaleza, CE, Brasil. E-mail: thalitasantagro@gmail.com; pedrozaagro@gmail.com; abeatrizmachadoufc@gmail.com
[2] Universidade Federal de Viçosa, Viçosa, MG, Brasil. E-mail: saulo.chaves@ufv.br
[3] Embrapa Amazônia Oriental, Belém, PA, Brasil. E-mail: rafael-moyses.alves@embrapa.br
[4] Universidade Federal de Lavras, Lavras, MG, Brasil. E-mail: andrezarafaelymartins@gmail.com

ABSTRACT: The evaluation of the genetic dissimilarity of access is a fundamental step in a genetic improvement program, allowing the direction of actions and supporting decision making that will define the program success. This study aimed to determine the most informative variables to analyze the diversity of the Active Germplasm Bank (BAG - Cupuassu) - Nova Ipixuna Collection, through the morphological characterization of leaves, flowers and fruits. For this purpose, 20 clonal access from fragments of primary forest in the municipality of Nova Ipixuna, Pará, were characterized for floral, leaf and fruit characters. In all, 34 characters were measured. The access were designed in randomized complete blocks with three useful plants per plot and five replications. The characters had their genetic variances and correlations analyzed. These two procedures were used to discard characters. Dissimilarity was estimated using the generalized Mahalanobis distance, which were used for clustering via UPGMA (Unweighted Pair Group Method using Arithmetic averages). In general, the traits have high heritability and a high coefficient of genetic variation. Correlations were more pronounced between characters belonging to the same group. Such analyzes allowed the discarding of 11 variables, leaving 23 for the calculation of dissimilarity. The grouping determined the formation of six groups, which did not depend on the place of collection. Therefore, the presence of genetic variability among access is verified, accrediting them for use in the breeding program.

Key words: Active Germplasm Bank; clustering; genetic dissimilarity; morphological characterization; *Theobroma grandiflorum*

Caracterização da estrutura genética de uma população de cupuaçuzeiro coletada em floresta primária

RESUMO: A avaliação da dissimilaridade genética de acessos é uma etapa fundamental em um programa de melhoramento genético, permitindo o direcionamento de ações e embasando tomadas de decisões que definirão o sucesso do programa. Este estudo objetivou determinar as variáveis mais informativas para analisar a diversidade do Banco Ativo de Germoplasma (BAG) do Cupuaçuzeiro - Coleção Nova Ipixuna, por meio da caracterização morfológica de folhas, flores e frutos. Para este fim, 20 acessos clonais provenientes de fragmentos de floresta primária no município de Nova Ipixuna, Pará, foram caracterizados quanto a caracteres florais, foliares e de fruto. Ao todo, 34 caracteres foram mensurados. Os acessos foram delineados em blocos completos ao acaso com três plantas úteis por parcela e cinco repetições. Os caracteres tiveram suas variâncias e correlações genéticas analisadas. Estes dois procedimentos foram utilizados para descarte de caracteres. A dissimilaridade foi estimada por meio da distância generalizada de Mahalanobis, as quais foram utilizadas para o agrupamento via UPGMA (Unweighted Pair Group Method using Arithmetic averages). De maneira geral, as características possuem alta herdabilidade e alto coeficiente de variação genético. As correlações foram mais pronunciadas entre caracteres pertencentes à um mesmo grupo. Tais análises permitiram o descarte de 11 variáveis, restando 23 para o cômputo da dissimilaridade. O agrupamento determinou a formação de seis grupos, os quais independeram do local de coleta. Portanto, constata-se a presença de variabilidade genética entre os acessos, credenciando-os para utilização no programa de melhoramento.

Palavras-chave: Banco Ativo de Germoplasma, agrupamento; dissimilaridade genética; caracterização morfológica; *Theobroma grandiflorum*

* Thalita Gomes dos Santos - E-mail: thalitasantagro@gmail.com (Corresponding author)
Associate Editor: Alberto Cargnelutti Filho

Introduction

The cupuassu tree (*Theobroma grandiflorum* (Willd. ex Spreng.) Schum) is an allogamous perennial species native to the Brazilian Amazon, where it is widely cultivated. From the 1970s on, commercial plantings of the crop became popular in Pará soil, a fact associated with the phytosanitary crises in the cultivation of the black pepper (*Piper nigrum* L.) caused by fusarium (*Fusarium solani* f. sp. *piperis* Albuquerque) (Homma, 2014). Since then, the products made from cupuassu pulp and almonds have been gaining not only notoriety in the Amazon region, but also throughout the country.

There is intra- and inter-population genetic diversity in native populations of cupuassu trees (Alves et al., 2007), a characteristic that qualifies the species for improvement. In fact, the use of such diversity by the cupuassu genetic improvement program is essential for the achievement of strong and impactful results, especially regarding fruit production and resistance to the witch's broom (*Moniliophthora perniciosa* (Stahel) Aime & Phillips-Mora) (Souza et al., 2012).

Active Germplasm Bank (BAG) play an important role in this regard, acting as maintainers of natural genetic diversity and variability, making alleles of interest immediately reachable by the improver, optimizing gains and conserving genetic resources that may eventually be lost, mainly due to landscape alteration in the species natural habitats (Migicovsky et al., 2019). The cupuassu has BAG in almost all the states of the Northern region of Brazil, which represent a sample of the natural diversity found in these states (Souza et al., 2012).

The process of characterizing and analyzing the genetic diversity of the germplasm of a species is of fundamental importance in a breeding program. From the evaluations performed, it is possible to identify genotypes holding alleles of agronomic interest, form heterotic groups directing crosses to capitalize on heterosis, and guide germplasm collections or re-colections, prioritizing different geographic origins, aiming at amplifying the representativeness of the BAG (Migicovsky et al., 2019; Adeigbe et al., 2021).

This study aimed to determine the most informative variables to analyze the diversity of the Active Germplasm Bank (BAG – Cupuassu) - Nova Ipixuna Collection, through morphological characterization and to estimate genetic parameters related to leaf, flower and fruit characters.

Materials and Methods

The access of the collection used in this study were collected from native populations (Table 1), located in remnants of primary forest in the municipality of Nova Ipixuna, Pará, Brazil (4° 55' 23" S, 49° 4' 19" W, and 97 m of altitude). The access were cloned by grafting and planted in the BAG cupuassu at the Embrapa Amazônia Oriental Physical Base, in Tomé Açu, Pará, Brazil (2° 35' 32" S, 48° 21' 22" W, and 45 m of altitude). The region climate is humid and mesothermal, corresponding to the Ami type of Köppen classification, with an average temperature of 26 °C, relative air humidity around 85%, and average annual precipitation of 2,300 mm. The soil is of 'Latossolo amarelo' type, with medium texture and low natural fertility.

Table 1. Origin and place of collection of the 20 clonal access of the active BAG cupuassu.

Access	Municipality, State	Latitude	Longitude
1077	Nova Ipixuna, PA	04° 57' 33" S	49° 11' 51" W
1080	Nova Ipixuna, PA	04° 57' 27" S	49° 11' 32" W
1085	Nova Ipixuna, PA	04° 52' 31" S	49° 20' 29" W
1089	Nova Ipixuna, PA	04° 52' 27" S	49° 20' 28" W
1093	Nova Ipixuna, PA	04° 52' 34" S	49° 20' 21" W
1098	Nova Ipixuna, PA	04° 52' 36" S	49° 20' 31" W
1103	Nova Ipixuna, PA	04° 53' 51" S	49° 22' 06" W
1111	Nova Ipixuna, PA	04° 53' 44" S	49° 22' 10" W
1118	Nova Ipixuna, PA	04° 54' 20" S	49° 21' 43" W
1119	Nova Ipixuna, PA	04° 54' 20" S	49° 21' 43" W
1122	Nova Ipixuna, PA	04° 53' 24" S	49° 22' 18" W
1124	Nova Ipixuna, PA	04° 54' 16" S	49° 21' 48" W
1125	Nova Ipixuna, PA	04° 54' 14" S	49° 21' 48" W
1126	Nova Ipixuna, PA	04° 54' 14" S	49° 21' 48" W
1127	Nova Ipixuna, PA	04° 54' 14" S	49° 21' 48" W
1132	Nova Ipixuna, PA	04° 54' 09" S	49° 21' 45" W
1133	Nova Ipixuna, PA	04° 54' 09" S	49° 21' 44" W
1136	Nova Ipixuna, PA	04° 55' 12" S	49° 18' 47" W
1137	Nova Ipixuna, PA	04° 56' 23" S	49° 18' 08" W
1145	Nova Ipixuna, PA	04° 56' 06" S	49° 16' 53" W

In February 2005, 20 clonal access of cupuassu were planted in 5 × 5 m spacing. As temporary shade, banana plants (*Musa* spp.) were installed between the rows of cupuassu trees at the same spacing. At the end of the sixth year after planting, all the banana trees had been eliminated, keeping the cupuassu tree in full sun. The experiment was conducted following good cupuassu cultivation practices (Souza, 2007).

The access were arranged in randomized complete block design, with three useful plants per plot and five repetitions. In all, 34 morphological characters were evaluated (Table 2).

The leaf characters were collected from nine mature, healthy and intact leaves of the clonal access, where the CL, LB, LM, LT, CP, DP, CA, LA, and DN characters were measured using a digital caliper, and the AB, AM, and AT characters were measured using a protractor. The floral characters were evaluated in five striated buds (near anthesis) and five fully open flowers, where all characters were measured with the aid of a digital pachymeter. Finally, the fruit characters were measured in five fruits per plant per harvest, where the CFR character was measured using a tape measure, the DFR and ECA characters were measured using a digital caliper, and the PFR character was measured using a semi-analytical scale. The approximate measurement locations are illustrated in Figure 1.

The data were tested for variance analysis (ANOVA) assumptions by the Kolmogorov-Smirnov test (normality) and Bartlett test (homoscedasticity) at 5% probability. Once the residues were found to be aligned with the assumptions, the variances were analyzed using Snedecor F-test, also at 5% probability.

Table 2. Morphological characters (C) and their acronyms (A), evaluated in 20 clonal access of the BAG cupuassu.

Leaf		Flower		Fruit	
C	A	C	A	C	A
Limbus length (cm)	CL	Knurled button length (mm)	CBE	Fruit length (cm)	CFR
Basal width of limbus (cm)	LB	Knurled button diameter (mm)	DBE	Fruit diameter (cm)	DFR
Median width of limbus (cm)	LM	Flower size (mm)	TF	Peel thickness (mm)	ECA
Apical width of limbus (cm)	LT	Stigma-anther distance (mm)	DEA	Fruit weight (g)	PFR
Petiole length (cm)	CP	Penduncle length (mm)	CPE	Number of seeds	NS
Petiole diameter (cm)	DP	Penduncle diameter (mm)	DPE		
Apex length (cm)	CA	Sepal length (mm)	CLS		
Apex width (cm)	LA	Sepal width (mm)	LLS		
Angulation of basal ribs (°)	AB	Petal length (mm)	CLP		
Angulation of median ribs (°)	AM	Petal width (mm)	LLP		
Angulation of apical ribs (°)	AT	Cogules length (mm)	CC		
Inter-nervous distance (cm)	DN	Cogules width (mm)	LC		
Number of rib pairs	NPN	Staminode length (mm)	CE		
		Stylus length (mm)	CET		
		Ovarian length (mm)	CO		
		Ovarian diameter (mm)	DO		

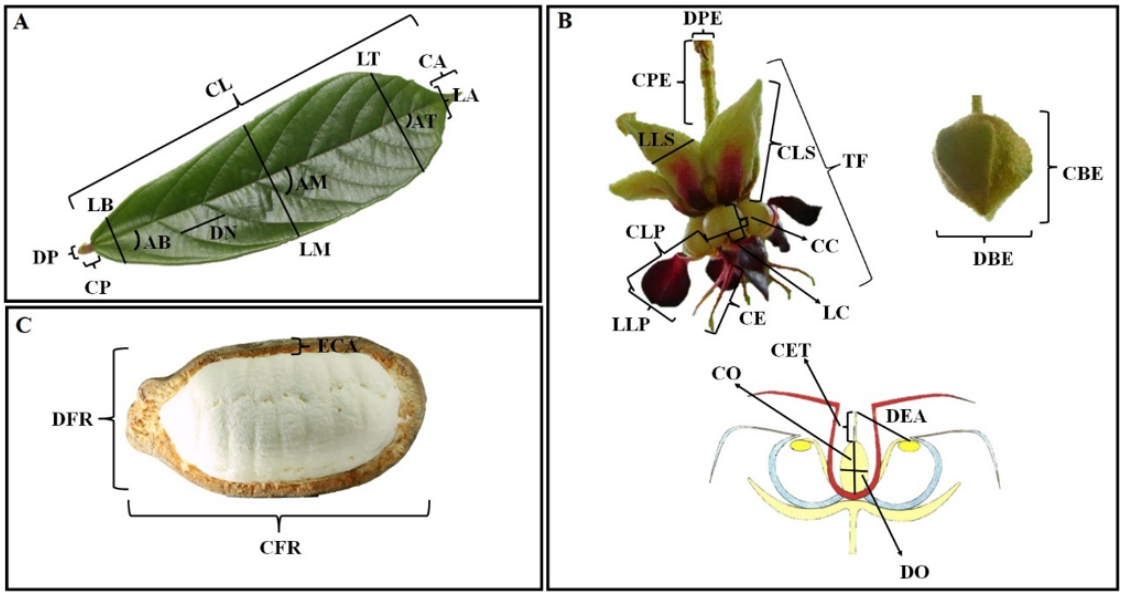


Figure 1. Approximate measurement locations of the 34 leaf, flower, and fruit characters (see Table 2 for acronyms) measured in 20 clonal access from the BAG cupuassu.

The phenotypic, genotypic, and residual variances were obtained from the mathematical hope of the mean squares of the model effects. From the variances, the heritability in the broad sense (h^2), and the residual (CV_e) and genetic coefficients of variation (CV_g) were estimated using Equations 1, 2, and 3, respectively:

$$h^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_f^2} \quad (1)$$

$$CV_e = \frac{\sqrt{\hat{\sigma}_e^2}}{\mu} \quad (2)$$

$$CV_g = \frac{\sqrt{\hat{\sigma}_g^2}}{\mu} \quad (3)$$

where: σ_g^2, σ_e^2, and σ_f^2 are the estimates of genotypic, residual, and phenotypic variances, respectively..

These variances were also used to calculate the genotypic correlations (r_g) between variables (Equation 4):

$$r_g = \frac{\hat{\sigma}_{g(xy)}}{\sqrt{\hat{\sigma}_{g(y)}^2 \hat{\sigma}_{g(x)}^2}} \quad (4)$$

where: $\sigma_{g(xy)}$ is the genotypic covariance between any x and y characters. Correlations were tested by Student t-test at 5% probability level.

The results of the analysis of variance (ANOVA) and the genotypic correlations were used as filters for character discarding, aiming for a better fit of the diversity analyses. Those variables where treatment effects were not significant in the ANOVA, or that had a high genetic

correlation (≥ 0.9) with another, more easily measured variable, were dropped.

With the final set of characters, the genetic diversity among access was studied by the generalized Mahalanobis distance (Mahalanobis, 1936) (Equation 5):

$$D^2_{ii'} = \delta' \psi^{-1} \delta \quad (5)$$

where: $D_{ii'}^2$ is the Mahalanobis distance between genotypes i and i', Ψ is the matrix of residual variances and covariances, and δ is the vector of variances.

The estimated distances were used to estimate the relative contribution of the characters to the divergence, using Singh (1981) criterion (Equation 6):

$$S_{.j}(\%) = \frac{S_{.j}}{\sum_{i}\sum_{i'} D^2_{ii'}} \quad (6)$$

where: $S_{.j}$ is the relative contribution of each variable to the dissimilarity.

Based on the distances, the access were grouped using the UPGMA method (Unweighted Pair-Group Method using Arithmetic averages), which uses the average distances. The dissimilarity of the access was represented by a dendrogram, and the formation of the groups took into account the criteria recommended by Mojena (1977). The consistency of the clustering was tested by the cohenetic correlation coefficient (CCC), which studies the relationship between the dissimilarity matrix and the cohenetic matrix, obtained from the clustering (Farris, 1969). CCC was tested by Student t-test at 5% probability. Statistical analyses were performed in Genes (Cruz, 2016) and R (R Core Team, 2021) software.

Results and Discussion

In six of the 34 characters, the treatment effect was non-significant, to name: CP, DP, AM, ECA, PFR, and NS, the first three being leaf characters and the remaining fruit characters (Table 3). The significance of this effect indicates the existence of differences between the averages of the access for a given character. In other words, the six non-significant characters for the access were not good indicators of variability in this study.

This result is based on the coefficients of genotypic variation (CV_g) and the heritabilities in the broad sense (h^2) of the discarded traits, which have the lowest values among the evaluated traits. The CV_g represents the genetic variability inherent in the experiment, for a trait (Resende & Alves, 2020). The h^2 represents the genetic proportion in the phenotypic variance. The six discarded characters have median and low h^2, following Resende & Alves (2020) criteria. In the face of diminished h^2 values, two non-mutually exclusive hypotheses are raised: i) the character has quantitative genetic control, that is, small additive effects of several alleles and with great

Table 3. Mean square of the treatment effect (MST), heritability in the broad sense (h^2), residual (CVe) and genotypic variation coefficients (CVg) of morphological characters† evaluated in 20 clonal access of cupuassu.

Character †	MST	h^2	CV_e	CV_g
CL	21.724 **	79.485	8.505	7.487
LB	1.325 **	71.274	13.687	9.642
LM	3.248 **	78.341	10.296	8.757
LT	2.278 **	72.700	11.421	8.335
CP	0.015 ns	30.950	9.000	2.695
DP	0.458 ns	39.899	13.397	4.882
CA	0.171 **	60.245	18.282	10.065
LA	0.017 **	63.288	16.409	9.635
AB	50.219 **	74.503	6.053	4.627
AM	20.714 ns	16.827	7.661	1.541
AT	36.861 *	51.941	7.330	3.408
DN	0.273 *	52.526	11.900	5.598
NPN	1.231 *	51.838	9.178	4.258
CBE	2.938 **	94.120	3.220	5.761
DBE	3.321 **	96.132	3.496	7.794
TF	8.197 **	85.336	4.204	4.535
DEA	1.738 **	97.650	5.284	15.232
CPE	13.433 **	87.414	7.991	9.419
DPE	0.046 **	62.150	6.734	3.859
CLS	4.248 **	93.425	3.625	6.111
LLS	1.184 **	84.411	6.099	6.346
CLP	0.809 **	65.089	6.446	3.936
LLP	1.623 **	90.229	5.936	8.067
CC	0.403 **	79.146	3.955	3.446
LC	0.523 **	84.961	5.310	5.644
CE	3.133 **	86.677	4.776	5.448
CET	0.215 *	52.217	14.443	6.752
CO	0.123 **	89.755	5.067	6.707
DO	0.041 **	72.869	5.378	3.942
CFR	821.383 **	59.336	9.042	4.885
DFR	111.597 *	44.684	7.012	2.818
ECA	2.121 ns	33.873	16.680	5.339
PFR	38,896.255 ns	15.952	14.848	2.893
NS	23.045 ns	.	14.564	.

*, ** and ns: Significant effect with p < 0.05 and p < 0.01, and non-significant effect respectively, according to Snedecor F test.
† See Table 2 for acronyms.

environmental influence (increased residual variance); and, ii) the character has low genetic variability (decreased genetic variance) (Schmidt et al., 2019). Both of these apply to dropped characters.

The other characters, on the other hand, have high heritability, denoting high variability and/or oligogenic genetic control. Studying genetic diversity from these traits will provide lower residual effects and more reliable results. If there is interest in selecting genotypes for any of these characters, the gains from selection will be of great magnitude (Dangi et al., 2018).

Another parameter to be noted is the residual coefficient of variation (CV_e). This represents the proportion of the experimental variation attributed to random errors (Cruz et al., 2014). Therefore, characters with lower CV_e values are desirable. A lower CV_e means less environmental noise in the results generated from the character in question. The

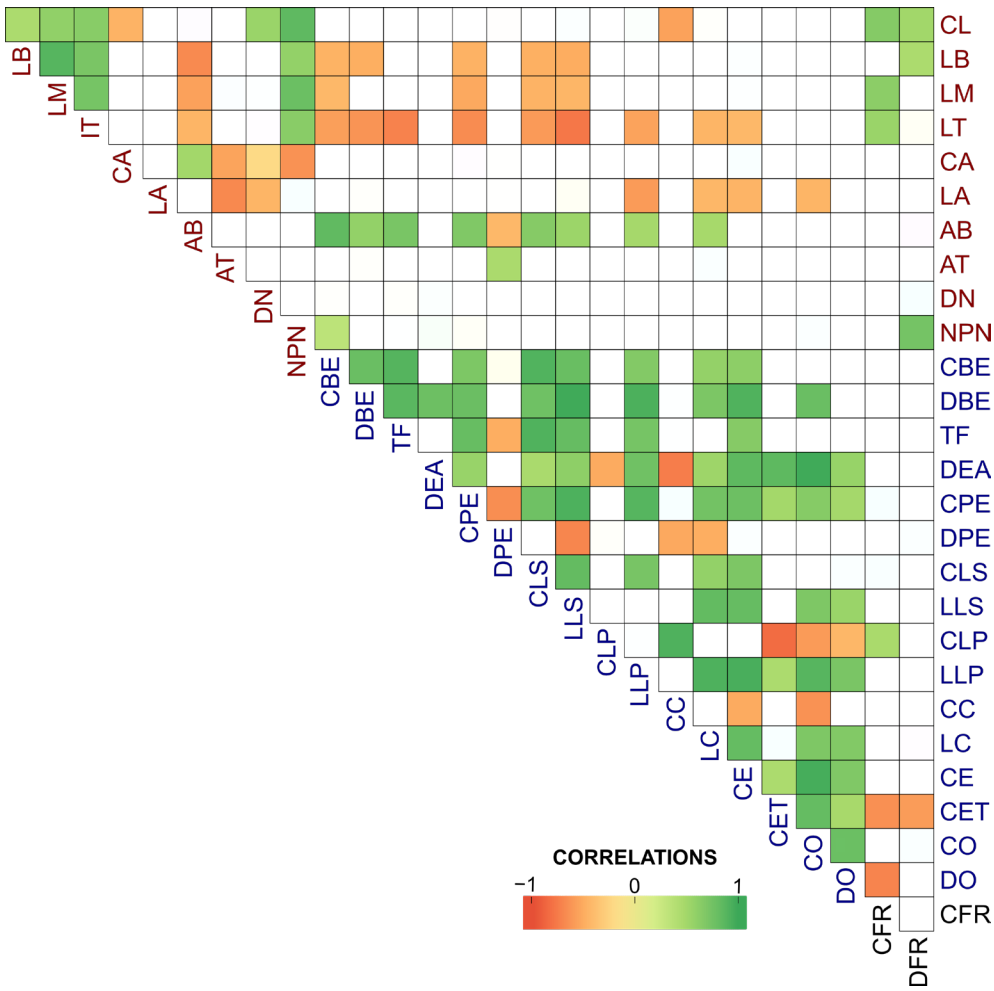
ratio between CV_g and CV_e reflects on the quality in genetic evaluation and the potential for gains (Resende & Alves, 2020). In this sense, the CBE, TF, CPE, CLS, LLS, LLP, and CE characters stand out, in which this ratio exceeds unity, a scenario considered optimal.

Most of the significant genetic correlations are between traits in the same group (Figure 2). There are several possible causes for genetic correlation, most notably pleiotropy, a permanent cause, and gene linkage, which can be reversible (Montesinos-López et al., 2019). It is expected that characters of the same group are influenced by the same genes, that is, that the gene products act in common biosynthetic pathways for several characters, or even that the final product of a chain of reactions that will determine a character affects a second route, determining other(s) characteristic(s) (Chen & Lübberstedt, 2010). Therefore, genetic correlation is a useful tool for indirect selection, prioritizing those traits that are easier to measure (Rehman et al., 2020).

The relationship between the floral characters illustrates the above facts. The characters DBE, DEA, LC, and CE are highly correlated with other floral characters and are more difficult to measure. The knurled button diameter (DBE) requires the use of an appropriate caliper. Measuring the distance between stigma and anther (DEA) requires expertise on the part of the collector. The width of the cogula and the length of the staminodes demand that the flower be opened, which must be done delicately, so as not to compromise the measurement. Therefore, these four floral characters were discarded.

The filtering performed from the aforementioned results was successful in discarding 10 variables (CL, LB, LM, LT, CA, LA, AB, AT, AT, DN, and NPN), leaving 24 for the dissimilarity computation. This was estimated by Mahalanobis distance, preferable over Euclidean distances because it considers the experimental precision and the relationship between variables by aggregating in its estimator the matrix of residual variances and covariances, obtained from ANOVA (Cruz et al., 2014).

Mahalanobis distances were used to make a dendrogram (Figure 3), using the UPGMA method, widely employed for this purpose (Andrade et al., 2017). Applying Mojena (1977) method to adopt a cutoff point, six groups are evident, proving the hypothesis of the presence of variability among access (Figure 3). The first group is the largest, comprising 11 of the 20 accesses. Another four groups consist of a single access, and a final group consists of five accesses.



† See Table 2 for acronyms.

Figure 2. Significant genetic correlations among morphological characters† evaluated in 20 clonal access of cupuassu. The acronyms of the foliar, floral and fruit characters are represented by the colors green, blue and black, respectively.

Analyzing the composition of the groups holistically, it is observed that the site of collection was not preponderant for the grouping, since in the same group there are acesses from different origins and in different groups, acesses from the same site. This pattern highlights the high intra- and inter-population variability observed in wild cupuassu populations by Alves et al. (2007). It also suggests that geographic location is not a good indicator of (dis)similarity, raising the importance of genetic divergence analysis, as performed in the present study (Cruz et al., 2014; Gopaulchan et al., 2019). The corroboration of this fact can be done after studying the divergence between acesses collected at more distant points.

Comparing the dendrogram with the heatmap (Figure 2), it can be seen that there is a difference between the clusters and the distances, which reduced the CCC value (0.643). However, these differences were not strong enough to make the CCC non-significant, giving reliability to the results presented. The cohenetic correlation coefficient (CCC) represents the degree of association between the dissimilarity matrix (Mahalanobis distance, in this study) and the cohenetic matrix generated by the UPGMA clustering. The CCC values are expected to be high and significant, indicating that there is a good coincidence between both matrices (Cruz et al., 2014). This property makes the CCC a good indicator of cluster consistency (Vittorazzi et al., 2018).

Mahalanobis distance can also be used to study the relative contribution of each character to the dissimilarity (Table 4), using the method proposed by Singh (1981). Among the five characters with the highest contribution, four are floral, to name: sepal length (CLS, 16.83%), knurled button length (CBE, 16.37%), petal width (LLP, 7.52%) and ovarian length (CO, 4.98%). The only non-floral character present among the first five is limbus length (CL, 4.31%). This highlights the importance of floral characters for differentiation between cupuassu access, a fact also observed by Alves et al. (2003).

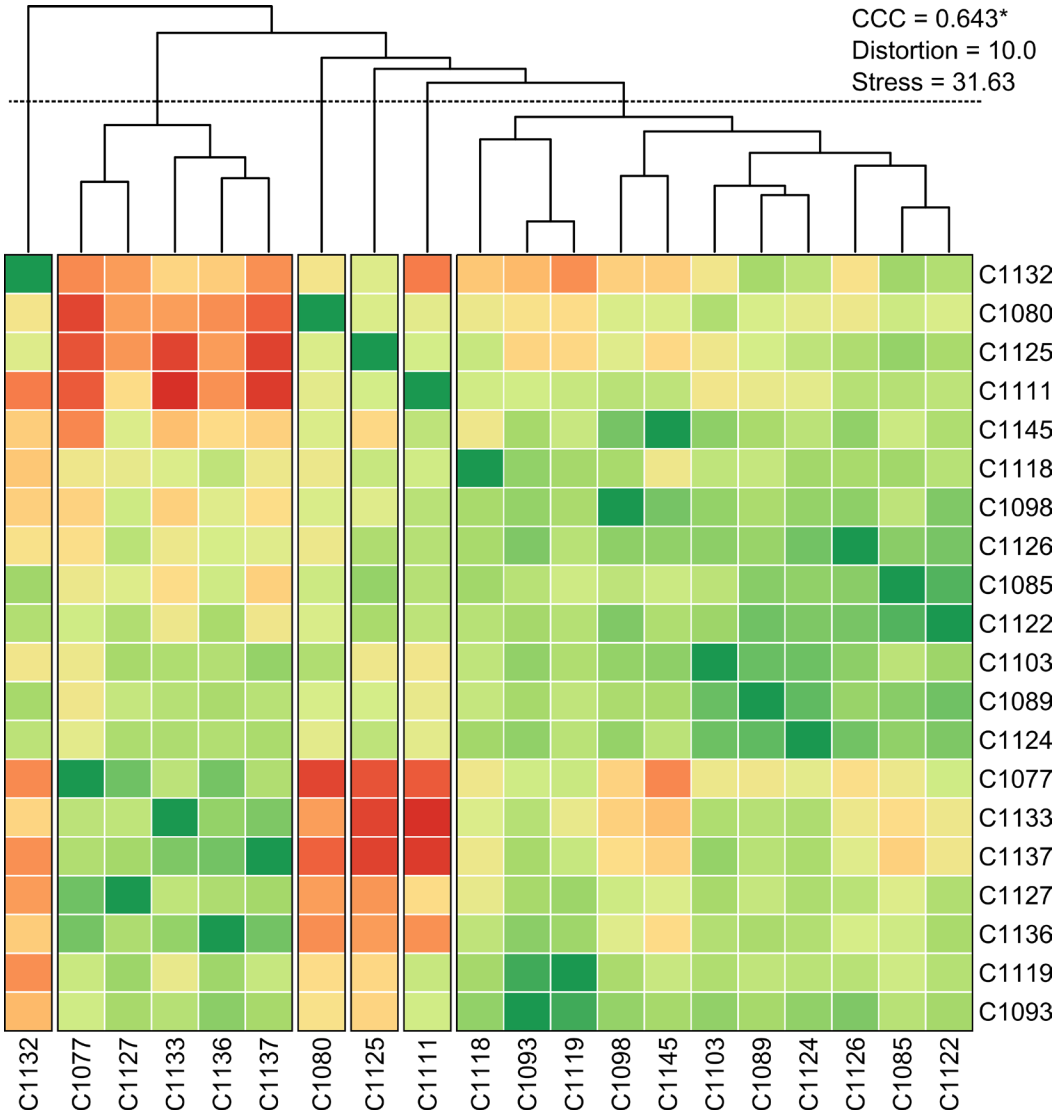


Figure 3. Genetic dissimilarity among 20 clonal access of cupuassu represented by the generalized Mahalanobis distance (Heatmap, smaller distances in greenish shades and larger distances in orange/reddish shades) and by the UPGMA clustering (Dendrogram); and cohenetic correlation coefficient (CCC) between the dissimilarity measuring methodologies. The dashed line in the dendrogram represents the division of the six groups formed.

Table 4. Decreasing order of S-value (de Singh, 1981) and relative contribution of morphological characters† evaluated for the dissimilarity in 20 clonal access of cupuassu.

Character†	S value	Relative contribution (%)
CLS	1,919.835	16.8341
CBE	1,866.835	16.3693
LLP	857.3289	7.5175
CO	568.0608	4.981
CL	491.7149	4.3116
CC	442.4215	3.8794
CPE	438.2241	3.8426
LM	434.0126	3.8056
CLP	383.4756	3.3625
AB	372.6659	3.2677
DPE	368.6083	3.2321
CA	359.7160	3.1542
DO	347.1528	3.044
LA	342.7344	3.0053
LT	316.3197	2.7736
TF	292.0769	2.5611
AT	285.5335	2.5037
CFR	268.3614	2.3531
LLS	264.2751	2.3173
DN	199.3742	1.7482
NPN	167.6851	1.4703
LB	153.6377	1.3472
DFR	145.5892	1.2766
Total	11,404.46	100

† See Table 1.

Conclusions

Genetic divergence was found among the 20 clonal cupuassu access from wild populations in Nova Ipixuna, PA, Brazil, which can be grouped into six distinct groups.

Sepal length, striated bud length, petal width, ovary length and limb length are the characters that contribute most to divergence.

Acknowledgments

To Embrapa Amazônia Oriental for yielding its productive space for the work, to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and to the Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) for granting the scholarships and to the technicians José Raimundo Quadros Fernandes, Antonio Fontel and José Oliveira who were essential in obtaining the data.

Compliance with Ethical Standards

Author contributions: Conceptualization: TGS; Formal analysis: RMA; Funding acquisition: RMA; Methodology: RMA; Project administration: RMA; Supervision: JLPN, SFSC; Software: TGS, JLPN, SFSC; Visualization: TGS, JLPN, SFSC; Writing - original draft: TGS; Writing - review & editing: TGS, JLPN, RMA, ABMS, ARMJ.

Conflict of interest: The authors declare that there is no conflict of interest

Financing source: The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, the Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP), the Embrapa Amazônia Oriental and the Universidade Federal do Ceará.

Literature Cited

Adeigbe, O. O.; Adewale, B. D.; Muyiwa, A. A. Variability and character association of bean biochemical traits of cocoa (*Theobroma cacao*) genotypes in four Nigerian field banks: Providing a platform for nutrition-based selection. Journal of Agriculture and Food Research, v. 3, e100115, 2021. https://doi.org/10/gj2pxh.

Alves, R. M.; Garcia, A. A. F.; Cruz, E. D.; Figueira, A. Seleção de descritores botânico-agronômicos para caracterização de germoplasma de cupuaçuzeiro. Pesquisa Agropecuária Brasileira, v. 38, n.7, p. 807-818, 2003. https://doi.org/10.1590/S0100-204X2003000700004.

Alves, R. M.; Sebbenn, A. M.; Artero, A. S.; Clement, C; Figueira, A. High levels of genetic divergence and inbreeding in populations of cupuassu (*Theobroma grandiflorum*). Tree Genetics & Genomes, v. 3, n. 4, p. 289-298, 2007. https://doi.org/10.1007/s11295-006-0066-9.

Andrade, E. K. V.; Andrade Júnior. V. C. de; Laia, M. L.; Fernandes, J. S. C.; Oliveira, A. J. M.; Azevedo, A. M. Genetic dissimilarity among sweet potato genotypes using morphological and molecular descriptors. Acta Scientiarum. Agronomy, v. 39, n. 4, p.447-455, 2017. https://doi.org/10.4025/actasciagron.v39i4.32847.

Chen, Y.; Lübberstedt, T. Molecular basis of trait correlations. Trends in Plant Science, v. 15, n. 8, p. 454-461, 2010. https://doi.org/10.1016/j.tplants.2010.05.004.

Cruz, C. D. Genes Software - extended and integrated with the R, Matlab and Selegen. Acta Scientiarum. Agronomy, v. 38, n. 4, p. 547-552, 2016. https://doi.org/10.4025/actasciagron.v38i3.32629.

Cruz, C. D.; Carneiro, P. C. S.; Regazzi, A. J. Modelos biométricos aplicados ao melhoramento genético. v. 2. 3.ed. Viçosa: UFV, 2014. 480p.

Dangi, R.; Kumar, A.; Khar, A. Genetic variability, heritability, and diversity analysis studies in short day tropical onion (*Allium cepa*). Indian Journal of Agricultural Sciences, v. 88, n. 6, p. 140-149, 2018. https://epubs.icar.org.in/index.php/IJAgS/article/view/80653/33642. 25 Aug. 2021.

Farris, J. S. On the cophenetic correlation coefficient. Systematic Biology, v. 18, n. 3, p. 279-285, 1969. https://doi.org/10.2307/2412324.

Gopaulchan, D.; Motilal, L. A.; Bekele, F. L.; Clause, S.; Ariko, J. O.; Ejang, H. P.; Umaharan, P. Morphological and genetic diversity of cacao (*Theobroma cacao* L.) in Uganda. Physiology and Molecular Biology of Plants, v. 25, n. 2, p. 361-375, 2019. https://doi.org/10/gj36bf.

Homma, A. K. O. Extrativismo vegetal na Amazônia: história, ecologia, economia e domesticação. Brasília: Embrapa, 2014. 467p. https://ainfo.cnptia.embrapa.br/digital/bitstream/item/124626/1/Livro-Extrativismo-HOMMA-ONLINE.pdf. 06 Sep. 2021.

Mahalanobis, P. C. On the gereralized distance in statistics. Proceedings of the National Institute of Sciencies of India, v. 2, n. 1, p. 49-55, 1936.

Migicovsky, Z.; Warschefsky, E.; Klein, L. L.; Miller, A. J. Using living germplasm collections to characterize, improve, and conserve woody perennials. Crop Science, v. 59, n. 6, p. 2365-2380, 2019. https://doi.org/10.2135/cropsci2019.05.0353.

Mojena, R. Hierarchical grouping methods and stopping rules: an evaluation. The Computer Journal, v. 20, n. 4, p. 359-363, 1977. https://doi.org/10.1093/comjnl/20.4.359.

Montesinos-López, O. A.; Montesinos-López, A.; Hernández, M. V.; Ortiz-Monasterio, I.; Pérez-Rodríguez, P.; Burgueño, J.; Crossa, J. Multivariate bayesian analysis of on-farm trials with multiple-trait and multiple-environment data. Agronomy Journal, v. 111, n. 6, p. 2658-2669, 2019. https://doi.org/10.2134/agronj2018.06.0362.

R Core Team. R: A Language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2021. https://www.r-project.org. 25 Aug. 2021.

Rehman, A.; Mustafa, N; Du, X.; Azhar, M. T. Heritability and correlation analysis of morphological and yield traits in genetically modified cotton. Journal of Cotton Research, v. 3, n. 1, e23, 2020. https://doi.org/10.1186/s42397-020-00067-z.

Resende, M. D. V.; Alves, R. S. Linear, generalized, hierarchical, bayesian and random regression mixed models in genetic/genomics in plant breeding. Functional Plant Breeding Journal, v. 2, n. 2, p. 1-31, 2020. https://doi.org/10.35418/2526-4117/v2n2a1.

Schmidt, P.; Hartung, J.; Bennewitz, J.; Piepho, H. P. Heritability in Plant Breeding on a Genotype-Difference Basis. Genetics, v. 212, n. 4, p. 991-1008, 2019. https://doi.org/10.1534/genetics.119.302134.

Singh, D. The relative importance of characters affecting genetic divergence. Indian Journal of Genetics and Plant Breeding, v. 41, n. 2, p. 237-245, 1981.

Souza, A. G. C. Boas práticas agrícolas da cultura do cupuaçuzeiro. Manaus: Embrapa Amazônia Ocidental, 2007. 56p. https://ainfo.cnptia.embrapa.br/digital/bitstream/CPAA-2009-09/18684/1/Livro_BPA.pdf. 23 Oct. 2021.

Souza, A. das G. C. de; Alves, R. M.; Sousa, N. R.; Sousa, A. M. G. de. Domestication and breeding of the cupuassu tree. In: Borém, A.; Lopes, M.T.G.; Clement, C.R.; Boda, H. (Eds.). Domestication and breeding: Amazonian species. 1.ed. Viçosa: Suprema, 2012. p.319-331.

Vittorazzi, C.; Junior, A. T. A.; Guimarães, A. G.; Silva, F. H. L; Pena, G. F.; Daher, R. F.; Gerhardt, I. F. S.; Oliveira, G. H. F.; Santos, P. H. A. D.; Souza, Y. P.; Kamphorst, S. H.; Lima, V. J. Evaluation of genetic variability to form heterotic groups in popcorn. Genetics and Molecular, v. 17, n. 3, e18083, 2018. https://doi.org/10.4238/gmr18083.