



SNP fingerprinting and farmer criteria for selection, multiplication, and traceability of cacao in Nicaragua¹

Huella genética y criterios de productores para apoyar la selección, multiplicación y trazabilidad del cacao en Nicaragua

Mélanie Bordeaux², Jonny Alonso Castillo², Abner Castro Olivás², Oswald R. Jiménez³, María José Torres⁴, Luis Orozco-Aguilar⁵

- ¹ Reception: September 5th, 2022. Acceptance: January 13th, 2023. This study was commissioned by the Nicaraguan Cacao Commission (COMCACAO), funded by COSUDE and led by the Nicaraguan Producers and Exporters Association (APEN).
- ² Fundación Nicafrance, Finca La Cumplida, Matagalpa, Nicaragua. research@fundacionnicafrance.org (corresponding author, <https://orcid.org/0000-0002-5444-2950>); jonny.castillo@fundacionnicafrance.org (<https://orcid.org/0000-0003-2410-0631>); abnercastrofnf@gmail.com (<https://orcid.org/0000-0001-8528-2892>).
- ³ University of Central America (UCA), The Interdisciplinary Institute of Natural Sciences, Managua, Nicaragua. oswalt.jimenez@uca.edu.ni (<https://orcid.org/0000-0001-6056-3700>).
- ⁴ Asociación de Exportadores de Nicaragua, Managua, Nicaragua. mtorres@apen.org.ni (<https://orcid.org/0000-0002-4432-7043>).
- ⁵ Corus International/Lutheran World Relief/Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Cartago Province, Costa Rica. lorozcoaguilar1980@gmail.com (<https://orcid.org/0000-0002-2731-3563>).

Abstract

Introduction. Genetic diversity, registration, and traceability of cacao planting material are the essential tripod to support its sustainable cultivation. In Nicaragua, cocoa planting material is mostly obtained from seeds, which leads to great variability in yield, bean quality, and tolerance to pests and diseases. Farmers, technical staff, development projects, and investors depend on a limited supply of elite trees as a reliable source of genetic material to support new cacao fronts and meet market quality standards. Therefore, the development of a national genetic improvement program and a planting material traceability system in Nicaragua are necessary to improve the sustainability of cacao cultivation. **Objective.** To evaluate the genetic resources on farms and provide information for future breeding programs, as well as to lay foundation for a national traceability and certification system. **Materials and methods.** Forty-nine elite trees selected by farmers in the main producing regions in Nicaragua between 2018 and 2020 were sampled. Ninety-three SNP markers were used to characterize them and resolve their genetic origins. **Results.** The evaluated cacao trees had a narrow genetic background, mainly composed of hybrids with Amelonado (36 %), Criollo (17 %) and Iquitos (15 %) origins. A set of trees with high genetic diversity that could be incorporated into a selection and breeding program was identified. **Conclusion.** Fingerprinting with SNP markers was a useful tool for evaluating the genetic links of cultivated cacao and can be used for varietal identity test at the farm level. The study provided the basis for developing both a breeding program and a traceability system of cacao planting material in Nicaragua.

Keywords: *Theobroma cacao*, registration, plant breeding, genetic groups, criollo.



Resumen

Introducción. La diversidad genética, el registro y la trazabilidad del material vegetal del cacao son esenciales para apoyar su cultivo sostenible. En Nicaragua, el material de plantación de cacao se obtiene mayoritariamente a partir de semillas, lo que provoca una gran variabilidad en rendimiento, calidad del grano y tolerancia a plagas y enfermedades. Los agricultores, el personal técnico, los proyectos de desarrollo y los inversores dependen de una oferta limitada de árboles élitos como fuente fiable de material genético para apoyar nuevos frentes de cacao y cumplir con los estándares de calidad del mercado. Por lo tanto, el desarrollo de un programa nacional de mejoramiento genético y un sistema de trazabilidad del material de siembra en Nicaragua son necesarios para mejorar la sostenibilidad del cultivo. **Objetivo.** Evaluar los recursos genéticos en las fincas y aportar información para futuros programas de mejoramiento, así como sentar las bases de un sistema nacional de trazabilidad y certificación. **Materiales y métodos.** Se muestrearon 49 árboles élitos seleccionados por los agricultores en las principales regiones productoras en Nicaragua entre 2018 y 2020. Se utilizaron 93 marcadores SNP para caracterizarlos y resolver sus orígenes genéticos. **Resultados.** Los árboles de cacao evaluados tenían un fondo genético estrecho, compuesto principalmente por híbridos con orígenes Amelonado (36 %), Criollo (17 %) e Iquitos (15 %). Se identificó un conjunto de árboles con alta diversidad genética que podría incorporarse a un programa de selección y mejoramiento. **Conclusión.** El fingerprinting con marcadores SNP fue una herramienta útil para evaluar los vínculos genéticos del cacao cultivado y puede servir para realizar pruebas de identidad varietal a nivel de finca. El estudio proporcionó las bases para desarrollar tanto un programa de mejoramiento como un sistema de trazabilidad del material vegetal de cacao en Nicaragua.

Palabras clave: *Theobroma cacao*, registro, fitomejoramiento genético, grupos genéticos, criollo.

Introduction

Globally, only about 30 % of cultivated cacao comes from breeding programs and other traceable sources. This limited diversity increases variability in crop productivity, bean quality, and susceptibility to climate change (Cilas & Bastide, 2020; Farrell et al., 2018; Motamayor et al., 2008). Despite the steadily increase in demand for cacao every year (End et al., 2021; Hütz-Adams et al., 2002), the chocolate industry faces multiple threats to supply of its key ingredient, cacao grains.

However, the lack of genetic diversity in current cacao cultivated varieties is often overlooked, even though it poses a threat to the entire cacao value chain. To address this issue, collecting, safeguarding, and facilitating access to cacao genetic diversity is essential. This will provide cacao breeders, scientists, and the industry with greater opportunities to access materials to produce varieties that comply with production and quality requirements, and ultimately to make these improved varieties available to small-scale farmers. A robust strategy to support regional and national breeding programs should focus on the neglected and underutilized genetic richness of cacao grown at the farm level (Ceccarelli et al., 2022; End et al., 2021).

Cacao farming in Nicaragua covers about 10,000 hectares and involves more than 11,000 small farmers which produce 6,600 metric tons annually, representing more than USD\$ 5 million in revenues (Wiegel et al., 2020). Nicaragua is ranked 25th as producing country, and since 2015 the country has been categorized as fine or aroma supply origin (International Cocoa Organization, 2015).

A large proportion of the cacao genetic material currently grown by small farmers comes from hybrid open-pollinated seeds, which are characterized by high variability in terms of yield, tolerance to pests and diseases,

and bean quality (Ji et al., 2013). Other producers were supplied with planting materials of higher quality sourced from the cacao research center “El Recreo”, managed by the Nicaraguan Institute of Agricultural Technology (INTA), and distributed via several cacao cultivation projects over a decade (1990-2000). Some farmers multiplied their own seeds from the best-performing trees, but the agronomic performance of these cacao plantations is below expectations. Additionally, Trinitarian clones such as UF-296, UF-613, UF-667, and ISC-95 were widely distributed during the 90s (Mata-Quirós et al., 2017).

Between 2005-2013, new planting material, including CC-137, PMCT-58, and three new clones that are descendants of UF-237 (CATIE-R1, CATIE-R4, and CATIE-R6), was introduced in to the country due to their high bean quality, productivity, and tolerance to moniliasis and black pod (Phillips-Mora et al., 2012). A small proportion of farmers, mainly medium-size farmers, bought or reproduced their own planting material using promising clones or varieties available at private clonal gardens nearby.

Since 2015, nursery companies such as ECOM have used advanced micro-grafting technology to mass-produce cacao plants distributed among ongoing cacao projects nationwide (ECOM Agroindustrial Corp. Lt, 2020). Although certified at the nursery and multiplication process level, no private provider has distributed duly certified cacao genetic material (Ceccarelli et al., 2022).

In 2018, the cacao research and breeding program of INTA developed a new productive clone named “PACAYITA”, which has been widely distributed. However, no data on agronomic and yield performances are currently available (Instituto Nicaragüense de Tecnología Agropecuaria, 2018; Martorell Mir, 2020).

In Nicaragua, the origin of several clones is unknown. Mislabeling of available genetic resources is one of the most significant problems within germplasm collections and clonal gardens, with some studies estimating labeling errors of up to 30 % in international collections (Turnbull et al., 2017).

Previous studies have shown that the genetic makeup of cacao in Nicaragua is mainly composed of Trinitarian and Criollo materials (Trognitz et al., 2011). However, these studies have limited geographic coverage or focused on a few germplasm collections (Herrera-García et al., 2015; Ruiz et al., 2011), leaving the genetic diversity of cultivated cacao across the country largely unexplored.

Recent advances in DNA sequencing technology, particularly Single Nucleotide Polymorphism (SNP) sequencing, allowed the detection of duplicates and mislabeling (Mata-Quirós et al., 2017). This technology has been successfully used to study the genetic diversity of various crops, including grapevine (De Lorenzis et al., 2019), corn (Adu et al., 2019), and beans (Jiménez, 2019), among others. In the case of cacao, SNP markers have been employed to characterize and classify the resources of different germplasm banks worldwide (Mahabir et al., 2017; Mata-Quirós et al., 2017).

According to a national baseline assessment conducted by the Ministry of Agriculture (MAG), substituting around 5000 ha of hybrid plantations with an average density of 600 plants/ha with clonal plantations at 1000 plants/ha would require producing at least 5 million grafted plants (End et al., 2021; Somarriba, 2013). With the current growth and invest in commercial cacao in Nicaragua, there is a need for a national breeding strategy, as well as a traceability and certification system. This would enable relevant authorities to better coordinate the supply and demand of planting material and ensure the genetic integrity of cacao plants being produced and sold at clonal gardens and private nurseries.

The objective of this study was to evaluate the genetic resources on farms and provide information for future breeding programs, as well as to lay foundation for a national traceability and certification system.

Materials and methods

Sampling strategy and selection criteria for elite trees

The genetic material (leaves) was sampled from selected cacao trees located in different farms owned by independent farmers or members of the following associations/cooperatives: La Campesina, Flor de Pancasán, Cooperativa Agropecuaria de Servicios Extracciones Esenciales (COOPESIUNA, R.L), and others working with the non-governmental organization Catholic Relief Services (CRS). Ideally, to obtain a representative sample of the genetic variability of cacao in each visited zone, the sampling strategy should include as many individuals as possible. However, due to the high cost of the analysis, trees with special features were prioritized, including (but not limited to) high yielding trees, elderly individuals, trees showing high tolerance to pests and diseases, vigorous individuals, and trees with white seeds (locally referred to as Criollo or Acriollado). Based on these criteria, a total of 49 trees were selected by farmers across various locations in the municipalities of Matiguas (15 samples), Rancho Grande (11 samples), Siuna (10 samples), Bonanza (7 samples), Waslala (6 samples), Rosita (5 samples), Matagalpa (2 samples), Rivas (2 samples), and Nueva Segovia (2 samples).

Collection, preparation, and shipment of samples

Between 2018 and 2020, several collection expeditions were carried out in selected cacao production areas, resulting in the sampling of a total of 49 cacao trees. From each tree, a leaf in good health was collected from the third node of the productive branch, which was located up to 2 meter above ground. The leaf samples were air-dried, certified for phytosanitary controls to ensure the absence of pests and diseases, and subsequently shipped to the laboratory for DNA extraction and fingerprinting. In addition to the leaf sampling, a survey was conducted with each farmer to document the local criteria used to select and evaluate elite tree at the farm level.

DNA extraction and sequencing

DNA was extracted from each leaf sample using the DNeasy® plant Mini kit Tissue Kit from Qiagen, Inc., Valencia, CA, USA, following the supplier's instructions. The samples were stored at -80 °C until sequencing was performed using a Caspar-fluidigm SNP microfluidic probe / QPCR / fluorescence reading. The sequences of the 96 SNP markers used in this study were provided by the USDA (D. Zhang, personal correspondence, 2018). These markers were previously selected based on their balanced distribution in the ten chromosomes of *T. cacao* and for their high level of polymorphism (Ji et al., 2013). This set of SNP markers has been partially or completely used in other similar studies conducted in Honduras (Ji et al., 2013), Puerto Rico (Cosme et al., 2016), Jamaica (Lindo et al., 2018), Peru (Arevalo-Gardini et al., 2019), and Madagascar (Li et al., 2021).

Data analysis

To assign each sampled tree to distinct genetic group, a principal component analysis was performed using the Adegenet package in R software (R Core team, 2020), with the parameters $pca = 10$ and $da = 10$. To analyze relationships between groups Structure software (version 2.3.4) (Pritchard Lab Stanford University, 2012) was used, assigning each tree to one of ten groups previously described by (Motamayor et al., 2008) based on data from 220 genotypes of *T. cacao* that were sequenced using the same SNP markers. Three markers were ignored in

which the expected alleles were not found (SNP_546, SNP_731, and SNP_1149). The parameters used for genetic grouping included 93 loci, a 50,000 burn-in period, and 100,000 reps with $K=10$. Additionally, a Bartlett test was conducted for each elite tree sample to determine its heterozygosity and consanguinity index. Tree origins were classified as pure when exhibited more than 75 % membership in a particular genetic group. Conversely, Trees with origin values below 75 % were classified as hybrids with origins from two or more groups, as per the method outlined by Lukman et al. (2014).

Results

Genetic diversity and inbreeding of elite cacao trees in Nicaragua

The heterozygosity index of the sampled elite cacao trees was found to be significantly lower than expected, as indicated by the Bartlett test results, with H_o and H_e values of 0.28 and 0.38, respectively. The analysis of the genetic diversity, using 93 loci, detected an average of 185 alleles, indicating that the SNP markers used were sufficiently polymorphic and appropriate for the analysis of the genetic diversity in this study. Additionally, a high consanguinity index (0.58) was observed among the elite trees, indicating a high presence of inbreeding.

The genetic relationships between sampled trees and the reference groups were revealed through Principal Component Analysis (PCA), as shown in Figure 1.

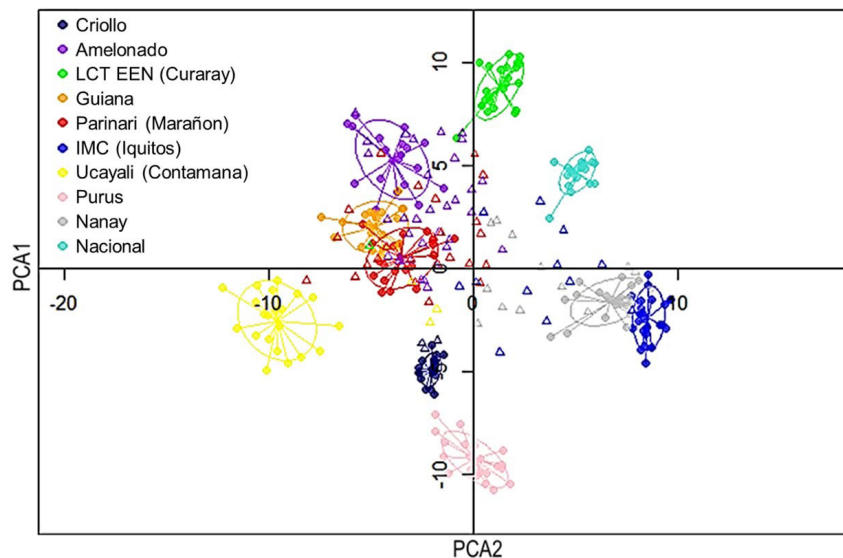


Figure 1. Principal component analysis (PCA) of the ten genetic groups of *T. cacao* and linkages of sampled elite trees in Nicaragua. Rounded marks represent the individuals within groups, while triangles represent the individuals sequenced in this study (Nicaragua, 2020).

Figura 1. Análisis de componentes principales (ACP) de los diez grupos genéticos de *T. cacao* y vínculos de árboles elite muestreados en Nicaragua. Las marcas redondeadas representan a los individuos dentro de los grupos. Los triángulos representan a los individuos secuenciados en este estudio (Nicaragua, 2020).

Genetic structure of elite trees

The majority of the sequenced trees did not have a pure origin and were identified as hybrids between various reference groups. Among the sampled trees, the most represented genetic groups were Amelonado (36 %), followed by Criollo (17 %) and IMC-Iquitos (15 %) (Figure 2). Other genetic groups such as Ucayali, Nacional, Nanay, and Parinari were represented in less than 15 % of the sampled trees.

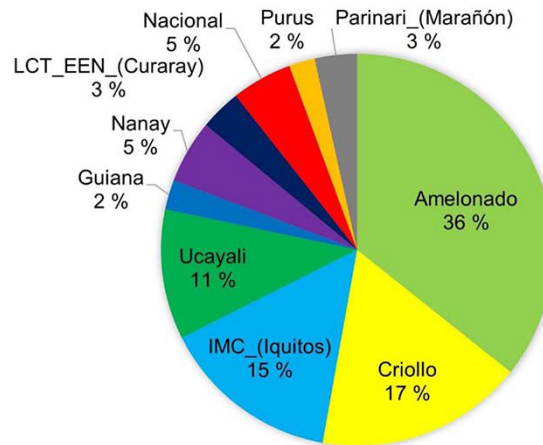


Figure 2. Percentage of representation of each genetic group among the elite cacao trees sampled (n = 49) across different growing locations in Nicaragua in 2020.

Figura 2. Porcentaje de representación de cada grupo genético entre los árboles de cacao elite muestreados (n = 49) en diferentes ubicaciones de cultivo en Nicaragua en 2020.

Twenty-seven of the sampled trees had more than half of their genetics background from a particular group, mostly Amelonado. On the other hand, 36 trees had origins from more than three groups. For instance, the tree labeled “M19-05W” had origins from six different groups. Nine trees exhibited genetic purity, with more than 75 % of their genetics belonging to one specific group. These trees were “El Bálsamo1”, “La union-El Corozo”, “A19-4R”, “La Campesina7”, “La Campesina1”, “M19-01W” (all from Amelonado group). In addition, two trees from La Cumplida and one tree from Mozonte municipality also showed a high level of purity (mainly Criollo) (Figure 3).

The analysis showed that approximately 50 % of the elite cacao trees had ancestral Criollo lineage, ranging from 5 % to 50 % of Criollo ancestry (Figure 4). Owners of the trees with the highest percentage of Criollo lineage described them as Criollo or Acriollado, indicating a local understanding of the genetic distinction of cultivated trees and the application of farmers’ criteria for on-farm selection and breeding purposes.

White-seeded trees were found to belong to three different types of genetic composition: (1) 100 % Criollo, as seen in the samples “La Cumplida1”, “La Cumplida2”, and “Mozonte”; (2) hybrids with Criollo lineage, such as samples “A19-5B”, “A19-4”, “A19-3” for example; and (3) trees without Criollo lineage, as defined in the Motamayor groups, including “La Campesina4”, “A19-1”, “A19-2R”, “El_Balsamo1”, and “La Campesina3”. These trees were mostly hybrids with Amelonado or Ucayali origins (Figure 4).

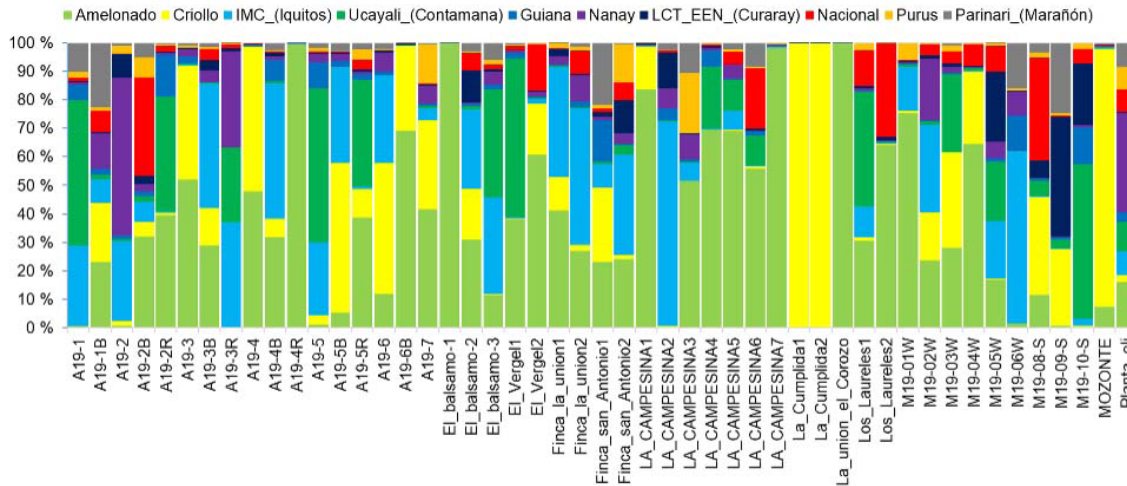


Figure 3. Genetic structure of 49 elite cacao trees from Nicaragua based on the genetic groups described by Motamayor et al. (2008), Nicaragua, 2020.

Figura 3. Estructura genética de 49 árboles elite de cacao de Nicaragua basada en los grupos genéticos descritos por Motamayor et al. (2008), Nicaragua, 2020.

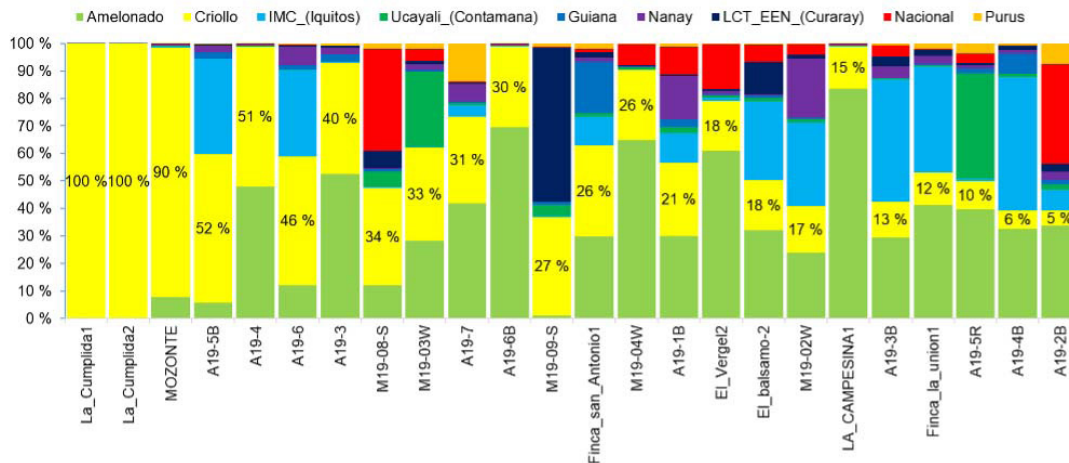


Figure 4. Genetic structure of “Criollo” and “Acriollado” trees selected by farmers in different growing locations in Nicaragua. 2020.

Figura 4. Estructura genética de árboles “Criollo” y “Acriollado” seleccionados por agricultores en diferentes ubicaciones de cultivo en Nicaragua. 2020.

Farmers’ criteria for the selection of elite trees

The selection of elite trees carried out by farmers appeared to be based on three main criteria: (1) productivity, (2) resistance or tolerance to pod diseases, and (3) white bean color, which farmers considered an indicator of quality. The frequency at which each of these criteria appeared in the sampled trees is shown in Figure 5. Most elite

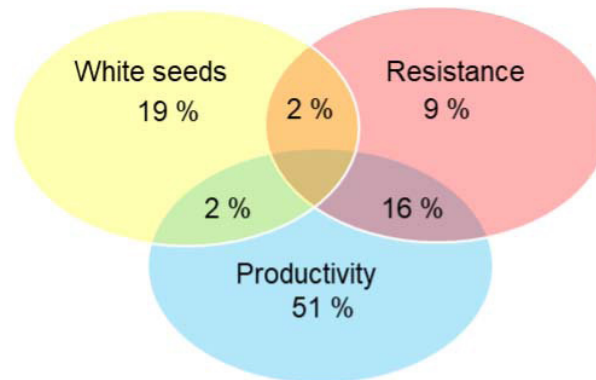


Figure 5. Selection criteria used by cacao farmers to classify an individual as elite cacao tree. The numbers indicate the frequency at which each criterion was applied by farmers. Nicaragua, 2020.

Figura 5. Criterios de selección utilizados por los productores de cacao para clasificar a un individuo como un árbol de cacao elite. Los números indican la frecuencia con la que cada criterio fue aplicado por los agricultores. Nicaragua, 2020.

trees were selected for their high productivity (69 %), while 23 % of the farmers selected them for their high-quality potential, which refers to the white seeds. Up to 27 % of farmers selected their trees based on apparent resistance/tolerance against moniliasis or the black pod disease. The plants selected for their productivity ranged from 3 to 40 years old, with an average age of 24 years. They produced between 50 and 280 pods per year, with an average of 123 pods per tree per year. Some farmers selected trees based on combined criteria: 2 % combined white seeds and resistance, 16 % combined resistance and productivity, and 2 % combined productivity and white seeds. None of them combined all three criteria.

Linkages between local criteria and genetic background

Elite trees which selection was based on productivity had mostly Amelonado origin (36 %) and IMC (Iquitos) origin (19 %) (Figure 6). The other groups were represented in a lower proportion (below 11 % each).

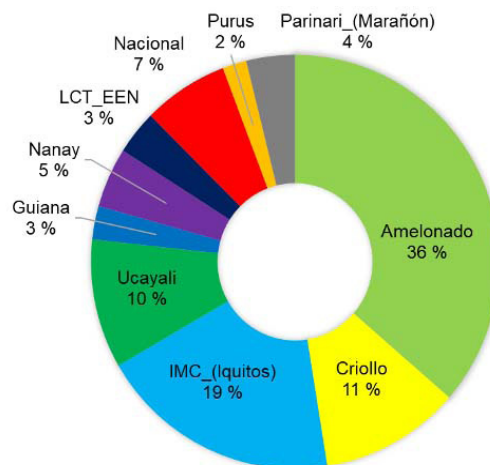


Figure 6. Genetic composition of the 34 elite cacao trees selected by the farmers for their high productivity in Nicaragua, 2020.

Figura 6. Composición genética de árboles seleccionados por los productores por su alta productividad en Nicaragua, 2020.

The genetic composition of trees selected based on resistance to black pod and moniliasis was very similar to that of trees selected based on productivity criteria: 47 % Amelonado and 18 % IMC-Iquitos (Figure 7).

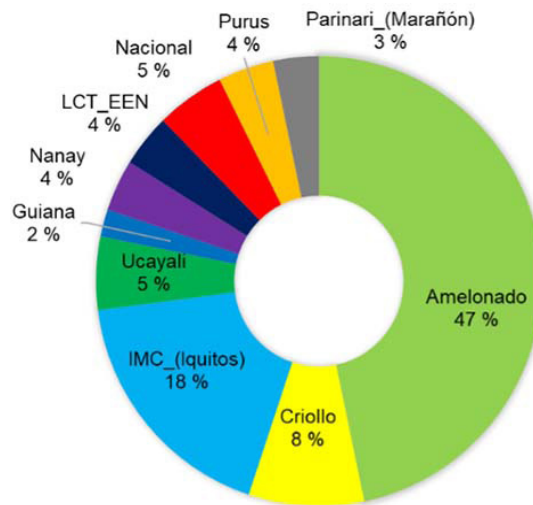


Figure 7. Genetic composition of elite cacao trees selected by the farmers to be more tolerant or resistant to monilia and black pod diseases than average. Nicaragua, 2020.

Figura 7. Composición genética de árboles seleccionados por los productores por ser más tolerante o resistente a monilia y mazorca negra. Nicaragua, 2020.

A higher percentage of Criollo lineage (36 %) was observed in trees selected for their white seeds, with Amelonado paternity (31 %) (Figure 8).

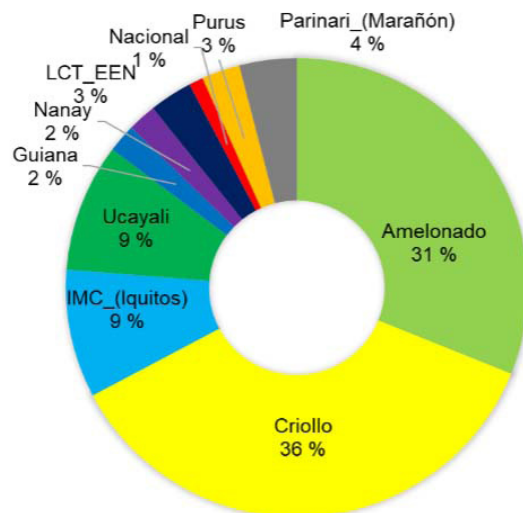


Figure 8. Genetic composition of elite cacao trees selected by the farmers for having white seeds. Nicaragua, 2020.

Figura 8. Composición genética de árboles seleccionados por los productores por tener semillas blancas. Nicaragua, 2020.

Discussion

Cacao has been cultivated in Latin America since pre-Columbian times (~1900 B.C.). Criollos were the first to be introduced and domesticated, while Forasteros and mostly Trinitarios (hybrids between Forasteros and Criollos, originally from Trinidad and Tobago), arrived later (Motamayor et al., 2013). However, they did not take long to replace Criollo trees across Central America due to their greater adaptability to cope with biotic and abiotic stresses, as well as their higher yield potential (Cornejo et al., 2018).

There is evidence of large cacao plantations on the pacific coast of Nicaragua since before the arrival of Christopher Columbus. The country's production area extended mainly to the isthmus of Rivas, where the Nicaraguan Criollo (*T. cacao*) and another variety known as Lagarto cacao (*T. pentagona*) were cultivated, both of high quality and white seeds.

It was not until the 19th century that the first documented massive introductions of cacao from abroad began to complement the two largest commercial plantations in Nicaragua: Las Mercedes and Valle Menier, both in Nandaime, Granada. Las Mercedes imported "Cauca" from Colombia and "Trinitario" from Trinidad, while Valle Menier introduced seeds of the genetic Trinitario type mixed with local Criollo and Lagarto trees (Radell, 1971).

Several waves of introduction of genetic material occurred during the 70s, 80s, and 90s, from various producing countries. The most recent wave of massive introduction was that of 2007-2013 under the PCC project (Proyecto Cacao Centroamérica), which introduced and distributed novel genetic material from CATIE, principally Trinitarios (Phillips-Mora et al., 2012).

Previous studies using Single-Sequence Repeats (SSR) markers in Waslala municipality revealed a prevalence of Trinitarian hybrids and identified old Criollo trees. The distribution of these hybrids suggested two origins: trees near the main roads and the city of Waslala were introduced recently, possibly in the 2000s by the cooperative Cacaonica and third parties, while trees located further away could have originated from older introduction, with trees adapted to the climatic conditions of those areas (Trognitz et al., 2011). Additionally, Ji et al. (2013) analyzed fine cacao clones in Nicaragua and Honduras with SNP markers, where they showed a majority of Trinitario with Amelonado origin.

The present study is consistent with historical records, as a prevalence of Amelonados (36 %) and Criollos (17 %) was observed (Figure 2), with most trees resulting from the hybridization between these two groups. The prevalence of Amelonados among the sampled trees could be explained by the ancestral selection of cacao "Indio" and "Indio Rojo" types, linked to the "Amelonado" group due to their high productivity (Ji et al., 2013).

Studies conducted in the Dominican Republic and Cuba have found a similar trend, with a clear dominance of trees with Amelonado and Criollo origins. The study carried out in the Dominican Republic showed a genetic composition very similar to what was found in this study, with the same three major groups represented but in different proportions: 72.1 % Amelonado, 9.5 % Criollo, and 7.8 % Iquitos (Boza et al., 2013). In Cuba, a prevalence of Amelonado (61.6 %) and Criollo (27.3 %) was also found, and the authors stated that those trees were introduced to the island from Central America (Bidot Martínez et al., 2015). Studies conducted in other producing countries reported other genetic groups: in Ghana, Padi et al. (2015) discovered a majority of Amelonados, Nanay, and Iquitos. In Indonesia, a majority of Trinitario, Parinari, and Nanay genetic groups were observed among small cacao farms (Dinarti et al., 2015).

Currently, cacao farms in Nicaragua are populated of two types of genetic material: a) commercial clones distributed by projects or cooperatives for their superior productive characteristics (30 %), and b) trees selected by farmers mainly based on phenotypical criteria (70 %). The genetic analysis highlights that farmer mostly selected Amelonado type trees, which can be explained by their greater adaptability, resistance, and availability in the country (Ceccarelli et al., 2022; Martorell Mir, 2020).

Unlike farmers who received cacao plants from development projects or their cooperative, very few independent farmers knew the variety or genetic origin of the cacao trees they grow. This lack of knowledge may explain the poor compatibility and low productivity seen in certain farms. Knowing the genetics of the trees could help understand the best combinations and avoid incompatibility issues between clones from the same farm.

The high inbreeding coefficient (0.58) also evidences human or natural hybridization with geographically and genetically close trees. Moreover, individuals from the two most represented groups (Amelonado and Criollo) have the capacity for self-compatibility (Lanaud et al., 2017), which increases the consanguinity in the offspring.

In this study, it was found that half of the trees described by farmers as elite trees with white seeds did not have a Criollo origin in their genetic background. Consequently, having white seeds does not necessarily mean that the tree is a native or original “Criollo” tree, as defined by Motamayor et al. (2008). Other groups, such as the Amelonado, may also have white or slightly purple seeds, due to their origin or local hybridization.

Many trees labeled as elite trees had genetic linkages with well-known clones (i.e. IMC 67, UF, ICS; data not shown) widely distributed in Nicaragua during the 80s and 90s. Therefore, it is very likely that they come from open pollination of commercial clones, whose seeds were harvested and planted (Mata-Quiros et al., 2017). This process of on-farm selection seeks “genetic improvement” based on production features but leads to uncertainty regarding the compatibility of selected trees with their neighbors (Ayestas et al., 2013; López et al., 2021).

The genetic analysis of cacao trees selected by farmers for different criteria such as productivity, disease resistance, and white seeds revealed that Amelonado trees were mostly chosen for their productivity and resistance to diseases. This indicates that farmers selected trees that fulfilled both criteria simultaneously. The presence of Criollo trees of between 6 to 10 % in these productive and diseases-resistant trees is likely due to the existence of native Criollo trees in the area, whose origin was mixed through local selection and multiplication over time. The trees chosen for their white seeds were mainly of Criollo and Amelonado origin, owing to the strong presence of these groups in Nicaragua.

It is worth noting that none of the tree sampled in this study combined all three selection criteria, i.e., having white seeds, being resistant to diseases, and being highly productive. This underscores a gap in the variety of cacao offered and emphasizes the need for further breeding efforts to enrich the genetic pool of local farms.

DNA analysis is essential for determining the origin or reliability of traded plants and serves as the foundation for an efficient, effective, and reliable traceability system (End et al., 2021). This type of analysis enables the identification of genetics variations that cannot be observed through visual or physical analyses of cacao plants. Such variations can impact the selection and breeding pipeline (Motamayor et al., 2013).

The set of 93 SNPs tested in this study was used to characterize and classify the sampled trees, creating a unique genetic fingerprint for each planted material. This set of markers was found to be suitable for certification and traceability purposes. The genetic make-up suggested that the number of SNPs markers could be reduced while achieving similar results, thereby reducing operational costs. For instance, Ji et al. (2013) reported that a smaller number of markers (26 SNPs) could provide 99.99 % confidence in identifying a unique cacao tree.

The knowledge of the region’s genetic diversity serves as baseline for the conservation and use of on-farm *T. cacao* germplasm (Ceccarelli et al., 2022). It is crucial to maintain a high genetic diversity to have various sources of resistance to pests and diseases, as well as quality attributes, among others factors (Padi et al., 2015).

Currently, the clones traded in Nicaragua mostly belong to the Amelonado genetic group, with a limited diversity. The present study did not identify significant proportions of germplasms from the Guiana, LCT-EEN, Nanay, Parinari, and Purus groups. Importing cacao from these genetic groups, such as from the international collections of the Cocoa Research Centre of Trinidad and Tobago and/or CATIE, could expand the genetic variety of cacao trees in Nicaragua and introduce genes of interest, such as resistance to pests and diseases or drought tolerance, into the Nicaraguan cacao trees gene pool. Clones from the Nanay, Parinari, and IMC groups were highlighted as sources of resistance to pests and diseases (Lukman et al., 2014).

Several trees with Criollo lineage could be part of the improvement plan to take advantage of genes of interest for organoleptic quality adapted to local climate. Trees such as “La_Cumplida1”, “La_Cumplida2”, “Mozonte”, “A19-4”, as shown in Figure 3, could play that role. However, a quality assessment of the cacao liquor from these trees is necessary.

Two samples, “M19-08S” and “M19-09S”, exhibited a genetic profile that was markedly different from the others, forming a distinct cluster (data not shown). These trees were sourced from Siuna, Floripon, and Rancho Alegre communities, and were described by farmers as “Cacao Mono” and “Native cacao that does not form horquet”. Given their description and genetics, they may belong to the same family as *Theobroma cacao* (Malvaceae), but to a different species such as *Herrania purpurea*. *H. purpurea* produces pods that are very similar to those of *T. cacao* but are smaller and produce a bitter drink (Belt, 2003).

Other trees selected by the farmers comprised various ancestries such as “A19-1B”, “M19-05W”, and “La_Campesina5”, among others. These well-adapted trees to local conditions might be useful for a national selection and breeding program. However, further assessment of these clones in terms of vegetative growth, yield potential, compatibility, resistance to pests and diseases, adaptation to the environment, and organoleptic quality is recommended.

Conclusion

The genetic analysis carried out in this study revealed the unique genetic make-up of the Nicaraguan cacao, which explains its exclusive profile. Additionally, the analysis exposed a low diversity of local cacao and a lack of knowledge regarding the genetic origin of the clones. This emphasizes the need for a certification and traceability system for cacao planting material in Nicaragua. The SNP fingerprinting of several local trees provided the necessary foundation to implement this traceability and certification system. Furthermore, the genetic study provided insights for a national breeding program. Several genetic materials that could integrate a local breeding program were identified, allowing for a better selection of clones that are more adapted to local conditions in the future. By having greater control over the compatibility between clones, higher yield can be achieved in the medium term.

Acknowledgement

The authors would like to express their gratitude to the cooperatives La Campesina, Flor de Pancasán, COOPESIUNA, as well as all the producers and the Instituto de Protección y Sanidad Agropecuaria (IPSA), who provided the genetic material analyzed in this study. The authors also like to thank CRS for collecting samples from the Siuna, Bonanza, and Rosita regions, and M.Sc. Allan Mata from CATIE for his assistance during authors visit to collect samples from various commercial clones in the cacao collection at CATIE. Finally, authors would like to acknowledge Dr. Dapeng Zhang from USDA for sharing the sequence list of the 96 SNPs used for this study, as well as the sequences of 220 individuals that allow authors to link the samples included in this study to the different genetic groups.

References

Adu, G. B., Badu-Apraku, B., Akromah, R., Garcia-Oliveira, A. L., Awuku, F. J., & Gedil, M. (2019). Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *PLoS ONE*, *14*(4), Article e0214810. <https://doi.org/10.1371/journal.pone.0214810>

- Arevalo-Gardini, E., Meinhardt, L. W., Zuñiga, L. C., Arévalo-Gardni, J., Motilal, L., & Zhang, D. (2019). Genetic identity and origin of “Piura Porcelana” a fine-flavored traditional variety of cacao (*Theobroma cacao*) from the Peruvian Amazon. *Tree Genetics & Genomes*, *15*, Article 11. <https://doi.org/10.1007/s11295-019-1316-y>
- Ayestas, E., Orozco, L., Astorga, C., Munguía, R., & Vega, C. (2013). Caracterización de árboles promisorios de cacao en fincas orgánicas de Waslala, Nicaragua. *Agroforestería En Las Américas*, *49*, 18–25.
- Belt, T. (2003). *El naturalista en Nicaragua* (Serie viajeros No. 4, Trad. J. I. Barquero). Colección Cultural de Centro America.
- Bidot Martínez, I., Riera Nelson, M., Flamand, M. -C., & Bertin, P. (2015). Genetic diversity and population structure of anciently introduced Cuban cacao *Theobroma cacao* plants. *Genetic Resources and Crop Evolution*, *62*(1), 67–84. <https://doi.org/10.1007/s10722-014-0136-z>
- Boza, E. J., Irish, B. M., Meerow, A. W., Tondo, C. L., Rodríguez, O. A., Ventura-López, M., Gómez, J. A., Moore, J. M., Zhang, D., Motamayor, J. C., & Schnell, R. J. (2013). Genetic diversity, conservation, and utilization of *Theobroma cacao* L.: Genetic resources in the Dominican Republic. *Genetic Resources and Crop Evolution*, *60*(2), 605–619. <https://doi.org/10.1007/s10722-012-9860-4>
- Ceccarelli, V., Lastra, S., Loor Solórzano, R. G., Chacón, W. W., Nolasco, M., Sotomayor Cantos, I. A., Plaza Avellán, L. F., López, D. A., Fernández Anchundia, F. M., Dessauw, D., Orozco-Aguilar, L., & Thomas, E. (2022). Conservation and use of genetic resources of cacao (*Theobroma cacao* L.) by gene banks and nurseries in six Latin American countries. *Genetic Resources and Crop Evolution*, *69*(3), 1283–1302. <https://doi.org/10.1007/s10722-021-01304-3>
- Cilas, C., & Bastide, P. (2020). Challenges to cocoa production in the face of climate change and the spread of pests and diseases. *Agronomy*, *10*(9), Article 1232. <https://doi.org/10.3390/agronomy10091232>
- Cornejo, O. E., Yee, M. -C., Dominguez, V., Andrews, M., Sockell, A., Strandberg, E., Livingstone III, D., Stack, C., Romero, A., Umaharan, P., Royaert, S., Tawari, N. R., Ng, P., Gutierrez, O., Phillips, W., Mockaitis, K., Bustamante, C. D., & Motamayor, J. C. (2018). Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process. *Communications Biology*, *1*, Article 167. <https://doi.org/10.1038/s42003-018-0168-6>
- Cosme, S., Cuevas, H. E., Zhang, D., Oleksyk, T. K., & Irish, B. M. (2016). Genetic diversity of naturalized cacao (*Theobroma cacao* L.) in Puerto Rico. *Tree Genetics & Genomes*, *12*, Article 88. <https://doi.org/10.1007/s11295-016-1045-4>
- De Lorenzis, G., Mercati, F., Bergamini, C., Cardone, M. F., Lupini, A., Maucri, A., Caputo, A. R., Abbate, L., Barbagallo, M. G., Antonacci, D., Sunseri, F., & Brancadoro, L. (2019). SNP genotyping elucidates the genetic diversity of Magna Graecia grapevine germplasm and its historical origin and dissemination. *BMC Plant Biology*, *19*, Article 7. <https://doi.org/10.1186/s12870-018-1576-y>
- Dinarti, D., Susilo, A. W., Meinhardt, L. W., Ji, K., Motilal, L. A., Mischke, S., & Zhang, D. (2015). Genetic diversity and parentage in farmer selections of cacao from southern Sulawesi, Indonesia revealed by microsatellite markers. *Breeding Science*, *65*(5), 438–446. <https://doi.org/10.1270/jsbbs.65.438>
- ECOM Agroindustrial Corp. Lt. (2020). *Breaking new ground in coffee & cacao farming in Nicaragua*. <https://bit.ly/3Rc19Tz>
- End, M. J., Daymond, A. J., & Hadley, P. (2021). *Directrices técnicas para el movimiento seguro del germoplasma del cacao* (Versión revisada de las Directrices técnicas de FAO/IPGRI No. 20). Bioversity International. <https://hdl.handle.net/10568/104540>
- Farrell, A. D., Rhiney, K., Eitzinger, A., & Umaharan, P. (2018). Climate adaptation in a minor crop species: is the cocoa breeding network prepared for climate change? *Agroecology and Sustainable Food Systems*, *42*(7), 812–833. <https://doi.org/10.1080/21683565.2018.1448924>

- Herrera-García, K., Aragón-Obando, E., & Aguilar-Bustamante, V. (2015). Diversidad genética en 105 accesiones de cacao (*Theobroma cacao* L.) colectadas en nicaragua, utilizando 10 marcadores moleculares tipo SSR. *La Calera*, 15(25), 54–62. <https://lacialera.una.edu.ni/index.php/CALERA/article/view/260>
- Hütz-Adams, F., Campos, P., & Fountain, A. C. (2022). *Barómetro del cacao - Base de referencia para Latinoamérica*. Cocoa Barometer. <https://cocoabarometer.org/wp-content/uploads/2022/11/220923-Cocoa-Barometer-Americas-ES.pdf>
- International Cocoa Organization. (2015, September 18). *Report by the chairman on the meeting of the ICCO Ad Hoc Panel on fine or flavour cocoa to review annex "C" of the international cocoa agreement, 2010*. <https://www.icco.org/wp-content/uploads/2019/06/FFP-4-5-Report-of-meeting-final-edited-w.pdf>
- Instituto Nicaragüense de Tecnología Agropecuaria. (2018). *Catálogo de clones de cacao*. <https://www.caja-pdf.es/2018/12/04/catalogo-de- clones-de-cacao-fida-grun2018/catalogo-de- clones-de-cacao-fida-grun2018.pdf>
- Ji, K., Zhang, D., Motilal, L. A., Boccara, M., Lachenaud, P., & Meinhardt, L. W. (2013). Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. *Genetic Resources and Crop Evolution*, 60(2), 441–453. <https://doi.org/10.1007/s10722-012-9847-1>
- Jiménez, O. R. (2019). Common bean (*Phaseolus vulgaris* L) breeding. In J. M. Al-Khayri, S. Mohan Jain, & D. V. Johnson (Eds.), *Advances in plant breeding strategies: Legumes* (Vol. 7, pp. 151–200). Springer. https://doi.org/10.1007/978-3-030-23400-3_5
- Lanaud, C., Fouet, O., Legavre, T., Lopes, U., Sounigo, O., Eyango, M. C., Mermaz, B., Da Silva, M. R., Loor Solorzano, R. G., Argout, X., Gyapay, G., Ebairey, H. E., Colonges, K., Sanier, C., Rivallan, R., Mastin, G., Cryer, N., Boccara, M., Verdeil, J. -L., ... Clément, D. (2017). Deciphering the *Theobroma cacao* self-incompatibility system: from genomics to diagnostic markers for self-compatibility. *Journal of Experimental Botany*, 68(17), 4775–4790. <https://doi.org/10.1093/jxb/erx293>
- Li, Y., Zhang, D., Motilal, L. A., Lachenaud, P., Mischke, S., & Meinhardt, L. W. (2021). Traditional varieties of cacao (*Theobroma cacao*) in Madagascar: their origin and dispersal revealed by SNP markers. *Beverage Plant Research*, 1, Article 4. <https://doi.org/10.48130/BPR-2021-0004>
- Lindo, A. A., Robinson, D. E., Tennant, P. F., Meinhardt, L. W., & Zhang, D. (2018). Molecular Characterization of Cacao (*Theobroma cacao*) Germplasm from Jamaica Using Single Nucleotide Polymorphism (SNP) Markers. *Tropical Plant Biology*, 11(3–4), 93–106. <https://doi.org/10.1007/s12042-018-9203-5>
- López, M. E., Ramírez, O. A., Dubón, A., Cherubino Ribeiro, T. H., Díaz, F. J., & Chalfun-Junior, A. (2021). Sexual compatibility in cacao clones drives arrangements in the field leading to high yield. *Scientia Horticulturae*, 287, article 110276. <https://doi.org/10.1016/j.scienta.2021.110276>
- Lukman, Zhang, D., Susilo, A. W., Dinarti, D., Bailey, B., Mischke, S., & Meinhardt, L. W. (2014). Genetic Identity, Ancestry and Parentage in Farmer Selections of Cacao from Aceh, Indonesia Revealed by Single Nucleotide Polymorphism (SNP) Markers. *Tropical Plant Biology*, 7(3–4), 133–143. <https://doi.org/10.1007/s12042-014-9144-6>
- Mahabir, A., Motilal, L. A., Gopaulchan, D., Sankar, A., & Umaharan, P. (2017, November 13-17). *Identification of a core SNP panel for cacao identity and population analyses* [Conference presentation abstract]. International Symposium on Cocoa Research (ISCR), Lima, Peru. <https://www.cabdirec.org/cabdirec/abstract/20203127067>
- Martorell Mir, J. (2020). *Caracterización a nivel nacional del sector cacao en Nicaragua. Gobernanza e incidencia en la cadena de valor de cacao en Nicaragua*. Asociación de Productores y Exportadores de Nicaragua, & Cooperación Suiza.

- Mata-Quirós, A., Arciniegas-Leal, W., Phillips-Mora, W., Meinhardt, L. W., & Zhang, D. (2017, November 13-17). *Understanding the genetic structure and parentage of the clonal series of cacao UF, CC, PMCT and ARF preserved in the International Cacao Collection at CATIE (IC3)* [Conference presentation abstract]. International Symposium on Cocoa Research (ISCR), Lima, Peru.
- Motamayor, J. C., Lachenaud, P., da Silva e Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). *PLoS ONE*, 3(10), Article e33111. <https://doi.org/10.1371/journal.pone.0003311>
- Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Livingstone III, D., Cornejo, O., Findley, S. D., Zheng, P., Utro, F., Royaert, S., Sasaki, C., Jenkins, J., Podicheti, R., Zhao, M., Scheffler, B. E., Stack, J. C., Feltus, F. A., Mustiga, G. M., Amores, F., ... Kuhn, D. N. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biology*, 14(6), Article r53. <https://doi.org/10.1186/gb-2013-14-6-r53>
- Padi, F. K., Ofori, A., Takrama, J., Djan, E., Opoku, S. Y., Dadzie, A. M., Bhattacharjee, R., Motamayor, J. C., & Zhang, D. (2015). The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. *Tree Genetics and Genomes*, 11, Article 44. <https://doi.org/10.1007/s11295-015-0875-9>
- Phillips-Mora, W., Arciniegas-Leal, A., Mata-Quiros, A., & Motamayor-Arias, J. C. (2012). *Catálogo de clones de cacao seleccionados por el CATIE para siembras comerciales* (Manual técnico 105). Centro Agronómico Tropical de Investigación y Enseñanza.
- Pritchard Lab, Stanford University. (2012). *Structure (2.3.4)* [Computer software]. <https://web.stanford.edu/group/pritchardlab/structure.html>
- R Core team. (2020). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Radell, D. R. (1971). *Historical geography of Western Nicaragua: The spheres of influence of León, Granada, and Managua, 1519-1965*. Berkeley.
- Ruiz, J. C., Roa Gamboa, O., & Marin Arguello, I. (2011). Molecular Ecology of genetic diversity of cacao cultivated in the south-east region of Nicaragua. *International Research Journal of Agricultural Science*, 1(1), 6–13.
- Somarriba, E. J. (2013). *Oferta mundial de tecnologías de producción de cacao prioritarias para elevar los rendimientos, mejorar la calidad del cacao y asegurar la sostenibilidad y seguridad alimentaria de las familias cacaoteras de Centroamérica*. Unidad Regional para el Desarrollo Sostenible.
- Trognitz, B., Scheldeman, X., Hansel-Hohl, K., Kuant, A., Grebe, H., & Hermann, M. (2011). Genetic population structure of cacao plantings within a young production area in Nicaragua. *PLoS ONE*, 6(1), Article e16056. <https://doi.org/10.1371/journal.pone.0016056>
- Turnbull, C. J., Daymond, A. J., Gutierrez, O., Hadley, P., Livingstone, D., Motamayor, J. C., Phillips, W., Umaharan, P., & Zhang, D. (2017, November 13-17). *Adopting Reference Genotypes to identify off-types in cacao collections* [Conference presentation abstract]. International Symposium on Cocoa Research, Lima, Peru.
- Wiegel, J., del Río, M., Gutiérrez, J. F., Claros, L., Sánchez, D., Gómez, L., González, C., & Reyes, B. (2020). *The cacao market system in Nicaragua. Opportunities for supporting renovation and rehabilitation*. International Center for Tropical Agriculture. <https://cgspace.cgiar.org/handle/10568/108120>