





Phenotypic and genotypic characterization of cacao cultivars (*Theobroma cacao* L.) from Dibulla, La Guajira, Colombia

 Angélica Ramos Ospino¹,  Margarita Gómez Alvarez¹,  Elwi Machado-Sierra²,  Yani Aranguren^{2*}

¹ Universidad Libre Seccional Barranquilla. Barranquilla, Colombia

² Universidad Simón Bolívar. Barranquilla, Colombia

* Corresponding author: Universidad Simón Bolívar. Facultad de Ciencias Básicas y Biomédicas. Laboratorio de Investigación en Microbiología. Headquarters Carrera 59 No. 59-65. Barranquilla, Colombia. yani.aranguren@unisimonbolivar.edu.co

Received: August 13, 2019

Accepted: March 30, 2020

Published: September 04, 2020

Subject editor: Andrés Cortes (Corporación Colombiana de Investigación Agropecuaria [AGROSAVIA])

How to cite this article: Ramos Ospino, A., Gómez Alvarez, M., Machado-Sierra, E., & Aranguren, Y. (2020). Phenotypic and genotypic characterization of cacao cultivars (*Theobroma cacao* L.) from Dibulla, La Guajira, Colombia. *Ciencia y Tecnología Agropecuaria*, 21(3), e1557. https://doi.org/10.21930/rcta.vol21_num3_art:1557

Abstract

In Sierra Nevada de Santa Marta, cacao plantations are comprised of commercial hybrid cultivars, and although native cacaos are found, they are not widely cultivated. Given the need to verify if these cacao varieties found in the Sierra region belong to the Criollo type genetic group, a phenotypic and genotypic characterization of cacao from the municipality Dibulla, La Guajira was carried out. For this, 11 cultivars were sampled in Mingueo. Phenotypic traits were evaluated using UPOV descriptors for cacao. The qualitative and quantitative parameters were compared through cluster and principal component analyses (PCA), and the quantitative variables through the non-parametric Mann-Whitney test. Molecular biology protocols were standardized, and the ITS region was sequenced to assess genetic relationships. From the sequences, groupings were carried out utilizing distance and phylogenetic methods. Finally, significant differences were found among the seeds ($p = 0.01$), and the white coloration of the cotyledon of the criollos or native stands out in contrast to the dark purple coloration of the hybrids. The cluster analysis, PCA, and sequence analysis groupings, showed differences between the group of native cacaos and commercial hybrids cultivated; in addition, native cacaos are related to the Criollo type group.

Keywords: Criollo cacao, genotype, germplasm, phenotype, UPOV descriptors

Caracterización fenotípica y genotípica de cultivares de cacao (*Theobroma cacao* L.) de Dibulla, La Guajira, Colombia

Resumen

En la Sierra Nevada de Santa Marta los cultivos de cacao están conformados mayoritariamente por cultivares híbridos comerciales y, aunque se encuentran cacaos nativos, estos son poco cultivados. Dada la necesidad de verificar si estos cultivares de cacao encontrados en la Sierra pertenecen al grupo genético tipo Criollo, se realizó una caracterización fenotípica y genotípica de cacaos del municipio Dibulla, La Guajira. Para esto, se muestrearon 11 cultivares en Mingueo. Los rasgos fenotípicos se evaluaron empleando descriptores UPOV para cacao. Los parámetros cualitativos y cuantitativos se cotejaron por análisis de conglomerado y análisis de componentes principales (ACP), y las variables cuantitativas se compararon a través de la prueba no paramétrica test de Mann-Whitney. Para evaluar las relaciones genéticas, se estandarizaron protocolos de biología molecular y se secuenció la región ITS. A partir de las secuencias, se realizaron agrupamientos por métodos de distancia y filogenéticos. Finalmente, se encontraron diferencias significativas entre las semillas ($p = 0,01$), y resalta la coloración blanca del cotiledón de los criollos en contraste con la coloración púrpura oscura de los híbridos. Asimismo, los análisis de conglomerados, ACP y los análisis de secuencias demostraron diferencias entre el grupo de los cacaos nativos y los híbridos comerciales cultivados; además, los cacaos nativos se emparentan con el grupo de cacao tipo Criollo.

Palabras clave: cacao Criollo, descriptores UPOV, fenotipo, genotipo, germoplasma

Introduction

Cacao *Theobroma cacao* L. is a tree belonging to the Malvaceae family, native to central and northern humid tropical regions of South America (López-Medina, 2017). According to historical records, the processes of domestication, cultivation, and consumption of cacao were initiated by the Mayans in Mexico and Central America, who consumed it as a drink called *xocoatl*, a precursor of modern chocolates (Motamayor et al., 2002; Young, 2008). However, genomic analyzes indicate that it is native from the Amazon basin in northwestern South America (Zarrillo et al., 2018).

Currently, cacao is grown in hot and rainy climates in the Americas, Africa and Asia (Bhattacharjee & Kumar, 2007; Rodríguez-Medina et al., 2019) in agroforestry systems that allow it to be combined with other crops and native species, in balance with the environment (Jagoret et al., 2011; Navarro-Prado & Mendoza-Alonso, 2006; Roa-Romero et al., 2009; Romero & Urrego, 2016). Cocoa is of great economic importance due to its versatility of uses in artisanal confectionery, the cosmetic industry, and the food industry, as is the case of the production of chocolates, oils, and liquors, respectively (Bennett, 2003). In Colombia, there has been an increase in cacao cultivation in recent years, and by 2017, it ranked as the seventh largest producer worldwide (Food and Agriculture Organization [FAO], 2017). This scenario makes cocoa one of the highest sustainable bets of the national and international market, from the environmental, social and economic point of view; this is because it allows satisfying all the links in the productive chain, guaranteeing a prosperous economy, mainly in post-conflict regions, promoting peace and the substitution of illicit crops (Rodríguez-Medina et al., 2019).

Three types of cacao are known: Criollo, originating in South America and Central America; Forastero, which comes from the Amazon basin; and Trinitario, which emerged in Trinidad & Tobago as a hybrid of the first two types (De La Cruz-Medina et al., 2012). The most relevant organoleptic characteristic of Criollo cacao is its bitter, acidic, and fruity taste, with white cotyledons (Andrade-Aguirre & Angulo-Reynoso, 2007). For its part, Forastero is the most widely cultivated cacao type, estimating that it covers more or less 85% of the world production because it is more resistant to diseases and pests; further, its flavor is strong, bitter and slightly acidic (Romero & Urrego, 2016). Trinitario is a more resistant and productive type of cacao, but of lower quality compared to Criollo (Andrade-Aguirre & Angulo-Reynoso, 2007).

Until the end of the 19th century, Criollo cacao was the most cultivated in Colombia, but the presence of diseases decimated this crop that was replaced by Forastero (Rodríguez-Medina et al., 2019). From that moment, native materials have been recovered and characterized, as well as conserved in germplasm banks; however, currently, the commercial cultivation of the Criollo type is restricted (Aranzazu et al., 2009; Oicatá, 1986; Perea et al., 2013; Rodríguez-Medina et al., 2019).

Plant varieties contain differential physiological, morphological, and genotypic characteristics; hence, the importance of being characterized, establishing the particular traits of an individual or population, and of developing plant improvement programs (Aranguren et al., 2010). Reliable morphological variables must be used to carry out this characterization, thus, allowing the differentiation between groups. These variables are established in the *Technical Guidelines for Varietal Description*, such as those issued by the International Union for the Protection of New Varieties of Plants (UPOV, 2011). The characterization of the germplasm is accompanied by the determination of the genetic traits that allow the measurement of genetic variability (Aranguren et al., 2010; Núñez-Colín & Espodedo-López, 2014). For this, molecular markers are used, which are DNA fragments that, alone or in combination, can be used to determine genetic diversity through the detection of their polymorphisms (Azofeifa-Delgado, 2006; Rocha, 2003).

Simple nucleotide polymorphisms (or SNP) are widely used to genotype and perform assisted selection in plant breeding from the sequencing and analysis of specific regions of the genome (Chagné et al., 2007; De Wever et al., 2019; Poland & Rife, 2012). ITS regions are ribosomal DNA sequences, easy to amplify and align, and used in the study of intra- and inter-population relationships since they show evolutionary differences and allow obtaining phylogenetic and taxonomic information on individuals (Avendaño-Sánchez et al., 2015; Quijada et al., 2017).

In this study, samples of cacao fruits cultivated in the town of Mingueo, department of La Guajira, were taken where commercial cacao cultivars and native cacao plants collected in Sierra Nevada are planted, and identified by farmers as native Criollo-type cacaos. However, there is no evidence that these are of the Criollo type, as this germplasm has not been agronomically characterized. Therefore, all these cultivars were phenotypically characterized using UPOV descriptors, and a genotypic evaluation was carried out through ITS sequences, to identify kinship relationships between native cacaos of the region and the commercial ones.

Materials and methods

The study was carried out from the materials grown in the cacao germplasm collection in the farm Brisas del Mar, of the Association of Organic Producers of the Municipality of Dibulla (APOMD, for its acronym in Spanish), located at 48 m above the sea level in the district (Corregimiento) of Mingueo, Dibulla, in the department of La Guajira. The farm is located in a tropical forest area, 10 km northeast of the town of Mingueo. Ten samples of cacao plants were collected, identified by the producers as four native cacaos, five commercial clones, and an unknown hybrid (table 1). Additionally, a sample of Criollo type cacao cultivated in Becerril, Cesar, was obtained and used as a reference. The sampling was of a simple random type, to collect information and samples. The morphological data of the cacao plants were taken *in situ*, according to the criteria established for *T. cacao* by UPOV (2011). Besides, young leaves were collected for genetic analysis.

Table 1. Cacao samples collected

Sample code	Type of cacao	Georeference	
001	Hybrid ICS 39	11°09'46"N	73°22'32"W
002	Hybrid ICS 95	11°09'44"N	73°22'33"W
003	Hybrid ICS 39	11°09'46"N	73°22'32"W
004	Hybrid ICS 60	11°09'45"N	73°22'33"W
005	Hybrid CCN 51	11°09'45"N	73°22'34"W
006	Criollo	11°09'39"N	73°22'29"W
008	Criollo de Becerril	9°42'10"N	73°16'42"W
009	Unknown hybrid	11°10'14"N	73°20'49"W
010	Criollo	11°10'12"N	73°21'04"W
011	Criollo	11°10'14"N	73°20'56"W
012	Criollo	11°10'14"N	73°20'58"W

Source: Elaborated by the authors

Leaf characterization was carried out, taking five samples per plant to measure each variable. Leaves with good phytosanitary status were selected, located in the fifth node of branches situated in the tree at the breast height, in the fruiting period. The length and width of the leaf blade, as well as some qualitative features, were measured: shape of the base, shape of the apex, and intensity of the green color. Likewise, physiologically mature fruits without disease symptoms were collected, and the length, diameter, thickness of the epicarp, the concentration of soluble solids in the pulp, and pH were measured. Similarly, the following qualitative characteristics were assessed: pod shape, apex shape, base shape, pod color, surface, depth of ridges, and color of the pulp. Besides, the quantity of all the whole seeds per fruit was counted. Five seeds were taken from each fruit, and length, width, and thickness were evaluated, as well as cotyledon color and shape in its longitudinal section. All fruit and seed measurements were made with a digital vernier caliper. In the concentration of soluble solids, a refractometer was used, and the pH was estimated with a portable pH meter. In the identification of qualitative features such as shapes and colors, UPOV's descriptors (UPOV, 2011) and the *Catalog of cacao cultivars* (García, 2009) were used.

The data collected in the phenotypic characterization were tabulated in Microsoft Excel™ and later analyzed with statistical tools. The quantitative variables of commercial hybrid and native cacaos were compared using the non-parametric U or Mann-Whitney test. Additionally, cluster analyzes were performed using the PC-ORD version 5.0 program. For this, matrices were constructed calculating the Euclidean distances and the groupings employing Ward's method. The quantitative data were normalized before calculating the distances, and dendrograms were constructed. Subsequently, multivariate statistics (Principal Component Analysis [PCA]) were applied using all the evaluated traits.

For the genotypic characterization, young leaves of each cacao plant were used and preserved in tubes with silica gel. Then, DNA was extracted following the protocol of Lodhi et al. (1994) with modifications. DNA purity was quantified and determined by spectrophotometry using a NanoDrop ND-2000. Using the available primers in the bank of primers of the Microbiology Research Laboratory of Universidad

Simón Bolívar, an *in silico* analysis was carried out on the different sequenced genomes of *T. cacao*, using the BLAST tool (Ye et al., 2012) and the Clustal W algorithm, from the Bioedit Sequence Alignment Editor program version 7.0.5.3.

Then, the primers were tested by PCR, and the most specific pair with the best amplification performance was selected. The selected primers are 86F 5'-GTGAATCATCGAATCTTTGAA-3 and 4R 5'-TCCTCCGCTTATTGATATGC-3' (White et al., 1990). Subsequently, the amplification reaction was optimized by adding DMSO (Miranda et al., 2010). The PCR products were purified and sequenced by the Sanger method, through the Macrogen company in South Korea. Once the sequences were obtained, the chromatograms were analyzed to verify their quality, and they were edited by eliminating the low-quality ends employing the BioEdit program. Then, a Blast sequence was carried out for each sequence (Ye et al., 2012) to verify its identity and rule out contamination.

Finally, dendrograms were constructed by distance (UPGMA and neighbor-joining) and phylogenetic methods (maximum likelihood and maximum parsimony) using the Mega X program (Kumar et al., 2018), to establish the kinship relationships between the ITS sequences of each cultivar. The best evolutionary model was calculated and, from the phylogenetic tree obtained by the maximum likelihood method under the parameters of the Tamura-3 evolutionary model and a Bootstrap of 1,500 repetitions (Kumar et al., 2018), a consensus tree was built to validate the results obtained by the other grouping methods. Furthermore, the cacao ITS sequence (JQ228376) reported in GenBank was used as an outgroup.

On the other hand, an *in silico* search was made for other markers that could be used in future work on the discrimination between cacao cultivars. For this, a search was done for the *matK* and *rbcL* genes, as well as the intergenic spacers *psbA-trnH* and *rpl32-trnL* in the cacao genomes present in the databases. From these, analyzes were made utilizing phylogenetic methods.

Results and discussion

Phenotypic characterization

The shape of the leaves was uniform within each individual, and each type of cacao evaluated. The length of the leaf varied between 271.2 and 392.6 mm, while the width had values between 90.0 and 132.6 mm. In total, two forms of the base and two of the apex were observed. Similarly, in the intensity of the leaf color, two variants were found. The leaf sample recorded from Becerril, identified with code 008, had the smallest size with 287.0 mm in length and 90.6 mm in width. Moreover, most of the leaves of the evaluated cultivars have an obtuse base, acuminate apex, with medium color intensity. Although the size of the leaf is not used to differentiate cacao cultivars, the differences observed in this study, especially in the width of the blade, reflect that there is a degree of variability between them.

Likewise, the shape of the fruits was homogeneous within each individual evaluated. The length in all cultivars varied between 110 and 277 mm; the diameter values ranged between 55 and 200 mm, and the thickness of the epicarp between 4.35 and 17.97 mm. Diversity was observed in fruit shapes, and the oblong shape was the most frequent (figure 1). In most cultivars, the basal fruit constriction was absent

or very weak; two apex forms were observed, and the predominant fruit surface was moderately rough. The fruit color varied, with yellow being the most common. The pulp colors recorded were white and light cream (table 2). This diversity of forms between cultivars shows the variability that exists between them, becoming traits that allow their differentiation and identification.



Figure 1. *Theobroma cacao* fruits from cultivars planted in the district of Mingueo. a. Criolo-12; b. Criolo-11; c. Criolo-6; d. CCN51-5; e. ICS60-4; f. ICS39-3; g. ICS95-2; h. ICS39-1.
Source: Elaborated by the authors

Concerning the number of seeds per fruit, the CCN51 clone is the one with the highest number of seeds, followed by Criollo de Becerril (008) with 42 seeds; the cultivar with the least amount of seeds was hybrid ICS95 (table 2). The average number of seeds was 30, with a length value that ranged between 29.76 and 21.53 mm, and a width value between 15.17 and 10.59 mm. The thickness of the seeds ranged from 7.9 to 12.2 mm. In this sense, the seed size of the native cacaos found in Mingueo was larger than that of the hybrids. Most of the seeds were oval or oblong, and the predominant cotyledon color was dark purple, followed by white. All the cacao samples identified as native or creole have white cotyledons (table 2). The morphological variants of the observed seeds are typical of each cultivar and are reflected in the productive and organoleptic characteristics of these cacaos. In this way, the coloration of creole cacaos stands out in relation to commercial hybrids (García, 2009).

Table 2. Outstanding phenotypic characteristics of cacao cultivars found in Dibulla, La Guajira

Sample	Blade width (mm)	Fruit diameter (mm)	Seed length (mm)	Seed width (mm)	Seed thickness (mm)	Brix degrees	Fruit shape	Fruit color	Pulp color	Seed shape	Cotyledon color
ICS39-1	126.8 ±9.2	92.0±14.1	27.78±2.35	13.54±0.53	10.50±0.85	21.5±0.5	Oblong	Greenish yellow	Light cream	Oblong	Dark purple
ICS95-2	123.2 ±7.4	82.0±18.3	23.36±1.08	14.07±0.65	10.47±0.80	19.0±0.1	Oblong	Medium red	White	Elliptic	Dark purple
ICS39-3	123.6 ±8.2	81.0±15.5	29.71±1.96	13.08±0.43	9.30±0.17	19.5±0.1	Oblong	Orange	Light cream	Oblong	Dark purple
ICS60-4	111.4 ±8.6	78.0±18.2	23.58±1.88	13.20±1.85	8.86±1.05	21.5±0.2	Elliptic	Greenish yellow	White	Oval	Dark purple
CCN51-5	121.4 ±25.3	96.5±19.8	24.72±1.07	13.59±1.06	8.80±1.09	16.2±0.5	Oblong	Dark red	Light cream	Oblong	Light purple
CRIO-6	116.8 ±14.9	60.0±17.2	21.53±1.16	14.35±1.16	9.34±1.81	16.0±0.1	Oblong	Yellow	Light cream	Oval	Dark purple
CRIO-8	90.6 ±12.2	73.8±6.4	22.04±1.71	10.59±0.29	7.90±0.44	16.7±0.0	Obovate	Yellow	Light cream	Oblong	Brown
HIB-9	132.6 ±13.7	200.1±29.1	29.76±1.17	14.76±1.31	10.60±1.63	13.5±0.6	Oblong	Orange	White	Oblong	Dark purple
CRIO-10	131.6 ±21.3	73.0±11.1	23.94±2.14	15.17±0.59	10.85±1.12	18.2±0.2	Obovate	Yellow	White	Oval	White
CRIO-11	129.0 ±24.6	71.0±13.7	23.36±1.44	13.12±1.50	10.15±0.96	15.3±0.1	Oblong	Yellow	White	Oval	White
CRIO-12	122.4 ±7.3	55.0±9.1	23.08±1.70	15.01±0.77	12.20±1.06	16.0±0.1	Oblong	Orange	Light cream	Oval	White

Source: Elaborated by the authors

The cacao cultivars studied were grouped into native and non-native cacaos, and the quantitative traits were compared between the two groups. According to the Mann-Whitney test, the null hypothesis H_0 is accepted for the variables *fruit length*, *blade length*, *epicarp thickness*, *number of seeds*, and *pH*, that is, there are no significant differences between the two cacao groups ($p = 0.01$). On the other hand, when comparing the variables *blade width*, *fruit diameter*, *°Brix* and *seed length*, *width and thickness*, the null hypothesis is rejected, noting that there are significant differences with 99 % confidence ($p = 0.01$) between creole or native and hybrid cacaos for these variables.

The cluster analysis using quantitative phenotypic characteristics through Ward's method forms two large clades (figure 2): in one clade, the creole cultivars were ordered, including the Criollo cacao variety used as a control, and in the other, the hybrid cultivars.

The PCA was performed by comparing the qualitative and quantitative characteristics (figure 3). The dispersion of the samples indicates that the parameters with the most significant influence on the variability are GS (seed thickness), CC (cotyledon thickness), LF (fruit length), CF (fruit color), FF (fruit shape), PLF (depth between ridges), LS (seed length), and FLSL (longitudinal shape of the seed). Furthermore, the segregation of creole cacaos from commercial cultivars is observed (figure 3).

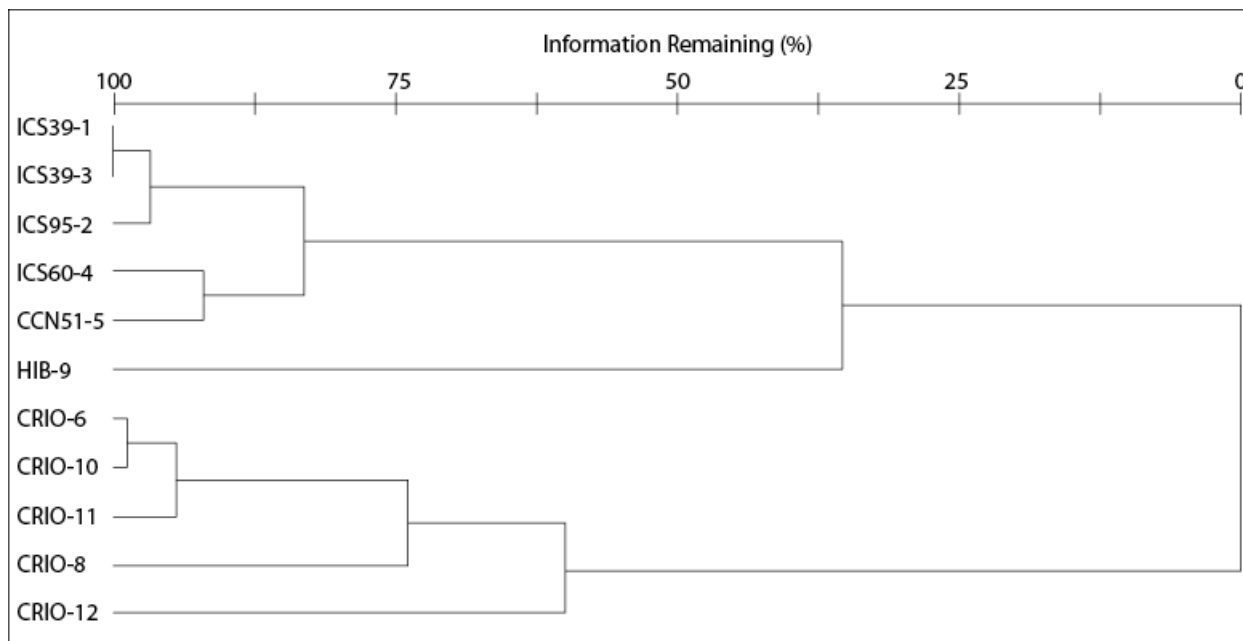


Figure 2. Dendrogram of cacao cultivars evaluated from quantitative phenotypic traits. Clustering through Ward's method and Euclidean distance.

Source: Elaborated by the authors with PCORD

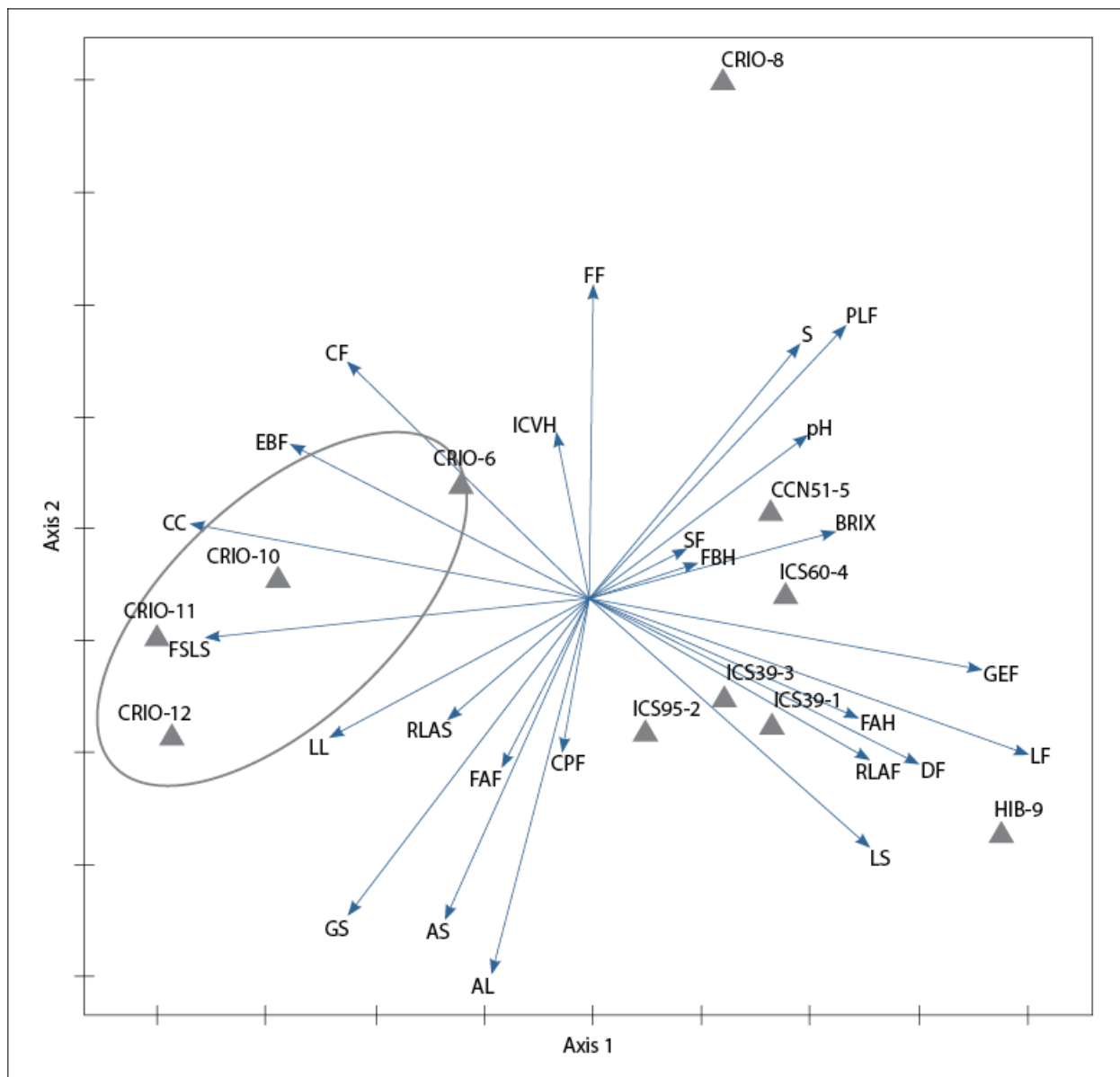


Figure 3. PCA multivariate analysis from quantitative and qualitative phenotypic characteristics. Source: Elaborated by the authors with PCORD

Genotypic analysis

Initially, some DNA extractions were carried out; however, the performance and quality were very poor. Within this procedure, the first organic extractions did not separate well the nucleic acids from the cellular residues since, at the end of the procedure, large amounts of a white pellet were observed, and it was not possible to amplify the ITS sequence from these DNAs. For this reason, several adjustments and tests were carried out using different concentrations of PVP, without obtaining significant changes. Then, the concentration of 2-β-Mercaptoethanol was increased to 1%, and the number of organic extractions

with chloroform: IAA up to four times. Thus, better results were obtained, which were verified by spectrophotometric quantification.

One of the factors that affect the extraction of nucleic acids from *T. cacao* leaves, and that interferes with the quality and quantity of DNA, is the high concentration of polyphenols and polysaccharides in the foliar tissue of the plant (Henaó et al., 2017; Martínez et al., 2013; Schrader et al., 2012). Once the adjustments to the extraction protocol were made, the DNA was obtained. According to the 260/280 ratio, these still contained some contaminants. Given the presence of these residues, the concentrations of the sample were adjusted by making dilutions to decrease the inhibitor effect and maintain a final DNA concentration of 10 ng/mL. As some authors have expressed, this type of problem is common when working with this species (Chia-Wong, 2009; Martínez et al., 2013; Ruiz, 2014). Then, the thermal profile of the PCR amplification was adjusted, and all the fragments were evidenced in the electrophoretic runs. The size of these sequences was 380 bp and corresponded to the ITS2 region, located between the 5.8S region and the largest subunit of the ribosome (White et al., 1990). This intergenic region was selected because it allows observing intraspecific variations, as it has less selective pressure than the coding regions (Quijada et al., 2017; Zambrano, 2017).

Once the sequences were obtained, grouping analyzes were performed by distance methods and UPGMA phylogenetic methods, neighbor-joining, maximum likelihood, and maximum parsimony. The nucleotide substitution models for these sequences were evaluated, and the Tamura-3 was identified as the evolutionary model with the best fit (Tamura, 1992). From these parameters, the dendrograms were constructed. In all the methods, a clade was formed in which the creole cacaos were grouped (including the Becerril reference), differing from the commercial hybrid cacao clade. Also, most clades are statistically supported by all methods. Thus, a tree was built using the maximum likelihood method, locating the consensus clades and indicating the probability for a Bootstrap of 1,500 replicas for each method (figure 4).

Even though the ITS used in the current work allowed differentiating the group of creole or native cacaos from the commercial ones, they do not permit us to answer the entire evolutionary history. For this reason, a prospective *in silico* analysis of other markers was carried out, allowing the complementation of this analysis and lay the foundations for future research. Initially, a search of the chloroplast genomes of cacao cultivars and other species of the Malvaceae family deposited in GenBank was made (Benson et al., 2012). The same cacao cultivars were used for all analyzes, and the species of the family varied in each marker. In these genomes, an *in silico* search was carried out, and by employing phylogenetic methods, the markers *matK* and *psbA-trnH* were found to have enough variability to see differences between cacao types. On the other hand, *rbcL* and *rpl32-trnL* do not allow resolving the differences between cacao cultivars.

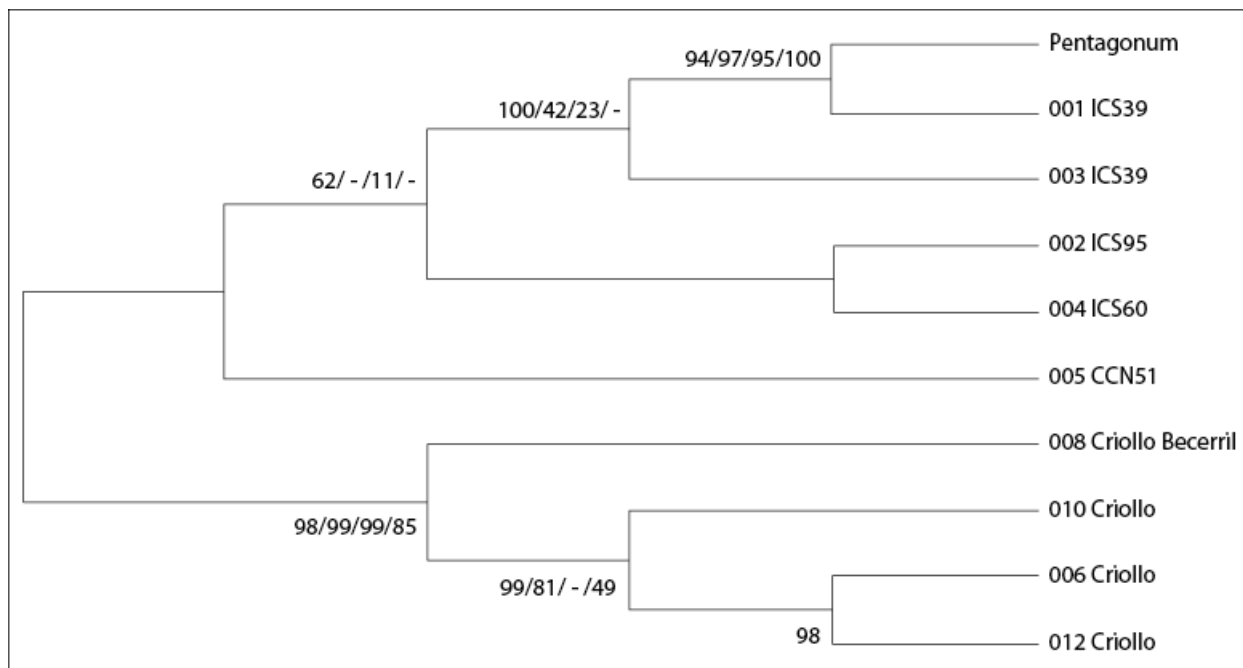


Figure 4. Kinship analysis between cacao cultivars from Dibulla, based on the ITS intergenic region. The phylogenetic tree was constructed by the maximum likelihood method under the parameters of the Tamura-3 evolutionary model and a Bootstrap of 1,500 repetitions. The clades were validated by the maximum likelihood, maximum parsimony, UPGMA, and neighbor-join methods, respectively.

Source: Elaborated by the authors with Mega X

In the area of Sierra Nevada de Santa Marta, the municipality of Dibulla has a biological diversity and environmental characteristics that favor a growing and dynamic agricultural activity. In this sense, it is necessary to carry out training programs for the production of quality cacao crops through the identification, selection, and use of internationally recommended materials and, in turn, promote the rescue of outstanding Criollo-type genotypes. This would allow the conservation and multiplication of germplasm to initiate genetic improvement programs to increase the quality of cacaos, as well as promoting efficient and sustainable cultivation systems over time. This work allows identifying the types of cacaos that are grown in the region to promote their cultivation and boost their final products.

When characterizing and identifying a cultivar, the phenotypic traits are the most distinctive. The phenotype is understood as the set of observable traits, whether morphological, physiological, or behavioral, within a species or population; this depends on the genotype, and can be influenced by environmental and nutritional factors (Botero & Arias, 2018). In this sense, the analyzed individuals were located in the same plot, in a uniform environment, under slightly variable temperature, humidity and radiation conditions, and with the utilization of the same fertilization regime and organic phytosanitary practices; furthermore, they were collected and measured on the same day. Therefore, the observed phenotypic variation should have little environmental influence.

Within the native cacaos found in Dibulla, typical Criollo type cacaos are found, and these are different from commercial cacaos (figures 2, 3 and 4). These are characterized by having leaves whose shape at the

base and apex are obtuse and acuminate, respectively; the length and width of these leaves are uniform. On the other hand, most of the fruits of this cacao have an oblong shape, the surface is moderately rough, and the basal constriction is weak (table 2). Additionally, two fruit shapes were found, one with a notched apex and the other with an acute apex. Regarding the size, the fruits are smaller in length, width and thickness compared to the commercial ones. The seeds are uniform, oval in shape, and moderately elongated, and their size is similar to that of the commercial ones (table 2). As for the commercial cultivars evaluated, they retain most of the morphological characteristics described by García (2009).

Likewise, one of the phenotypic traits used to distinguish creole or native cacaos from other types is the color of the cotyledon (Avendaño et al., 2014; Ventura et al., 2004). The cotyledon color showed a considerable variation that is reflected in the distribution effect of the samples in the PCA (figure 3). In this sense, the commercial cacao cultivars showed dark purple coloration, and the native cacaos called *criollos*, had white cotyledons (table 2).

In the case of Creole 006, the qualitative phenotypic characterization showed that it shares characteristics with some commercial cacaos, such as the color of the cotyledon and the shape of the base of the leaf that is acute, which suggests that this creole cacao is not pure, and that it is related to a hybrid. Likewise, Creole 008 from Becerril has some characteristics similar to clone CCN 51, indicating that it originates from crosses between Forasteros and Criollos (García, 2009). However, few phenotypic characteristics relate these cultivars, and these may be polygenic; therefore, additional genetic analyses should be carried out to determine the degree of kinship. for this reason, it is necessary to carry out population studies to cover a higher number of samples, a wider distribution, and the segregation of traits in the offspring, to examine distinction, homogeneity, and stability.

The ITS used in this work, allowed differentiating the group of creole cacaos from the commercial ones. However, the results do not finish answering the entire evolutionary history (Zambrano, 2017). The small nuclear region chosen does not represent the entire genome and, therefore, does not account for the entire genetic evolution of the species. The analysis of chloroplast DNA markers (cpDNA) has been used increasingly in population genetics to determine their structure, gene flow, haplotype frequency, and phylogenetic relationships (Gutiérrez-López et al., 2016). For this reason, an *in silico* evaluation of other sequence markers of the chloroplast genome was made, which will allow more in-depth and precise analyzes between cacao cultivars in the future. The phylogenetic analyzes carried out on the markers *matK* and *trnH-psbA* could better resolve the differences between cacao cultivars and would allow a better understanding of the origin and kinship relationships of cacaos in the region. Likewise, in future studies, other markers such as microsatellites could be used, which would allow estimating the genetic diversity (Aranguren-Díaz et al., 2018; Lanaud & Risterucci, 1999).

Conclusions

The collection of creole cacaos of Asociación de Productores Orgánicos del Municipio Dibulla is an important plant genetic resource that must be preserved. The phenotypic and genotypic analyses show that these native cacao cultivars have differences compared to commercial cultivars. Furthermore, this native germplasm can be classified within the genetic group of Criollo-type cacaos. However, more studies must be carried out to categorize and certify these.

The current work was a baseline for the knowledge of the creole or native cacao of Sierra Nevada de Santa Marta, which will allow the development of new studies to deepen the denomination and certification of these types of cacao. In this sense, we propose to carry out more studies using a higher number of creole cacao individuals and using *matK* and *trnH-psb* markers, to better establish the relationships between cacao cultivars and intra-population variations.

Finally, the conservation of native cacao cultivars is of vital importance since it contributes to the species variability persistence over time and can favor genetic improvement programs, for which phenotypic and genotypic characterization is essential.

Acknowledgments

The authors thank Asociación de Productores del Municipio Dibulla for its support in carrying out this project.

Disclaimers

The statements, judgments, and opinions expressed in this article correspond to the authors, who state that there are no conflicts of interest.

References

- Andrade-Aguirre, C. M., & Angulo-Reynoso, V. (2007). *La viabilidad económica del cultivo del cacao en México a través de una economía sostenible*. [Undergraduate thesis, Universidad de las Américas Puebla]. Repositorio Colección de tesis digitales. http://catarina.udlap.mx/u_dl_a/tales/documentos/lri/andrade_a_cm/
- Aranguren-Díaz, Y. C., Varani, A. M., Michael, T. P., & Miranda, V. F. O. (2018). Development of microsatellite markers for the carnivorous plant *Genlisea aurea* (Lentibulariaceae) using genomics data of NGS. *Molecular Biology Reports*, 45(1), 57-61. <https://doi.org/10.1007/s11033-017-4140-1>
- Aranguren, Y., Briceño, A., & Fermin, G. (2010). Assessment of the variability of Venezuelan guava landraces by microsatellites. *Acta Horticulturae*, 849, 147-154. <http://doi.org/10.17660/ActaHortic.2010.849.16>
- Aranzazu, H. F., Martínez, G. N., Palencia, C. G., Coronado, R., & Rincón, G. D. (2009). *Manejo del recurso genético para incrementar la producción y productividad del sistema de cacao en Colombia*. Federación Nacional de Cacaoteros-Corporación Colombiana de Investigación Agropecuaria-Ministerio de Agricultura y Desarrollo Rural.
- Avendaño-Sánchez, M., Espinoza-Velázquez, J., Gutiérrez-López, A., Flores-Gallegos, A. C., & Rodríguez-Herrera, R. (2015). Secuencias nucleotídicas de la región ITS en familias S. y PL de maíces poliembriónicos. *Revista Mexicana de Ciencias Agrícolas*, 6(3), 509-521. <http://www.scielo.org.mx/pdf/remexca/v6n3/v6n3a6.pdf>
- Avendaño, C., Cueto, J., Mendoza, A., Lopez, P., Sandoval, A., & Aguirre, J. (2014). *Graphic Handbook for Description of Cocoa Varieties (Theobroma cacao L.)*. Instituto de investigación forestales, agrícolas y pecuarias.
- Azofeifa-Delgado, Á. (2006). Uso de marcadores moleculares en plantas ; aplicaciones en frutales el trópico. *Agronomía Mesoamericana*, 17(2), 221-241. <https://doi.org/10.15517/am.v17i2.5163>
- Bennett, A. B. (2003). Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends in Plant Science*, 8(12), 561-563. <https://doi.org/10.1016/j.tplants.2003.10.004>

- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, 41(D1), D36-D42. <https://doi.org/10.1093/nar/gks1195>
- Bhattacharjee, R., & Kumar, P. L. (2007). Cacao. En C. Kole (Ed.), *Technical Crops* (pp. 127-142). Springer Berlin Heidelberg. <http://doi.org/10.1007/978-3-540-34538-1>
- Botero, K., & Arias, T. (2018). Uso de las ciencias ómicas para el mejoramiento genético de cultivos. *Revista de Ciencias Agrícolas*, 35(2), 64-78. <https://doi.org/10.22267/rcia.183502.92>
- Chagné, D., Batley, J., Edwards, D., & Forster, J. W. (2007). Single Nucleotide Polymorphism Genotyping in Plants. En N. C. Oraguzie, E.H.A. Rikkerink, S. E. Gardiner, H. N. De Silva (Eds.), *Association Mapping in Plants* (pp. 77-94). Springer New York. https://doi.org/10.1007/978-0-387-36011-9_5
- Chia-Wong, J. A. (2009). *Caracterización molecular mediante marcadores ISSR de una colección de 50 árboles clonales e híbridos de cacao (Theobroma cacao L.) de la UNAS-Tingo María*. [Master's thesis, Universidad Nacional Mayor de San Marcos]. <http://cybertesis.unmsm.edu.pe/handle/cybertesis/244>
- De La Cruz-Medina, J., Vargas-Ortiz, M., & Del Ángel-Coronel, O. (2012). *Cacao: operaciones poscosecha*. Food and Agriculture Organization Of the United Nations (FAO). <http://www.fao.org/3/a-au995s.pdf>
- De Wever, J., Everaert, H., Coppeters, F., Rottiers, H., Dewettinck, K., Lefever, S., & Messens, K. (2019). The development of a novel SNP genotyping assay to differentiate cacao clones. *Scientific Reports*, 9(1), 9512. <https://doi.org/10.1038/s41598-019-45884-8>
- Food and Agriculture Organization (FAO). (2017). *FAOSTAT Statistical Database*. <http://www.fao.org/faostat/en/#data/QC>
- García, L. (2009). *Catálogo de cultivares de cacao*. Ministerio de Agricultura.
- Gutiérrez-López, N., Ovando-Medina, I., Salvador-Figueroa, M., Molina-Freaner, F., Avendano-Arrazate, C. H., & Vázquez-Ovando, J. A. (2016). Unique haplotypes of cacao trees as revealed by trnH-psbA chloroplast DNA. *PeerJ*, 4, e1855. <https://doi.org/10.7717/peerj.1855>
- Henao, A., Salazar, H., & Urrea, A. (2017). Quality of cocoa (*Theobroma cacao* L.) DNA from foliar tissue at different stages of development. *Acta Agronomica*, 67(2), 311-318. <https://doi.org/10.15446/acag.v67n2.63046>
- Jagoret, P., Michel-Dounias, I., & Malézieux, E. (2011). Long-term dynamics of cocoa agroforests: a case study in central Cameroon. *Agroforestry Systems*, 81(3), 267-278. <https://doi.org/10.1007/s10457-010-9368-x>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547-1549. <https://doi.org/10.1093/molbev/msy096>
- Lanaud, C., & Risterucci, A. M. (1999). Isolation and characterization of microsatellites in *Theobroma cacao* L. *Molecular Ecology*, 8(12), 2141-2152. <https://doi.org/10.1046/j.1365-294x.1999.00802.x>
- Lodhi, M. A., Ye, G.-N., Weeden, N. F., & Reisch, B. I. (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Reporter*, 12(1), 6-13. <https://doi.org/10.1007/BF02668658>
- López-Medina, S. E. (2017). Características germinativas de semillas de *Theobroma cacao* L. (Malvaceae) "cacao". *Arnaldoa*, 24(2), 609-618. <https://doi.org/10.22497/arnaldoa.242.24212>
- Martínez, A., Leshner, J., & Jiménez, M. (2013). Comparación de tres métodos para la extracción de ARN total a partir de hojas de cacao. *Artículo Original Biotecnología Vegetal*, 13(2), 93-98. <https://revista.ibp.co.cu/index.php/BV/article/view/100/82>
- Miranda, V. F. O. de, Martins, V. G., Furlan, A., & Bacci Jr., M. (2010). Plant or fungal sequences? An alternative optimized PCR protocol to avoid ITS (nrDNA) misamplification. *Brazilian Archives of Biology and Technology*, 53(1), 141-152. <https://doi.org/10.1590/S1516-89132010000100018>

- Motamayor, J. C., Risterucci, A. M., López, P. A., Ortiz, C. F., Moreno, A., & Lanaud, C. (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity*, 89(5), 380-386. <https://doi.org/10.1038/sj.hdy.6800156>
- Navarro-Prado, M., & Mendoza-Alonso, I. (2006). *Cultivo del Cacao en Sistemas Agroforestales*. ProDeSoc.
- Núñez-Colín, C., & Espodedo-López, D. (2014). Caracterización de germoplasma vegetal: la piedra angular en el estudio de los recursos fitogenéticos. *Acta Agrícola y Pecuaria*, 1(1), 1-6.
- Oicatá, M. (1986). *Caracterización de 62 clones nacionales de cacao existentes en el banco de germoplasma del CNI Palmira*. Instituto Colombiano Agropecuario (ICA). <http://hdl.handle.net/20.500.12324/31979>
- Perea, A., Aranzazu, F., & Martínez, N. (2013). *Características de calidad del cacao de Colombia: Catalogo de 26 cultivares*. Federación Nacional de Cacaoteros.
- Poland, J. A., & Rife, T. W. (2012). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome*, 5(3), 92-102. <https://doi.org/10.3835/plantgenome2012.05.0005>
- Quijada, A., Méndez-Cárdenas, G., Hernández-Baños, B., & Álvarez-Buylla, E. (2017). La región de los ITS del ADN ribosomal del núcleo (mrADN), fuente de caracteres moleculares en la sistemática de las gimnospermas. *Botanical Sciences*, 60, 159. <https://doi.org/10.17129/botsci.1527>
- Roa-Romero, H., Salgado-Mora, M., & Álvarez-Herrera, J. (2009). Análisis de la estructura arbórea del sistema agroforestal de cacao (*Theobroma cacao* L.) en el Soconusco, Chiapas - México. *Acta Biológica Colombiana*, 14, 97-109. <https://doi.org/10.15446/abc>
- Rocha, P. (2003). Marcadores moleculares, una herramienta útil para la selección genética de palma de aceite. *Palmas*, 24(42), 11-25.
- Rodríguez-Medina, C., Arana, A. C., Sounigo, O., Argout, X., Alvarado, G. A., & Yockteng, R. (2019). Cacao breeding in Colombia, past, present and future. *Breeding Science*, 69(3), 373-382. <https://doi.org/10.1270/jsbbs.19011>
- Romero, C., & Urrego, E. (2016). *Estudio del cacao en el Perú y en el Mundo*. Ministerio de Agricultura y Riego.
- Ruiz, X. (2014). *Diversidad Genética de cacao Theobroma cacao L. con Marcadores moleculares microsatélites*. [Master's thesis, Universidad Nacional de Colombia, sede Palmira]. Repositorio UNAL. <http://bdigital.una.edu.co/39793/1/7211504.2014.pdf>
- Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors - occurrence, properties and removal. *Journal of Applied Microbiology*, 113(5), 1014-1026. <https://doi.org/10.1111/j.1365-2672.2012.05384.x>
- Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution*, 9(4), 678-687. <https://doi.org/10.1093/oxfordjournals.molbev.a040752>
- Unión Internacional para la Protección de las Obtenciones Vegetales (UPOV). (2011). *Directrices para la ejecución del examen de la distinción, la homogeneidad y la estabilidad. Cacao 'Theobroma cacao' L.* Unión Internacional para la Protección de las Obtenciones Vegetales. https://www.upov.int/edocs/mdocs/upov/es/tc/47/tg_cacao_proj_4.pdf
- Ventura, M., González, A., & Batista, L. (2004). *Selección de árboles de cacao (Theobroma cacao) nativo e híbrido de buena calidad y rendimiento*. Instituto Dominicano de Investigaciones Agropecuarias y Forestales (IDIAF).
- White, T., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. En I. MA, D. Gelfand, J. Sninsky & T. White (Eds.), *PCR Protocols: A Guide to Methods and Applications* (pp. 315). New York: Academic Press.
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13(1), 134. <https://doi.org/10.1186/1471-2105-13-134>
- Young, A. M. (2008). The Chocolate Tree: A Natural History of Cacao. *Culture & Agriculture*, 30(1-2), 63-64. <https://doi.org/10.1111/j.1556-486X.2008.00009.x>

- Zambrano, J. (2017). *Relaciones filogenéticas entre tipos de cacao (Theobroma cacao L.): forastero, trinitario y nacional, basadas en marcadores morfológicos y secuencias nucleotídicas de la región ITS; y su posible uso en la identificación de clones*. [Undergraduate thesis, Universidad Técnica Estatal de Quevedo]. Repositorio Digital. UTEQ. <http://repositorio.uteq.edu.ec/handle/43000/2722>
- Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., Fouet, O., Argout, X., Guichoux, E., Salin, F., Loor-Solorzano, R., Bouchez, O., Vignes, H., Severts, P., Hurtado, J., Yopez, A., Grivetti, L., Blake, M., & Valdez, F. (2018). The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nature Ecology & Evolution*, 2(12), 1879-1888. <https://doi.org/10.1038/s41559-018-0697-x>