Ciência e Agrotecnologia
SCIENCE AND AGROTECHNOLOGY

Ciência e Agrotecnologia, 47:e004923, 2023
http://dx.doi.org/10.1590/1413-7054202347004923

Agricultural Sciences
eISSN 1981-1829

# Cacao genotypes cultivated in agroforestry systems in Bahia have wide genetic variability in morpho-agronomic characters

## Cacaueiros cultivados em sistema agroflorestais da Bahia revelam ampla variabilidade genética em caracteres morfoagronômicos

Marc Ernso Georges[1], Cláusio Antônio Ferreira Melo[2], Margarete Magalhães de Souza[3], Ronan Xavier Corrêa[3*]

[1]Universidade Estadual de Santa Cruz/UESC, Programa de Pós-Graduação em Genética e Biologia Molecular, Ilhéus, BA, Brasil
[2]Centro de Ciência do Cacau, Mars Wrigley, Ilhéus, BA, Brasil
[3]Universidade Estadual de Santa Cruz/UESC, Departamento de Ciências Biológicas, Ilhéus, BA, Brasil
[*]Corresponding author: ronanxc@uesc.br
*Received in March 28, 2023 and approved in September 15, 2023*

### ABSTRACT

Cacao fruits are agronomically important and show wide variability for several morphological descriptors. The present study aimed to characterize the genetic distance of 51 genotypes of *Theobroma cacao* L. in the cocoa-producing region of Southern Bahia (Brazil) based on morpho-agronomic descriptors. The inference of genetic similarity was performed from the phenotypic data derived simultaneously from qualitative and quantitative variables, using the Ward-MLM procedure (Modified Location Model) with the SAS® software. The distance matrix was obtained using the Gower logarithmic function. For this, 28 descriptors were evaluated, five qualitative and 23 quantitative. Furthermore, using the likelihood function procedure, the optimal number of groups was five indicating wide variability. Of 23 quantitative descriptors evaluated by fruits and leaves, 95% showed significant differences. The exception was seed width. Groups 1 and 5 were the most distant, while groups 2, 3, and 4 were closest to each other. The greatest difference between increments was 35.32 for the fifth group. The analysis of the first two canonical axes revealed that both represented 81.88% of the variation, with CAN1 and CAN2 responsible for 53.66% and 28.22% of the variation. The most promising genotypes for breeding programs belong to group 5 due to their superior performance for almost all characteristics analyzed.

**Index terms:** *Theobroma cacao*; Ward-MLM; improvement; genetic distance; morphological diversity.

### RESUMO

Os frutos de cacau são de grande importância agronômica e apresentam ampla variabilidade em vários descritores morfológicos. O presente estudo teve como objetivo caracterizar a distância genética de 51 genótipos de *Theobroma cacao* L. na região produtora de cacau do Sul da Bahia (Brasil) com base em descritores morfo-agronômicos. A inferência de similaridade genética foi realizada a partir dos dados fenotípicos derivados simultaneamente de variáveis qualitativas e quantitativas, utilizando o procedimento Ward-MLM (Modified Location Model) com o software SAS®. A matriz de distância foi obtida utilizando a função logarítmica de Gower. Para isso, 28 descritores foram avaliados, cinco qualitativos e 23 quantitativos. Além disso, usando o procedimento de função de verossimilhança, o número ótimo de grupos foi cinco, indicando ampla variabilidade. Dos 23 descritores quantitativos avaliados em frutos e folhas, 95% apresentaram diferenças significativas, exceto a largura da semente. Os grupos 1 e 5 foram os mais distantes, enquanto os grupos 2, 3 e 4 foram mais próximos. A maior diferença entre os incrementos foi de 35,32 para o quinto grupo. A análise das duas primeiras variáveis canônicas representou um total de 81,88% da informação para a distribuição da variação, com CAN1 responsável por 53,66% e CAN2 por 28,22% da variação. Os genótipos mais promissores para programas de melhoramento pertencem ao grupo 5 devido ao seu desempenho superior para quase todas as características analisadas.

**Termos para indexação:** *Theobroma cacao*; Ward-MLM; melhoramento; distância genética; diversidade morfológica.

## INTRODUCTION

Cacao (*Theobroma cacao* L.) is a neotropical understory tree native to South America that belongs to the *Malvaceae* family (Motamayor et al., 2002). Currently, cacao is cultivated in more than 50 humid tropical countries (Prazeres; Lucas; Marta-Costa, 2021) and is considered one of the most valuable agrobiological goods inherited from the Mesoamerican culture (Oztürk; Young, 2017). *T. cacao* is a perennial commercial crop and has a popular fruit known as cocoa, which provides the raw material for the cosmetic, pharmaceutical, and food industries, mainly for the manufacture of chocolate (Whitlock; Bayer; Baum, 2001; Zuidema et al., 2005). Furthermore, cocoa farming plays a vital social and economic role, covering over 10 million hectares and providing income for countless families. Approximately 40 to 50 million people depend on cocoa production for subsistence (Somarriba; López Sampson, 2018). Therefore, *T. cacao's* genetic resources require conservation actions (Angiosperm Phylogeny Group - APG II, 2003).

The International Cocoa Organization (ICCO) estimated world production of cocoa beans at 4.8 million tons over 2021-2022. This represents a drop of around 8% compared to the 2020-2021 harvest and may be due to several challenges, such as the effect of old and new diseases affecting this crop; price variations on the international market; the low amount of inputs; and the lack of improved planting material that brings greater productivity and support to family farming (Lembang et al., 2019). Brazil is the second largest cocoa producer in the Americas behind Ecuador, with 290 thousand tons, and also stands out as the seventh in the world ranking (*International Cocoa Organization* - ICCO, 2022). Brazil's main cocoa producing States are Bahia, Pará, Rondônia, Espírito Santo, Amazonas, and Mato Grosso, which jointly have an estimated area of 620,215 hectares. Bahia is the largest producing State in the Northeast region, occupying 70.95% of the national area, that is, 440,050 hectares with 126,050 tons of cocoa beans. At the same time, the North leads the national production with 52.35% (Instituto Brasileiro de Geografia e Estatística - IBGE, 2022).

Initially, cacao varieties were classified into two groups, according to their geographic distribution and seed and fruit characteristics: Forastero and Criollo (Cheesman, 1944; Cuatrecasas, 1964). Subsequently, a third genetic group called Trinitario appeared, revealing intermediate characteristics (Engels, 1986). The Trinitario hybrid emerged naturally in Trinidad, but its classification is unclear (Cheesman, 1944). A new classification was proposed, dividing cacao into ten distinct genetic groups based on genetic diversity inferred by molecular markers (Motamayor et al., 2008).

Carrying out crosses between varieties and cacao clones is important for the genetic improvement of the crop and the quality of agronomic traits. Therefore, conducting research and having accurate information on the diversity of cacao varieties and clones is crucial for achieving high-yielding and high-quality plant materials. Therefore, information on the morphological characteristics of cocoa can be used to discriminate among different groups and hybridization can lead to cultivar development (Lembang et al., 2019). In this context, if special attention is paid to genetic selection programs, cocoa productivity can increase considerably. In addition, farmers must have access to disease-resistant cultivars exhibiting high productivity and controlled growth, which can effectively contribute to increased yield and income (Mustiga et al., 2018). This research aimed to evaluate the phenotypic variability based on morpho-agronomic descriptors, including agronomically important traits related to seeds, fruits, leaves, and seedling growth in 51 local cacao varieties in Bahia. The research seeks to identify the most promising variations for productivity at a commercial scale in this region.

## MATERIAL AND METHODS

### Characterization and description of biological materials in the collection areas

Genotype samples were collected in a completely randomized design in an agroforestry system known as 'cabruca' (i.e., predominantly grown under the shade of native trees) in five rural properties located in the Southern Bahia, Brazil, in the municipalities of Arataca (15º 15'49.43" S; 39º 30' 42.65") and Ilhéus (14º 45› 20.89›› S; 39º 09› 08.51››). In Ilhéus, samples were collected in three different locations: Almada, Banco Central, and Salobrinho. In Arataca and Banco Central, the sampled plants were organic crops. In other areas, conventional chemical fertilization schemes are eventually adopted but plants have not been fertilized in the last eight years. All cacao plantations were covered by native trees with around 50% insolation, mechanical mowing of spontaneous plants three times a year, and pruning undesirable cacao shoots once or twice a year in the last four years. These genotypes are locally recognized and commercially available in the region, and most of their resistance capacity against *Moniliophthora perniciosa* was previously assessed (Pinto; Pires, 1998; Marssaro et al., 2020). In total, 51 genotypes were evaluated: Comum, Maranhão, Pará, Parazinho, Redondo, and clones CCN51 and PS1319. These genotypes present distinct external morphological characteristics (Table 1).

### Morpho-agronomic descriptors analyzed

The pods used in this study were carefully selected, considering their representativeness and quality. Specifically, we used cocoa pods from open pollination (OP) for this research. From June to December 2022, three ripe fruits and five leaves per genotype were collected and transported to the Plant Improvement Laboratory. Samples were measured with a millimeter ruler, caliper, and digital scale. A total of 28 characters were evaluated for the fruits (18 characters were quantitative and four qualitative), leaves (two characters were quantitative and one qualitative), and seedling growth (three quantitative characters). The cocoa descriptors used in this study were previously described by Engels et al. (1980) Restrepo and Urrego (2018), including qualitative and quantitative descriptors (Table 2).

### Growth parameters

The seeds were selected for growth evaluation, washed with water to remove the pulp, and left in a vase with water in the laboratory at room temperature. Three days later, with roots between five and seven millimeters long, seeds were transplanted into germination tubes of 175 $cm^3$ filled with a substrate (sieved soil and commercial substrate tropstratum forest 2:1 v/v, one seed per tube) located at the greenhouse. One month later, 26 seedlings measuring 7 to 10 cm and/or having two leaves were selected and transplanted into polyethylene bags measuring 15 x 28 x 80 cm with the same substrate. The experiment was conducted at 25 to 27 °C, with watering every two days. Data on the plant growth (height, diameter, and the number of leaves) were collected fortnightly from 15 to 75 days after transplanting.

### Statistical analysis

The data collected were submitted to analysis of variance (ANOVA) and Scott-Knott mean test at 5% probability, using the Sisvar® *software* version 5.8 (Ferreira, 2011). Phenotypic data derived simultaneously from qualitative and quantitative variables were used to infer genetic similarity by applying the Ward-MLM multivariate procedure using SAS® *software* (SAS Institute, 2004). The distance matrix was obtained through the Gower logarithmic function (Gower, 1971), as used before by Melo et al. (2015). This defines the optimal number of groups based on pseudo-F and pseudo-$T^2$ criteria combined with the likelihood profile associated with the likelihood ratio test.

## RESULTS AND DISCUSSION

The morpho-agronomic characterization in an agricultural population is extremely important to assess the existing morphological diversity(Ramírez-Guillermo et al., 2018). Among 20 quantitative descriptors evaluated for fruits and leaves, 95% (except SW) showed significant differences by the F test. In the individual analysis of the TNVS per fruit, the Scott & Knott test revealed a significant effect. Genotypes 2050 and 105707 had the highest number of seeds, with approximately 6 and 4 non-viable seeds, respectively. Based on the different results, we observed that the highest TNSF was found in genotypes CCN51 and 7016 (53 seeds on average). The most elevated LS (29.3 mm) was obtained in 7015. CCN51 had the highest TWF (1016.3 g), followed by 2074 (826.8 g) and 2058 (776.3 g). On the other hand, 3003 had the lowest TWF (220.3 g). The data obtained for the weights of fruits were superior to those found by Reges et al. (2021), who evaluated clones CCN51, CEPEC 2004, CEPEC 2005, and PS 1319 harvested in an orchard located in the municipality of Russas, State of Ceará, Brazil. However, these authors also observed a great variation in the weights of the fruits, which is related to the specific characteristics of the population studied.

A significant effect ($P<0.05$) was observed between almost all descriptors within each genotype (Table 3). The coefficients of variation (CV %) ranged from 6.75% to 26.88% for the different samples of this study. The total weight of CPS was significantly different between genotypes, with the CCN51 clone showing the highest weight (709.3 g) followed by the genotypes 2074 (696.6 g) and 2085 (603 g). Concerning the PW, CCN51 had the highest weight (30.9 g). The Scott-Knott test revealed that the genotype 2002 showed a significantly higher value

**Table 1**: Local names and morphological characteristics of fruits of 51 analyzed *Theobroma cacao* genotypes in southern Bahia, Brazil.

| Local variety name | Number of genotypes | External morphological characteristics of fruits |
|---|---|---|
| Comum | 18 | Oblong, smooth, rounded ends |
| Maranhão | 14 | Oblong, grooved, constricted base, pointed apex |
| Pará | 06 | Rounded, smooth, rounded ends |
| Parazinho | 10 | Rounded, smooth, small, thin shell |
| Redondo | 01 | Rounded, smooth, straight ends |
| Clone: CCN51; PS1319 | 02 | Oblong, very large and red, apex pointed |

for MinTM (15 mm), while 5008 had the lowest value (7 mm). In contrast, genotype 7015 had higher MaxTM values (19.6 mm). For ET, genotype 2041 had the highest value (9 mm) and PS1319 had the lowest (2 mm). We also found that CCN51 (8.33 mm) and 2058 (5 mm) had the highest and lowest values for GD, respectively.

In contrast to the individual analyses, the Ward-MLM multivariate procedure allowed identifying almost all the highest mean values. The double Ward-MLM strategy enabled us to define the number of groups in which the analyzed germplasm is distributed by detecting the highest increment of the likelihood function. In that case, the

**Table 2:** Qualitative and quantitative morphological variables recorded in 51 *Theobroma cacao* genotypes grown in 'cabruca' agroforests in Arataca and Ilhéus, Bahia.

| Name/Code | Biometric measurement mode or assessment / Scale or unit |
|---|---|
| Qualitative descriptors | |
| Fruit color | Predominant color of the fruit (epidermis) - Yellow (1) and Red (2) |
| Presence of basal constriction | Absent (1), Present (2) |
| Pod shell apex shape | Rounded (1), Intermediate (2), Pointed (3) |
| Roughness | Smooth (1), Intermediate (2), Rough (3) |
| Leaf apex shape | Short acuminate (1), Long acuminate (2), Obtuse (3) Acute (4) |
| Quantitative descriptors | |
| Total fruit weight (TFW) | Weight of the whole fruit (g) |
| Fruit length (FL) | Linear distance from the fruit apex to the base(mm) |
| Fruit diameter (FD) | Cross section of the equatorial line of the shell (mm) |
| Groove depth (GD) | Line perpendicular inside the point of the secondary groove (mm) |
| Fruit longitudinal perimeter (FLP) | Longitudinal perimeter of the whole fruit (apex, base, apex) (mm) |
| Fruit transverse perimeter (FTP) | Transverse perimeter of the fruit on the equatorial line (mm) |
| Cocoa pod shell (CPS) | Weight of pod shell without seeds and placenta (g) |
| Placenta weight (PW) | Weight of the placenta with moisture (g) |
| Total weight of seeds per fruit (TWS) | Weight of seeds with pulp (g) |
| Seed weight with pulp (10SWP) | Weight of 10 seeds with pulp (g) |
| Seed weight without pulp (10S/OP) | Weight of 10 seeds without pulp (g) |
| Total number seed per fruit (TNSF) | Number of viable and non-viable seeds per fruit |
| Total number of non-viable seeds (TNVS) | Number of non-viable seeds per fruit |
| Maximum thickness of the mesocarp (MaxTM) | Maximum external thickness of the mesocarp (mm) |
| Minimum thickness of the mesocarp (MinTM) | Minimum external thickness of the mesocarp (mm) |
| Endocarp thickness (ET) | Internal thickness of shell / endocarp (mm) |
| Seed length (SL) | Measured from the embryo to the seed apex (mm) |
| Seed width (SW) | Measured from the widest part of the seed (mm) |
| Plant height ***** (PH) | Measured from the crown to the terminal bud (mm) |
| Plant diameter ***** (PD) | Diameter of the trunk (mm) |
| Number of leaves per plant ***** (NLP) | Total number of leaves per plant |
| Leaf length (LL) | Linear distance from the leaf apex to the base (mm) |
| Leaf width (LW) | Measured from the widest part of the leaf (mm) |

*****Young plants (1 to 4 months old), 26 progenies of each genotype.

pseudo-F and pseudo-$t^2$ statistical scores revealed five groups (Figure 1a and Table 4). The greatest difference between increments was 35.32 for the fifth group. The analysis of the first two canonical axes revealed that both represented 81.88% of the information for the variation distribution, with 53.66% in CAN1 and 28.22% in CAN2 (Figure 1b). Thus, it was possible to visualize the genetic diversity and the relationship among the groups based on the genetic distances among cacao types. The distribution of the five groups is shown in the dispersion (Figure 1b). Genotypes in groups 2, 4, and 5 were close to each other, whereas the ones from groups 1 and 3 were more distant from the other groups. The dissimilarity of the groups based on the Mahalanobis distance by the Ward-MLM strategy demonstrated a wide range of variation, with groups 2 and 4 being the closest (8.95). In contrast, groups 1 and 5 were the more distant (51.62) (Table 5).

The Ward-MLM strategy grouped 12 genotypes in groups 1 and 4, 10 in group 2, 9 in group 3, and 8 in group 5 (Tables 6 and 7). The genotypes with higher average values for fruit characteristics were in group 5: TWF (734.46 g), FL (177.01 mm), FD (94.67 mm), GD (4 mm), PW (18.63 g), FLP (443.7 mm), FTP (308.51 mm), and CPS (535.57 g). Still, in group 5, high values were observed the descriptors TWS (161.25 g), 10SWP (35.58), 10S/OP (24.66 g), TNSF (almost 45 seeds), and MinTM (10.96 mm). Group 3 had higher values for MaxTM (15.9 mm), LS (26.91 mm), and ET (6.8 mm). Group 4 had a higher SW (14.16 mm). Finally, for leave descriptors, group 2 had the highest value for LL (348.38 mm) and LW (122.58 mm), while, group 4 had the lowest values for these descriptors (278.13 mm and 98.55 mm, respectively).

**Table 3:** Analysis of variance of morpho-agronomic descriptors (fruits and leaves) for 51 genotypes of *Theobroma cacao* grown in 'cabruca' agroforests in Arataca and Ilhéus, Bahia.

| Characters | Mean Square | | Mean | CV (%) |
|---|---|---|---|---|
| | Genotypes | Error | | |
| TFW (g) | 67842.78* | 10181.37 | 508.64 | 19.84 |
| FL (mm) | 2108.47* | 330.87 | 142.70 | 12.75 |
| FD (mm) | 151.44* | 34.80 | 87.42 | 6.75 |
| GD (mm) | 4.92* | 0.50 | 2.73 | 26.08 |
| PW (g) | 73.13* | 11.08 | 12.38 | 26.88 |
| FLP (mm) | 8687.44* | 880.97 | 374.16 | 7.93 |
| FTP (mm) | 1926.15* | 435.47 | 374.16 | 7.35 |
| CPS (g) | 42569.70* | 6828.60 | 370.01 | 22.33 |
| TWS (g) | 3379.11* | 670.01 | 122.05 | 21.21 |
| 10SWP (g) | 102.21* | 12.63 | 29.23 | 12.16 |
| 10S/OP (g) | 40.95* | 6.81 | 20.78 | 12.56 |
| TNSF | 96.38* | 60.83 | 41.30 | 18.88 |
| MaxTM (mm) | 11.68* | 4.36 | 14.43 | 14.48 |
| MinTM (mm) | 10.53* | 2.64 | 10.66 | 15.26 |
| ET (mm) | 4.84* | 1.26 | 5.99 | 18.74 |
| LS (mm) | 7.75* | 3.53 | 25.79 | 7.29 |
| WS (mm) | 3.05$^{NS}$ | 2.43 | 13.33 | 11.71 |
| LF (mm) | 12802* | 2065.72 | 316.10 | 14.38 |
| WF (mm) | 2008.53* | 224.15 | 113.63 | 13.18 |

* Significant at 5% by F test.

$^{NS}$No significant.

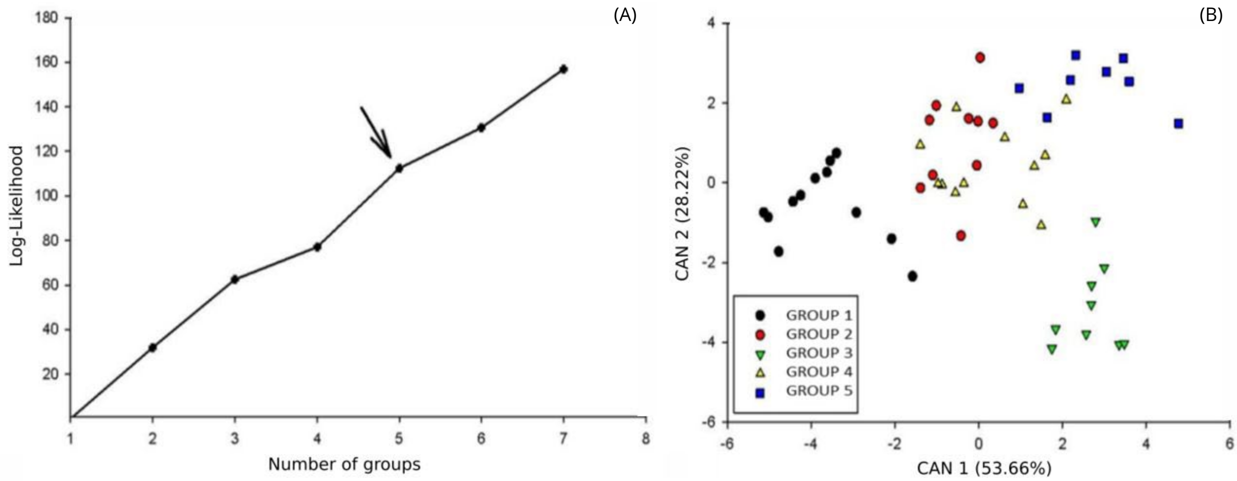


**Figure 1:** Optimal number of groups and dispersion of 51 genotypes of *Theobroma cacao* grown in 'cabruca' agroforests in Arataca and Ilhéus, Bahia. (a) Log-likelihood graph showing the optimal number of groups for the cacao genotypes analyzed. (b) Multivariate analysis by the dual Ward-MLM strategy.

**Table 4:** Number of groups formed by the logarithmic probability function (Log-Likelihood) and its increment for quantitative variables based on the morpho-agronomic characteristics of 51 genotypes of *Theobroma cacao* grown in 'cabruca' agroforests in the Southern of Bahia, Brazil.

| Number of groups | *Log-Likelihood* | increment |
|---|---|---|
| 1 | -2993.26 | 0.00 |
| 2 | -2961.30 | 31.96 |
| 3 | -2930.74 | 62.51 |
| 4 | -2916.21 | 77.04 |
| 5 | -2880.89 | 112.36* |
| 6 | -2862.59 | 130.66 |
| 7 | -2836.25 | 157.00 |
| 8 | -2784.40 | 208.85 |
| 9 | -2750.37 | 242.88 |

***** Largest increment for the formation of five groups by the logarithmic function.

Analyzing the variables length, diameter, and longitudinal and transverse perimeter of the fruit, genotypes in groups 1 and 2 were largely rounded, while those in groups 3 and 4 were rounded with rounded ends. Finally, the fruit of genotypes in group 5 had an oblong shape.

The pod shell, as well as its morphology, plays an important role in defining fruit types. It is important to highlight that a combination of independent inheritance characteristics determines the morphology of the pod shell, and the same tree or the same clone has identical pod shell characteristics (Wood; Lass, 1985). The parameters total of weight CPS, PW, ET, and MinTM showed higher average values in group 5. The cocoa pod shell has nutrients such as potassium to be recycled and used in cacao culture (Indiarto, 2021). They can also be transformed into fuel briquettes, adding value to the product and simultaneously solving the waste disposal problem (Ofori; Akoto, 2020). The shell thickness plays an important role in resistance to *Phytophthora palmivora*; susceptible cacao fruits have a thinner shell and higher moisture percentage (Reges et al., 2021).

**Table 5:** Distance between the groups formed by the Ward-MLM method verified for evaluating 51 genotypes of *Theobroma cacao* grown in 'cabruca' agroforests in the Southern of Bahia, Brazil.

| Groups | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 14.97 | 47.97 | 26.13 | 51.62 |
| 2 | | 0 | 28.70 | 8.95 | 15.96 |
| 3 | | | 0 | 26.18 | 32.72 |
| 4 | | | | 0 | 23.59 |
| 5 | | | | | 0 |

The weight of the placenta, a by-product of the extraction of cocoa beans, can potentially increase the bean's value as a food additive or ingredient due to its high fiber content. Despite being an important source of income for small and medium farmers, the use of the placenta needs to be addressed (Scheuer et al., 2022). Thus, the cocoa placenta can be an essential additive in food products and a useful bioactive for human and animal nutrition and health (Goude; Adingra; Gbotognon, 2019).

The yield of valuable raw almond material has a considerable influence on the value of cocoa for the chocolate industry, as the price and its importance vary according to the size and weight of the almond (CAOBISCO/ECA/FCC, 2015). Therefore, a fruit with more seeds and a larger size will imply a more favorable yield. The almond sizes

**Table 6:** Groups formed by Ward-MLM methods based on quantitative data on 51 genotypes of *Theobroma cacao* grown in 'cabruca' agroforests in the Southern of Bahia, Brazil.

| Groups | Genotypes |
|---|---|
| 1 | CCN51; 2058; 2041; 2002; 7015; 7016; 5012; 7011; 3009; 2079; 2000; 7018 |
| 2 | 2056; 2096; 2087; 2066; 5016; 2049; 2076; 5011; 2037; 3006 |
| 3 | 2046; PS1319; 5004; 5006; 2048; 2039; 2050; 5008; 3005 |
| 4 | 2074; 2075; 7017; 2001; 2003; 3008; 3004; 5014; 3002; 3003; 102202; 105614 |
| 5 | 2011; 105707; 7019; 105710; 3007; 3010; 2021; 2089 |

**Table 7:** Means of 19 quantitative descriptors for each of the five groups formed by the Ward-MLM method and the first two canonical axes.

| Descriptors | 1 | 2 | 3 | 4 | 5 | CAN1 | CAN2 |
|---|---|---|---|---|---|---|---|
| TFW | 337.42 | 502.31 | 595.68 | 473.96 | 734.46 | 0.825128 | 0.240658 |
| FL | 116.54 | 140.18 | 152.22 | 143.02 | 177.01 | 0.698677 | 0.264823 |
| FD | 80.5 | 89.25 | 91.48 | 85.78 | 94.67 | 0.65909 | 0.150247 |
| GD | 1.57 | 2.21 | 3.52 | 2.65 | 4 | 0.66247 | 0.045324 |
| PW | 7.6 | 11.66 | 14.17 | 11.91 | 18.63 | 0.675928 | 0.237769 |
| FLP | 313.84 | 370.76 | 397.18 | 373.5 | 443.7 | 0.776507 | 0.259128 |
| FTP | 256.23 | 285.96 | 298.2 | 279.91 | 308.51 | 0.723617 | 0.136944 |
| CSP | 237.95 | 365.41 | 447.44 | 335.4 | 535.57 | 0.822014 | 0.186128 |
| TWS | 87.44 | 123.78 | 125.63 | 125.24 | 161.25 | 0.65024 | 0.359366 |
| 10SWP | 23.85 | 29.51 | 31.2 | 27.8 | 35.58 | 0.627745 | 0.222522 |
| 10S/OP | 17.1 | 21.09 | 22.23 | 20.2 | 24.66 | 0.660643 | 0.209289 |
| TNSF | 38.98 | 42.22 | 39.86 | 42.07 | 44.51 | 0.204591 | 0.290184 |
| MaxTM | 12.04 | 13.9 | 15.9 | 15 | 15.21 | 0.691704 | -0.076248 |
| MinTM | 9.22 | 10.45 | 10.78 | 11.75 | 10.96 | 0.352306 | 0.102169 |
| ET | 5.59 | 5.96 | 6.8 | 6.03 | 5.87 | 0.250288 | -0.245525 |
| SL | 24.46 | 26.06 | 26.91 | 26.26 | 25.41 | 0.402818 | -0.186676 |
| SW | 12.49 | 13.28 | 12.6 | 14.16 | 13.68 | 0.251024 | 0.443804 |
| LL | 332.25 | 348.38 | 297.44 | 278.13 | 318.15 | -0.25423 | 0.157029 |
| LW | 118.93 | 122.58 | 115.33 | 98.55 | 111.3 | -0.154423 | -0.07506 |

TFW – Total Fruit Weight; FL- Fruit Length; FD – Fruit Diameter; GD – Groove Depth; TWS – Total Weight of Seeds per Fruit; PW – Placenta Weight; FLP – Fruit Longitudinal Perimeter; FTP – Fruit Transverse Perimeter; CPS– Cocoa Pod Shell Weight; MaxTM – Maximum Thickness of Mesocarp; MinTM – Minimum Thickness of Mesocarp; ET – Endocarp Thickness; 10SWP – Weight of 10 Seeds with Pulp; 10S/OP – Weight of 10 Seeds Without pulp; TNSF – Total Number of Seeds per Fruit; SL – Length of 4 Seeds per Fruit; SW – Width of 4 Seeds per Fruit, LL – Leaf Length; LW – Leaf Width.

interesting for the processing industries have an estimated variable format, on average 20 mm in length and 10 mm in width, and each fruit contains between 20 and 50 seeds (Oetterer, 2006). Genotypes 7016, 5011, 2002, 5012, and 2046 were mainly distinguished by their high number of seeds and sizes for primary trait selection for the chocolate industry. However, other plants within these same varieties could be selected because their characteristics were similar, such as genotype 7018. On the other hand, genotype 7017 and others had a considerable weight of the pulp. After removing their pulps, they lost almost 20% of their weight. The cocoa pulp can produce juices, soft drinks, citric acid, vinegar, cocoa jelly, and in the alcohol industry (Oddoye; Agyente-Badu; Gyedu-Akoto, 2013; Indiarto et al., 2021). Thus, it is possible to direct the production of these materials to increase the importance of this culture, creating additional values and new opportunities.

Concerning the evaluated characters, the ones that most contributed to the genetic diversity based on the canonical variables were TFW, CPS, and FLP. Thus, these characteristics have a greater influence on the distribution of genetic variation and hence are indispensable in characterizing the *T. cacao* genetic resources. Nevertheless, the descriptors LW, TNSF, and ET contributed the least to the distribution of genetic variation. They may be suppressed in subsequent phenotypic analyses, except when genetic improvement aims to increase values in these specific descriptors.

Considering the leaf analyses, this organ presents dimorphic characters corresponding to the different types of stems on which they appear. However, the highest mean values were observed in group 2. Non-shaded leaves grow more intensely than shaded ones, probably due to greater water stress and higher ambient temperature. Therefore, shaded leaves are longer and greener than those growing in full sun (Wood; Lass, 1985).

When selecting plants for production, using quantitative and qualitative characteristics is recommended. In addition to plants with high fruit weight values, genotypes from groups 4 and 5 had interesting values for fruit diameter, fruit length, longitudinal and transversal perimeters, total weight of seeds, total number of seeds, and width of seeds. These characteristics are essential for plant selection and improvement in terms of bean yield. The morpho-agronomic characterization of these parameters during this study is relevant and favors the mass selection of cacao genotypes.

Qualitative characters varied for fruit color and basal constriction in most groups, as described in Table 8. We observed that 96.07% of the genotypes evaluated had yellow fruits. Only genotypes in groups 1 and 2 contained red fruits and a large pod shell apex shape variation. Furthermore, 100% of the groups had the three types of apexes. The intermediate shape was predominant (58.82% of genotypes), followed by the rounded shape (21.56%) and pointed shape (19.62%). Rough characters also varied between groups, except for group 4, which did not show genotypes with this characteristic. However, concerning the presence and absence of the basal constriction of the fruit, group 4 contained more genotypes lacking basal constriction, whereas group 2 had more genotypes with basal constriction. For the leaf format, groups 1, 2, and 5 did not have obtuse format leaves, but great variability in leaves with different formats was observed in the other grups.

**Table 8:** Absolute frequency of qualitative variables in each of the five groups formed by the Ward-MLM method evaluating 51 genotypes of *Theobroma cacao* grown in 'cabruca' agroforests in the Southern of Bahia, Brazil.

| Qualitative descriptors | 1(12) | 2(10) | 3 (9) | 4 (12) | 5(8) |
|---|---|---|---|---|---|
| Ripe fruit color | | | | | |
| Yellow | 11 | 9 | 9 | 12 | 8 |
| Red | 1 | 1 | --- | --- | --- |
| Shape of the pod shell apex | | | | | |
| Rounded | 4 | 1 | 2 | 3 | 1 |
| Intermediary | 7 | 8 | 2 | 7 | 6 |
| Pointed | 1 | 1 | 5 | 2 | 1 |
| Roughness | | | | | |
| Smooth | 3 | 5 | 4 | 3 | 5 |
| Intermediary | 7 | 4 | 4 | 9 | 2 |
| Rough | 2 | 1 | 1 | --- | 1 |
| Presence of basal constriction | | | | | |
| Absent | 4 | 2 | 4 | 7 | 2 |
| Present | 8 | 8 | 5 | 5 | 6 |
| Leave format | | | | | |
| Short acuminate | 5 | 2 | 2 | 5 | 1 |
| Long acuminate | 5 | 5 | 5 | 4 | 4 |
| Obtuse | --- | --- | 1 | 1 | --- |
| Acute | 2 | 3 | 1 | 2 | 3 |

This study evaluated the intensity of vegetative growth by measuring seedling height and the trunk's diameter (Figure 2). The factorial ANOVA indicated a significant difference among the cultivars in relation to growth, among the fortnights of evaluation, and the interaction between these factors (Table 9). The seedling growth was highly significant, indicating that growth is influenced in relation to time, even in the initial periods of germination and seedling development. CCN51 (278.20 mm) and 2079 (276.80 mm) plants showed the highest height values in the first 75 days of evaluation. Genotypes 3007 (154.76 mm) and 3004 (155.56 mm) showed slower growth. Genotypes 2089 and 2066 had larger trunk diameters, whereas 2001 had smaller. Cultivating cacao as a high-yield crop requires several years of field tests to identify predictive characteristics of high cumulative yields to accept yield efficiency in terms of plant attractiveness, tree vigor, and planting density (Olufemi et al., 2020).

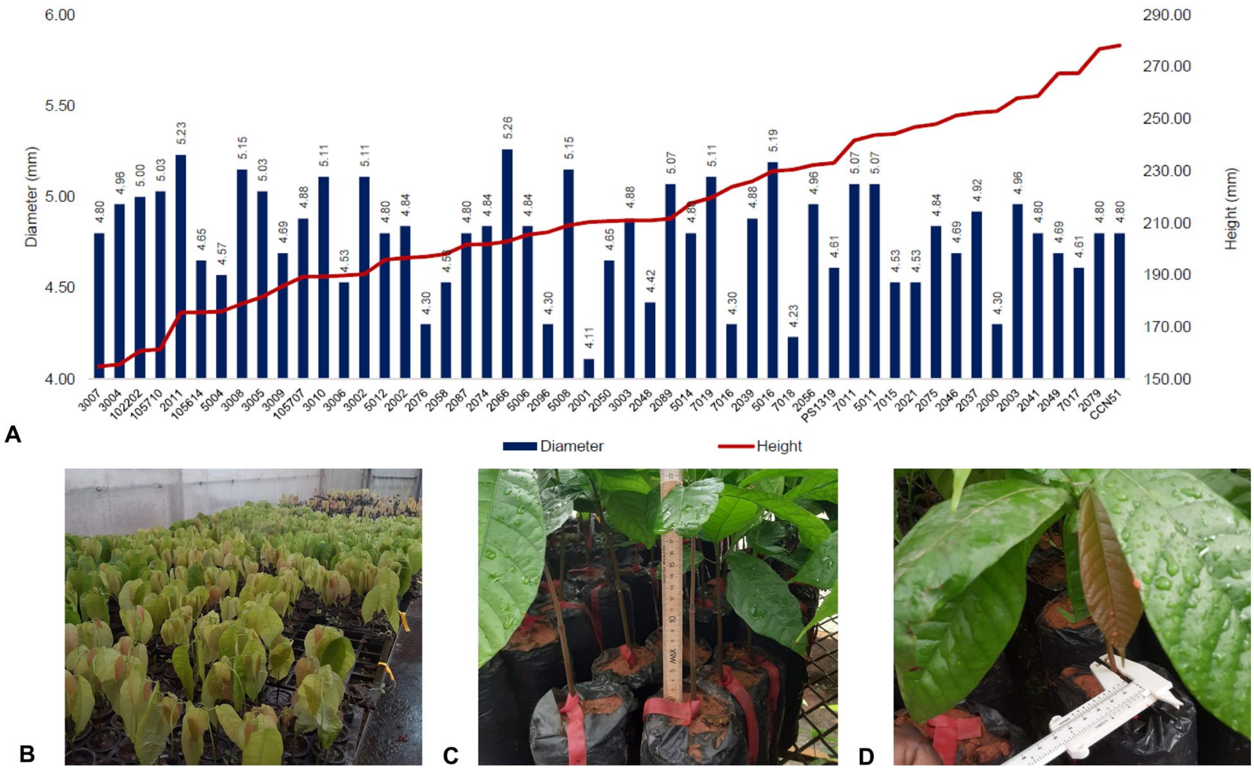


**Figure 2**: Growth and development of 51 genotypes of *Theobroma cacao* grown in the greenhouse. (A) Evaluation of the seedling height and trunk diameter (n = 26 seedlings per genotype) (B)Seedlings before transplanting (one month old); (C, D) Measurement of seedling height and trunk diameter in the last evaluation(four month old).

**Table 9:** Analysis of variance for plant growth in 51 genotypes of *Theobroma cacao* grown in 'cabruca' agroforests in the Southern of Bahia, Brazil.

| | | Mean Square | |
|---|---|---|---|
| Source of variation | Degree of freedom | Seedling height | Stem diameters |
| Genotypes | 50 | 136671.30* | 8.144* |
| Age of grow | 4 | 1574936.58* | 1425.63* |
| Age of grow * Genotypes | 200 | 2437.44* | 1.82* |
| Error | 6552 | 1428.58 | 0.34 |
| Total corrected | 6806 | 6629 | 6629 |
| Cv (%) | | 17.63 | 16.50 |
| Mean | | 214.37 | 3.55 |

***Significant at 5% by F-test.

## CONCLUSIONS

The 51 cacao genotypes studied have genetic variability considering the morpho-agronomic variables for fruits, leaves, and seedling growth. Due to their good performance in almost all variables, the most promising genotypes for breeding programs belong to group 5. This research identified genotypes 2074, 2041, and 2058 promising to improve regional cocoa production.

## AUTHOR'S CONTRIBUTIONS:

Conceptual idea: Georges, M.E; Corrêa, R.X.; Melo, C.A.F.; Souza, M.M; Methodology Design: Georges, M.E; Corrêa, R.X.; Data collection: Georges, M.E.; Melo, C.A.F.; Corrêa, R.X.; Data analysis and interpretation: Georges, M.E.; Melo, C.A.F.; Corrêa, R.X.; Souza, M.M; Writing and editing: Georges, M.E.; Melo, C.A.F.; Corrêa, R.X.; Souza, M.M.

## ACKNOWLEDGEMENTS

The authors are grateful for the logistical support and permission to collect biological material on a farm located in the municipality of Arataca, BA, especially technicians Reginaldo Barbosa Gomes and Idelbrandando de Jesus Fernandes. MEG would like to thank the Brazil PAEC-OEA-GCUB, UESC and CAPES scholarship program for the travel assistance and the master's scholarship at Programa de Pós-Graduação em Genética e Biologia Molecular in UESC. RXC thanks CNPq (process 308959/2019-1) for the scientific productivity grant.

## REFERENCES

ANGIOSPERM PHYLOGENY GROUP - APG. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. Botanical Journal of the Linnean Society, 181(1):1-20, 2006.

CAOBISCO/ECA/FCC. Fèves de cacao: Exigences de qualité de l'industrie du chocolat et du cacao, Royaume-Uni. 2015. 106p. Available in: <https://cocoaquality.eu/data/Feves%20de%20cacao%20Exigences%20de%20qualite%20de%20l%20industrie%20Apr%202016_Fr.pdf>. Access in. September 22, 2023.

CHEESMAN, E. E. Notes on the nomenclature, classification, and possible relationships of cacao populations. Tropical Agriculture, 21:144-159, 1944.

CUATRECASAS, J. Cacao and its allies: A taxonomic revision of the genus *Theobroma*. Smithsonian Institution, 35(6):394, 1964.

ENGELS, J. M. M. The systematic description of cacao clones and its significance for taxonomy and plant breeding. Wageningen University and Research, 1986. 125p.

ENGELS, J. M. M.; BARTLEY, B. G. D.; ENRÍQUEZ, G. A. Cacao descriptors, their states, and modus operandi. Centro Agronomico de Investigacion y Enseñanza, CATIE, p.1-26, 1980.

FERREIRA, D. F. Sisvar: A computer statistical analysis system. Ciência e Agrotecnologia, 35(6):1039-1042, 2011.

GOUDE, K. A.; ADINGRA, K. M.; GBOTOGNON, O. J. et al. Biochemical characterization, nutritional and antioxidant potentials of cocoa placenta (*Theobroma cacao* L.). Annals Food Science and Technology, 20(3):603-613, 2019.

GOWER, J. C. A general coefficient of similarity and some of its properties. Biometrics, 27(4):857-871, 1971.

INDIARTO, R. et al. A review of innovation in cocoa bean processing by-products. International Journal of Emerging Trends in Engineering Research, 9(8):1162-1169, 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Levantamento sistemático da produção agrícola. 2022. Available in: <https://sidra.ibge.gov.br/home/lspa/brasil>. Access in. September 22, 2023.

INTERNATIONAL COCOA ORGANIZATION - ICCO. Quartely bulletin of cocoa statistics, XLVIII, (4), 2021/2022. Available in: <https://www.icco.org>. Access in. September 22, 2023.

LEMBANG, R. P. et al. Identification and analysis of relationship several local promising clones cocoa (*Theobroma cacao* L.) based on morphological characters in South Sulawesi. IOP Conference Series: Earth and Environmental Science, 270:012024, 2019.

MARSSARO, A. L. et al. Resistance to witches' broom in adult plants and progeny of local varieties of cacao in Southern Bahia. Bragantia, 79(4):546-557, 2020.

MELO, C. A. F. et al. Multivariate analysis of backcross progeny of *Passiflora* L. (Passifloraceae) for pre-breeding genotype selection. Genetics and Molecular Research, 14(4):15376-15389, 2015.

MOTAMAYOR, J. C. et al. Cacao domestication I: The origin of the cacao cultivated by the Mayas. Heredity, 89:380-386, 2002.

MOTAMAYOR, J. C. et al. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). PloS one, 3(10):e3311, 2008.

MUSTIGA, G. M. et al. Phenotypic description of *Theobroma cacao* L. for yield and vigor traits from 34 hybrid families in Costa Rica based on the genetic basis of the parental population. Frontiers in Plant Science, 9:808, 2018.

ODDOYE, E. O. K.; AGYENTE-BADU, C. K.; GYEDU-AKOTO, E. Cocoa and Its by-products: Identification and utilization. In: WATSON, R.; PREEDY, V.; ZIBADI, S. (eds). Chocolate in health and nutrition. Nutrition and Health, vol 7. Humana Press, Totowa, NJ. p.23-37, 2013.

OETTERER, M. Fundamentos de ciência e tecnologia de alimentos. Barueri/SP: Editora Manole Ltda, 2006. 632p.

OFORI,P.; AKOTO,O. Production and characterisation of briquettes from carbonised cocoa pod husk and sawdust. Open Access Library Journal, 7(02):e6029, 2020.

OLUFEMI, A. K. et al. Effect of high density planting on the vigour and yield of *Theobroma cacao* L. in the Southwest of Nigeria. World Journal of Advanced Research and Reviews, 8(1):217-223, 2020.

OZTÜRK, G.; YOUNG, G. M. Food evolution: The impact of society and science on the fermentation of cocoa beans. Comprehensive Reviews in Food Science and Food Safety, 3:431-45, 2017.

PINTO, L. R. M.; PIRES, J. L. Selecao de plantas de cacau resistentes a vassoura de bruxa. Ilhéus: CEPLAC, 1998. 34p. (CEPLAC. Boletim Técnico, 181).

PRAZERES, I.; LUCAS, M. R.; MARTA-COSTA, A. Cocoa markets and value chain: Dynamics and challenges for São Tomé and príncipe organic smallholders. International Journal of Innovation and Economic Development, 7(2):64-76, 2021.

RAMÍREZ-GUILLERMO, M. Á. et al. Morphological variation of cacao (*Theobroma cacao* L.) fruits and seeds from plantations in Tabasco, Mexico. Revista Fitotecnia Mexicana, 41(2):117-125, 2018.

REGES, B. M. et al. Quality variables for technological application of cocoa clones from the Brazilian semiarid region. Revista Brasileira de Engenharia Agrícola e Ambiental, 25(3):203-208, 2021.

RESTREPO, T. I.; URREGO, J. E. Protocolo para la caracterización morfológica de árboles élite de cacao (*Theobroma cacao* L.). Medellín, Colombia: Compañía Nacional de Chocolates S.A.S. 2018. 28p.

SAS INSTITUTE. Institute Inc.SAS/STAT® 9.1 User's Guide. Cary, NC: SAS Institute Inc, 2004. 83p. Available in: <https://www.support.sas.com/documentation/onlinedoc/91pdf/sasdoc_91/stat_ug_7313.pdf>Access in: October 2, 2023.

SCHEUER, P. M. et al. Quality properties of a bread made with levain and cocoa waste. Journal of Culinary Science & Technology, 20(5):409-420, 2022.

SOMARRIBA, E.; LÓPEZ SAMPSON, A. Coffee and cocoa agroforestry systems: Pathways to deforestation, reforestation, and tree cover change. CATIE Turrialba, Costa Rica: Background Paper, 2018. 51p.

WHITLOCK, B. A.; BAYER, C.; BAUM, D. A. Phylogenetic relationships and floral evolution of the Byttnerioideae ("*Sterculiaceae*" or *Malvaceae* sl) based on sequences of the chloroplast gene, *ndhF*. Systematic Botany, 26(2):420-437, 2001.

WOOD, G. A. R.; LASS, R. A. Cocoa. United States,Wiley: Blackwell, p.13-14, 1985.

ZUIDEMA, P. A. et al. A physiological production model for cocoa (*Theobroma cacao*): Model presentation, validation and application. Agricultural Systems, 84(2):195-225, 2005.