



## OPEN ACCESS

## EDITED BY

Shreekar Pant,  
Baba Ghulam Shah Badshah University, India

## REVIEWED BY

Arun K. Jugran,  
Govind Ballabh Pant National Institute of  
Himalayan Environment and Sustainable  
Development, India  
Padamnabhi Shanker Nagar,  
Maharaja Sayajirao University of Baroda, India

## \*CORRESPONDENCE

Sixto A. Imán

✉ siman@inia.gob.pe

Angelo F. Samanamud

✉ angelo.samanamud@unapiquitos.edu.pe

Juan C. Castro

✉ juan.castro@unapiquitos.edu.pe

RECEIVED 13 February 2025

ACCEPTED 29 May 2025

PUBLISHED 16 June 2025

## CITATION

Imán SA, Samanamud AF, Ramirez JF,  
Cobos M, Paredes C and Castro JC (2025)  
Development and phenotypic  
characterization of a native *Theobroma*  
*cacao* L. germplasm bank from the  
Loreto region of the Peruvian Amazon:  
implications for *Ex situ* conservation  
and genetic improvement.  
*Front. Conserv. Sci.* 6:1576239.  
doi: 10.3389/fcsc.2025.1576239

## COPYRIGHT

© 2025 Imán, Samanamud, Ramirez, Cobos,  
Paredes and Castro. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Development and phenotypic characterization of a native *Theobroma cacao* L. germplasm bank from the Loreto region of the Peruvian Amazon: implications for *Ex situ* conservation and genetic improvement

Sixto A. Imán<sup>1,2\*</sup>, Angelo F. Samanamud<sup>1,3\*</sup>, José F. Ramirez<sup>3</sup>,  
Marianela Cobos<sup>4,5</sup>, Cleydi Paredes<sup>1</sup> and Juan C. Castro<sup>4,5\*</sup>

<sup>1</sup>Direction of Genetic Resources and Biotechnology, San Roque Agricultural Experiment Station, Instituto Nacional de Innovación Agraria (INIA), Iquitos, Peru, <sup>2</sup>Academic Department of Ecology and Conservation, Faculty of Forestry Sciences, Universidad Nacional de la Amazonía Peruana (UNAP), Iquitos, Peru, <sup>3</sup>Academic Department of Soils and Crops, Faculty of Agronomy, Universidad Nacional de la Amazonía Peruana (UNAP), Iquitos, Peru, <sup>4</sup>Specialized Unit of Biotechnology Research Laboratory, Natural Resources Research Center of UNAP, Universidad Nacional de la Amazonía Peruana (UNAP), Iquitos, Peru, <sup>5</sup>Academic Department of Biomedical Sciences and Biotechnology, Faculty of Biological Sciences, Universidad Nacional de la Amazonía Peruana (UNAP), Iquitos, Peru

**Introduction:** The *ex situ* conservation and characterization of native *Theobroma cacao* L. genetic resources are critical for sustainable cacao production and breeding programs in the face of climate change and escalating disease pressures. This study aimed to establish and characterize a novel germplasm bank from the Loreto region of the Peruvian Amazon, a key center of cacao diversity.

**Methods:** We collected 140 native cacao accessions across 15 river basins in eight provinces of the Loreto region. Accessions were propagated using optimized grafting techniques with IMC 67 rootstock. Phenotypic evaluation was conducted on 402 plants using 36 standardized descriptors (25 quantitative and 11 qualitative). Data analysis included multivariate analysis using Uniform Manifold Approximation and Projection (UMAP) and Shannon-Weaver diversity indices to assess morphological diversity patterns.

**Results:** Grafting achieved 100% survival rate, establishing a comprehensive germplasm bank. Phenotypic characterization revealed exceptional morphological diversity, with quantitative traits exhibiting substantial variation, particularly in fruit characteristics (CV = 15.82–50.82%) and pod index (CV = 144.82%). Multivariate analysis identified five distinct phenotypic groups, with reproductive traits showing stronger differentiation than vegetative traits. Shannon-Weaver diversity indices highlighted high overall phenotypic diversity ( $H' \approx 0.7$ ), with seed longitudinal shape and fruit apex form displaying the highest trait-specific diversity ( $H' > 1.0$ ).

**Conclusion:** This comprehensive characterization establishes a foundation for future multiomics studies and advanced breeding strategies. The documented diversity offers opportunities to leverage CRISPR-Cas-based editing and omics technologies to develop climate-resilient, high-yielding cacao varieties with superior quality traits, contributing significantly to global cacao conservation and improvement programs.

#### KEYWORDS

cacao, fruit, genetic variation, multivariate analysis, phenotype, plant breeding, seed bank

## 1 Introduction

As a cornerstone of global agriculture and rural economies, *Theobroma cacao* L. underpins both commercial chocolate production and the livelihoods of millions of smallholder farmers. Originally domesticated in the Upper Amazon approximately 5,300 years ago, cacao is now cultivated in over 50 countries across Central and South America, Africa, and Asia. West Africa accounts for approximately 70% of the global production (Motamayor et al., 2002; Clement et al., 2010; Díaz-Valderrama et al., 2020). The economic importance of this crop is particularly evident in the chocolate industry's global market value. This reached USD 136.0 billion in 2020 and is projected to grow to USD 192.12 billion by 2028 (Zion Market Research, 2025). Beyond its commercial significance, cacao cultivation supports approximately six million smallholder farmers (Díaz-Valderrama et al., 2020). It simultaneously contributes to environmental sustainability through diverse agroforestry systems. These systems enhance biodiversity and promote ecological balance (Kristanto et al., 2024).

Despite its economic and ecological importance, contemporary cacao cultivation is facing unprecedented challenges that threaten both production stability and genetic diversity. Climate change projections indicate significant shifts in suitable growing regions, whereas emerging diseases and pests pose increasing threats to production (Ceccarelli et al., 2024; Morán et al., 2024). These challenges are particularly concerning, given that over 96% of global cocoa production depends on smallholder farmers, predominantly in West Africa, whose livelihoods are increasingly vulnerable to these threats (Aikpokpodion et al., 2009). Furthermore, the genetic diversity present in wild populations remains inadequately represented in cultivation areas, primarily due to the 'founder effect' during domestication (Aikpokpodion, 2010), limiting genetic basis for adaptation to emerging challenges.

Conservation and characterization of cacao genetic resources have emerged as critical priorities for addressing these challenges. *Ex-situ* germplasm banks serve as essential repositories of genetic diversity and sources of traits to address emerging challenges through breeding programs (Iwaro et al., 2003; Bekele et al., 2021; Ceccarelli et al., 2022). Within these conservation efforts,

morphological characterization continues to play a fundamental role in germplasm evaluation, particularly for highly heritable traits that directly influence yield and quality (Bidot Martínez et al., 2015b; Bekele et al., 2019; Oliva-Cruz et al., 2021; Kongor et al., 2024; Moundanga et al., 2024). The significance of morphological traits in differentiating cacao populations and elucidating genetic relationships has been extensively documented since Engels' foundational work (Engels, 1986), which has been further validated through modern methodologies that integrate morphological and molecular characterization (Crouzillat et al., 1996; Georges et al., 2023; Nieves-Orduña et al., 2024).

The Amazon Basin, as the evolutionary cradle of *T. cacao* L., has exceptional genetic diversity that remains largely untapped for crop improvement (Zhang and Motilal, 2016a). This vast region contains numerous wild cacao populations distributed along intricate river systems that exhibit remarkable morphological and genetic variation (Motamayor et al., 2008). Recent genomic analyses have revealed a deep evolutionary history, with the five major cacao genetic clusters diverging during the early to middle Pleistocene, approximately 1.83 to 0.69 million years ago (Nousias et al., 2024). This extended period of evolution has generated unique allelic combinations in indigenous Amazonian cacao populations, potentially conferring enhanced climate resilience and disease resistance traits that are crucial for modern breeding programs (Zhang et al., 2006; Nieves-Orduña et al., 2023). Within this evolutionarily significant landscape, the Loreto region in the Peruvian Amazon stands out as a significant center of cacao diversity, harboring unique genetic resources that remained largely unexplored for systematic conservation and breeding programs.

To address this critical gap in Amazonian cacao conservation, we established the first comprehensive native cacao germplasm bank in the Loreto Region of the Peruvian Amazon in 2017. Our research had four primary objectives: (1) to develop an *ex situ* conservation system for native cacao genetic resources from this key center of diversity; (2) to systematically document the morphological diversity through a comprehensive set of standardized descriptors across vegetative, floral, and reproductive traits; (3) to identify phenotypically distinct groups and characterize their spatial distribution across the region; and (4) to establish the baseline data necessary for future breeding

programs targeting climate resilience and disease resistance. This integration of *ex situ* conservation with systematic phenotypic evaluation positions this germplasm bank as a vital resource for developing climate-resilient cacao varieties to address future production challenges while preserving the unique genetic heritage of Amazonian cacao.

## 2 Materials and methods

### 2.1 Germplasm bank development

#### 2.1.1 Germplasm bank site preparation

The germplasm bank was strategically established at the Experimental Field “El Dorado” of the Instituto Nacional de Innovación Agraria (INIA), which is characterized by highly weathered Amazonian soils and a humid tropical climate. El Dorado is located at kilometer 25½ of the Iquitos-Nauta Highway (03°57'01" S, 73°24'59" W, 115 m.a.s.l.), San Juan Bautista District, Peru (Supplementary Figure S1). The site is characterized by challenging edaphic conditions. Soil analysis at 0–40 cm depth revealed a clayey texture (42.64% clay, 32.80% sand, and 24.56% silt) with a 1.28 g/cm<sup>3</sup> bulk density. This soil analysis was conducted prior to establishment to characterize baseline edaphic conditions, though subsequent systematic soil analyses were not included in the methodology. The soil is extremely acidic (pH 4.39) with poor fertility, characterized by low organic matter content (1.67%), low nitrogen (0.08%), and limited available phosphorus (3.81 g/kg). The exchangeable bases showed low values for calcium (1.77 Cmol<sup>+</sup>/L), magnesium (0.15 Cmol<sup>+</sup>/L), and potassium (0.04 Cmol<sup>+</sup>/L), with total exchangeable bases of 1.96 Cmol<sup>+</sup>/L. The soil presented high aluminum saturation (78.08%) and low electrical conductivity (0.03 mmhos/cm), indicating negligible salinity effects. These soil conditions are typical of the highly weathered Amazonian soils. The local climate is classified as humid tropical, with relative humidity exceeding 80%, a mean annual temperature of 26°C, and annual rainfall between 2,500–3,000 mm. While basic climate data was recorded during the study period based on the nearest weather station, systematic on-site environmental monitoring was not part of the study design.

During its early development stages, a temporary nursery (220 m<sup>2</sup>) was established in January 2016 to produce the shade species required by cacao, a semi-umbrophytic species. Two shade species were established from January to August 2016: banana (*Musa × paradisiaca* L.) at 3 m × 3 m spacing and ice cream bean (*Inga edulis* Mart.) at 12 × 12 m intervals (Supplementary Figure S2). Planting holes measuring 0.25 m (length) × 0.25 m (width) × 0.25 m (depth) were prepared with basal fertilization consisting of 200 g rock phosphate and 2 kg poultry manure.

#### 2.1.2 Rootstock development

Rootstocks were developed from August to December 2016 using the IMC 67 (Iquitos Mixed Calabacillo) cacao clone, a native cultivar from the Loreto region, selected for its superior agronomic characteristics. This genotype exhibited resistance to major fungal

diseases (moniliasis, brown rot, and witches' broom) and tolerance to flooding conditions. This demonstrates early flowering with high-quality fruit production and efficient pollen donation capacity.

The growing substrate was prepared following an established protocol (N'zi et al., 2023) by thoroughly mixing agricultural soil (2:1:1 ratio), wood organic matter, and poultry manure. This mixture was used to fill black polyethylene bags (15 × 30 cm). To induce germination of cocoa seeds from IMC 67, they were soaked for 24 h, sheltered for 24 h, and germinated at 48 h when the radicle appeared. Germinated cocoa beans were individually planted in the prepared substrate, and the plants were protected in a shade house using a plastic sunshade mesh (Raschel®) under 80% shading (Supplementary Figure S3). For phytosanitary control, the organic insecticide rotenone (rotenone) was applied at a dose of 0.2%, adding foliar nutrient and 0.2% and agricultural adherent at 0.05%. Manual weeding was performed every two weeks to eliminate competing vegetation and optimize seedling development.

After five months of growth, rootstock candidates were carefully selected based on uniform morphological characteristics, including stem diameter (8.77 ± 0.88 mm), plant height (78.74 ± 4.22 cm), and leaf number (25.28 ± 1.97), ensuring standardization for the subsequent grafting process. These rootstocks were transplanted into the definitive field in January 2017 and arranged in a systematic grid pattern with 3 × 3 m spacing to ensure optimal growth conditions.

#### 2.1.3 Plant material collection

Native cacao scions were collected between February and August 2017 from 140 locations distributed across eight provinces of the Loreto region of Peru (Figure 1). The collection sites were strategically selected based on three criteria: (1) preliminary ethnobotanical surveys identifying areas with traditional cacao cultivation, (2) reported morphological distinctiveness by local farmers, and (3) maximizing geographic representation across the region's river basins. This approach aimed to capture the greatest possible range of phenotypic diversity while ensuring comprehensive spatial coverage of the Loreto region (Supplementary Table S1). The study region is situated in the lowland Amazon basin, characterized by predominantly flat terrain interspersed with gentle hills, with elevations ranging from 70 m to 220 m above sea level. The soils are primarily of fluvial origin, exhibiting a clayey texture and marked acidity, with notably low fertility in older geological formations. The climate is classified as tropical humid, with minimal seasonal temperature variation (mean 27°C; range 20–33°C), substantial annual precipitation (2,500–3,000 mm), and consistently high relative humidity (87–90%) throughout the year (Dourojeanni, 2021).

Field collection was conducted by a trained team of technicians who accessed the cacao trees using climbing techniques when necessary to reach the optimal scion material. Five different cacao plants were sampled at each collection site. Ten healthy scions from each plant were collected, specifically from the middle third of the canopy, to ensure uniformity in physiological maturity. The selected scions were approximately 30 cm in length and 1 cm in diameter, displaying a dark brown coloration indicating proper

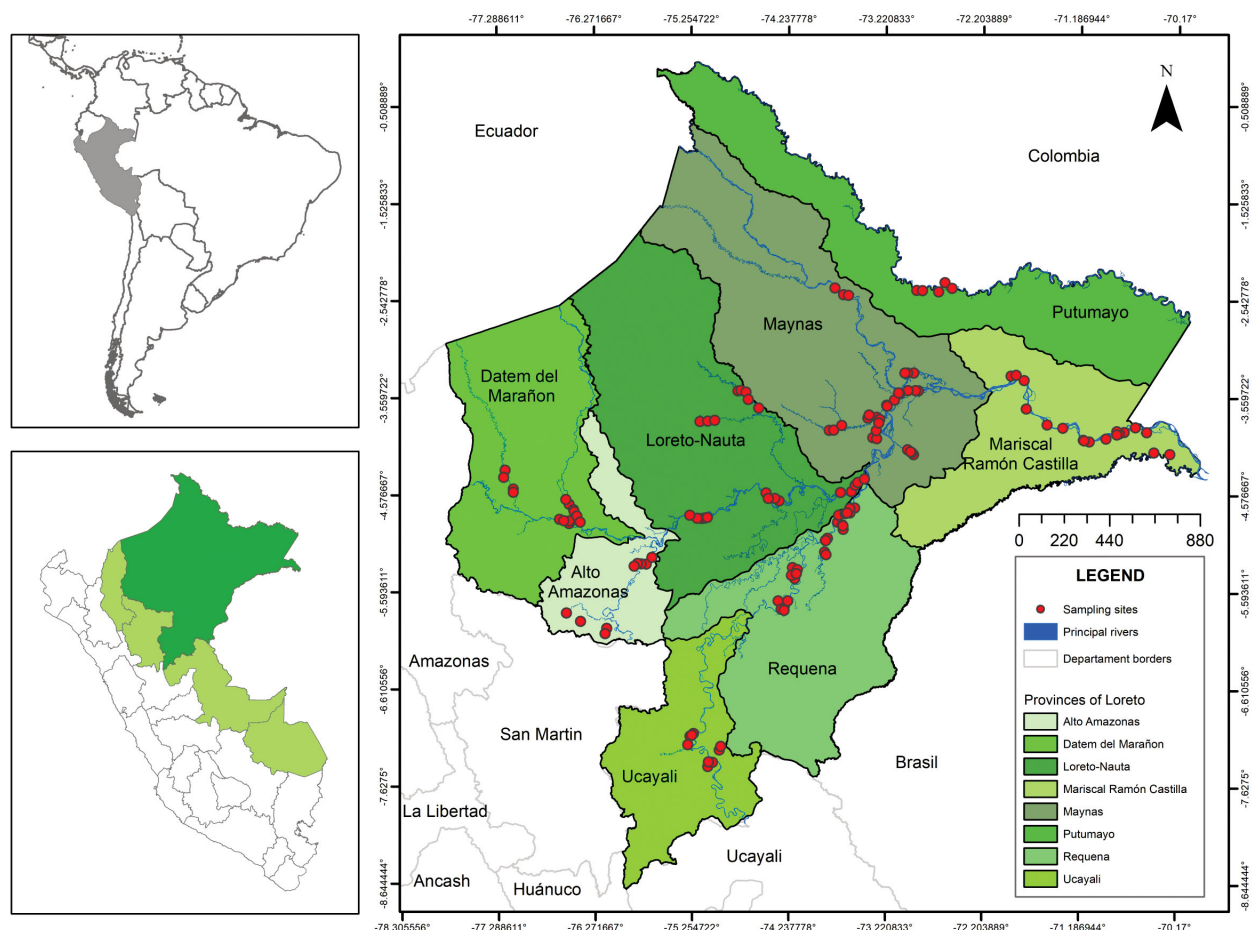


FIGURE 1

Geographic distribution of native *Theobroma cacao* L. collection sites in the Loreto region, Peru. Left panels: Location of Peru in South America (top) and the Loreto region within Peru (bottom). Main panel: Map showing the distribution of 140 sampling sites (red dots) across the eight provinces of Loreto, with principal river networks (blue lines) and provincial boundaries. Samples were collected between February and August 2017, spanning elevations from 70 to 220 m above sea level. Scale bar represents distance in kilometers.

lignification, bearing 14–16 mature leaves, and having 2–4 viable dormant buds. Only materials free from visible disease symptoms, pest damage, mechanical injuries, and with appropriate physiological maturity (as evidenced by leaf texture, stem hardness, and tissue turgidity) were harvested using sterilized pruning clippers. Immediately after collection, individual scions were wrapped in moistened newspapers to prevent desiccation and labeled with their collection codes and source plant identifiers. The bundled materials were then placed in portable coolers to maintain the optimal temperature and humidity and protect them from light during transportation ([Supplementary Figure S4](#)). This storage method preserves scion viability during collection expeditions, which sometimes extends to several days in remote areas.

#### 2.1.4 Grafting and germplasm bank establishment

Grafting procedures followed established protocols ([Munjuga et al., 2013](#)). Selected scions were sectioned into 2–4 pieces (15 cm length), ensuring that each segment contained 2–4 healthy, viable buds. Two grafting techniques were used: top-cleft grafting and side

grafting. Top-cleft grafting consisted of making a clean transverse cut across the rootstock using sterilized pruning shears, creating a 3–4 cm vertical slit in the center of the rootstock, beveling the scion base into a wedge shape, inserting it into the prepared slit to ensure cambial alignment, and securing the graft union with transparent plastic film to maintain humidity and promote healing. For side grafting, the process involved making a 3–4 cm downward-slanting cut on the side of the rootstock, creating a small flap by making a second cut at the base of the first cut, preparing the scion with a matching slanted cut, inserting it under the flap to ensure cambial alignment, and securing the graft union with transparent plastic film to maintain humidity and promote healing ([Supplementary Figure S5](#)). Graft unions were monitored weekly, and success rates were evaluated three weeks post-grafting. Both grafting techniques, top-cleft and side grafting, demonstrated exceptional efficiency, with 100% successful establishment for all 700 grafting attempts. Weekly monitoring revealed that by the third week post-grafting, all unions displayed successful integration as evidenced by bud sprouting and subsequent shoot development. This uniformly

high success rate across both methods provides strong validation of the grafting protocols employed and demonstrates the excellent compatibility between the IMC 67 rootstock and the diverse native scions collected from the Loreto region.

The grafting phase occurred between February and August 2017, and the germplasm bank was established in September 2017 at the Experimental Field “El Dorado”. The collection encompasses 10,000 m<sup>2</sup> and houses 140 unique accessions, with five different plants per accession, totaling 700 plants (Supplementary Figures S1, S5). Plants were distributed in a systematic grid pattern with 3 × 3 m spacing to optimize growth conditions and facilitate maintenance operations. Three years after the establishment of the germplasm bank, temporary shade species were removed as a preventive phytosanitary measure to reduce humidity and mitigate the risk of fungal pathogens that could compromise fruit development.

Follow-up soil analyses conducted 24 months after germ plasm bank establishment showed modest improvements in soil properties, with organic matter increasing to 2.35%, available phosphorus to 5.62 g/kg, and a slight decrease in aluminum saturation to 70.45%, demonstrating partial amelioration of the challenging edaphic conditions through the implemented management practices.

## 2.2 Phenotypic characterization

Phenotypic evaluation was conducted from April 2018 to December 2021, beginning with accessions 027 and 051–14 months after grafting. The characterization included 402 plants, comprising of three representatives from 134 accessions. Six accessions were excluded due to the absence of flowering and fruiting during the evaluation period. The assessment employed 36 standardized descriptors (25 quantitative and 11 qualitative) that are widely accepted in cacao research (Bekele et al., 2019; Gopalulchan et al., 2019), ensuring data compatibility and reliability across different germplasm studies.

The evaluated descriptors encompassed four main organs: the leaves, flowers, fruits, and seeds. The leaf measurements included length, width, and petiole length (Supplementary Tables S1, S2). Floral characteristics were assessed using 10 quantitative parameters (pedicel length, sepal dimensions, petal dimensions, filament length, staminode length, style length, and ovary measurements) and two qualitative traits (presence of anthocyanin in filaments and sepals). Fruit characterization involved both quantitative (weight, length, diameter, pericarp weight, pericarp thickness, and furrow depth) and qualitative (ripe fruit color, unripe fruit color, shape, apex form, basal constriction, and rugosity) parameters. Seed traits were evaluated using seven quantitative measures (number, fresh and dried weights, length, width, and thickness) and two qualitative descriptors (transversal and longitudinal shape).

All measurements were performed according to standardized protocols (Engels, 1983; Bekele and Bekele, 1996; Bekele et al., 2019) to minimize measurement errors and ensure data consistency, facilitating comparative analyses with other cacao germplasm collections.

## 2.3 Statistical analysis of phenotypic data

### 2.3.1 Data processing, basic statistical analysis and distance computation

The analytical process began with the systematic data preprocessing of the phenotypic dataset. Raw data were initially imported using the Pandas library from a specified sheet in the Excel workbook. Column names were standardized by stripping any leading or trailing whitespace to ensure consistency for subsequent processing, followed by a thorough quality assessment to identify missing values and potential inconsistencies. Missing values were systematically addressed using a forward-fill method to maintain data integrity while minimizing information loss. Following data cleaning, a comprehensive exploratory data analysis was conducted to assess data quality and examine the distribution patterns of all variables (Supplementary Table S3). Basic descriptive statistical analysis (Supplementary Table S4, Supplementary Figure S6) and Rho Spearman correlation analysis (Supplementary Table S5) of quantitative phenotypic descriptors. Given the mixed nature of the phenotypic descriptors (both quantitative and qualitative), the Gower distance metric was employed to compute dissimilarity between samples (Supplementary Table S6). This metric is well suited for datasets with mixed data types and was calculated using the Gower package. The resulting distance matrix provided a framework for downstream analysis. The Mantel test was also performed, considering the Gower distance matrix and geographic distances.

### 2.3.2 Multivariate analysis and clustering

Following Gower distance matrix computation, dimensionality reduction was performed using Uniform Manifold Approximation and Projection (UMAP). UMAP was configured to reduce the data to three dimensions, facilitating the creation of interactive 3D visualizations. Notably, the precomputed Gower distance matrix was used as the metric to ensure that the mixed data types were appropriately considered (Supplementary Table S6). After dimensionality reduction, the K-means clustering algorithm was applied to the UMAP embeddings. Five distinct clusters ( $k = 5$ ) were identified as optimal clustering solutions (Supplementary Figure S7), representing different phenotypic groups. The resulting cluster assignments were incorporated into the dataset (Supplementary Table S1), allowing for stratified analysis and visualization of group-specific characteristics. In addition, interactive 3D scatter plots were generated using Plotly to display UMAP embeddings on a percentage scale. For each phenotypic group, the centroids and their corresponding 95% confidence ellipsoids were computed and visualized. These visualizations provided insights into the spatial distribution and clustering patterns observed within the dataset (Video S1).

To compare phenotypic groups, the statistical evaluation proceeded systematically, with ANOVA and Kruskal-Wallis tests, to identify significant differences in quantitative phenotypic descriptors (Supplementary Table S7). Where significant differences were detected ( $p < 0.05$ ), pairwise comparisons were performed using HSD Tukey or Mann-Whitney U tests. These

comparisons were adjusted using Bonferroni correction to control for multiple testing effects, ensuring statistical rigor in the identification of significant differences between specific phenotypic groups. For qualitative phenotypic descriptors, the initial analysis involved chi-square tests of independence to examine the relationships between phenotypic groups. Fisher's exact test was used for variables with zero values in the contingency tables to ensure accurate probability calculations. The significance level was set at  $p = 0.05$ . *Post hoc* analyses and Dunn's test with Bonferroni correction were performed to conduct multiple comparisons to identify specific group differences (Supplementary Table S8).

### 2.3.3 Diversity metrics analysis

Diversity analysis began by computing the Shannon-Weaver (H) diversity index for all samples and each phenotypic group. A bootstrap resampling procedure was implemented with 10000 iterations for each diversity metric within the phenotypic groups to ensure statistical robustness. From these iterations, 95% confidence intervals were calculated to provide reliable estimation ranges for the diversity measure. To compare the phenotypic groups, the chi-square test and *post hoc* test were conducted (Supplementary Table S8).

The final phase involved comprehensive documentation and visualization of the analytical results. Detailed boxplots were generated to illustrate the distribution patterns of diversity metrics across phenotypic groups, incorporating the mean markers and interquartile ranges. Separate visualizations were created for each key diversity index, emphasizing the specific patterns and relationships. All statistical outputs, including bootstrap confidence intervals and test results, were systematically compiled and exported in the Excel format. High-resolution versions of all visualizations were preserved as JPEG files, ensuring the publication-quality documentation of the findings.

Statistical analyses were performed using Python (version 3.11.6) with specialized packages: *scipy.stats* for core statistical computations, *scikit-posthoc* for *post-hoc* analyses, and *plotly* for data visualization. All computations were executed using the Julius computational platform (<https://julius.ai/>), which ensured reproducibility and standardization of the analytical workflow.

## 3 Results

### 3.1 Germplasm bank development

The collection effort in the Loreto region of Peru yielded a comprehensive geographical representation, successfully gathering genetic material from 140 distinct locations distributed across eight provinces and 15 river basins (Figure 1; Supplementary Table S1). From each location, scions were collected from five individual plants, culminating in 700 source plants. The implemented collection and transportation protocol proved highly effective, with the combination of moistened newspaper wrapping and portable coolers maintaining scion viability, even during extended

collection periods in remote areas (Supplementary Figure S4). Upon arrival at the grafting facility, quality assessment revealed that over 95% of the collected material maintained optimal physiological conditions, exhibiting firm tissue consistency, healthy green foliage, and no signs of dehydration.

Both grafting techniques, top cleft and side grafting, demonstrated exceptional success rates in the establishment phase (Supplementary Figure S5). Within three weeks of grafting, all unions displayed successful integration, as evidenced by bud sprouting, and subsequently developed into viable plants with robust shoot growth. This uniform success in graft establishment laid a strong foundation for the subsequent field-establishment phase.

Field establishment at the Experimental Field "El Dorado" achieved optimal results, with all transplanted specimens successfully adapting to their new environment (Supplementary Figure S1). The implementation of a systematic grid pattern with a  $3 \times 3$  m spacing proved effective, allowing uniform establishment across the entire 10,000 m<sup>2</sup> area. Careful spacing design facilitates proper plant development and efficient maintenance.

The current status of the germplasm bank reflects complete preservation of the initially collected diversity (Figure 2). All 140 accessions maintained their original complement of five plants per accession, resulting in a total of 700 thriving individuals, representing 100% survival of the planted material. This exceptional establishment success ensures a comprehensive representation of the genetic diversity collected from the 15 surveyed river basins, fulfilling the bank's primary objective of *ex situ* conservation of the native cacao germplasm from the Loreto region.

## 3.2 Phenotypic characterization

### 3.2.1 Characterization based on quantitative phenotypic descriptors

Quantitative analysis revealed pronounced morphological diversity across vegetative, floral, and reproductive traits, with fruit characteristics demonstrating the highest variability (Table 1; Supplementary Table S3, Supplementary Figure S6). Regarding vegetative characteristics, leaf length averaged  $35.84 \pm 5.45$  cm (CV = 15.21%) with values ranging from 23.30 to 54.20 cm, while leaf width showed a mean of  $12.46 \pm 1.93$  cm (CV = 15.52%) with a range of 7.60 to 19.20 cm. Petiole measurements exhibited high variability, with petiole length showing the highest coefficient of variation (CV = 32.52%) among vegetative traits, ranging from 1.10 to 9.00 cm with a mean of  $2.41 \pm 0.78$  cm.

Floral characteristics demonstrated moderate to high variability. Sepal dimensions showed similar coefficients of variation (CV  $\approx$  15–16%), with sepal length averaging  $0.70 \pm 0.11$  cm and width  $0.23 \pm 0.04$  cm. Petal measurements revealed a higher variability in length ( $0.33 \pm 0.05$  cm) than width ( $0.20 \pm 0.02$  cm). Notably, the filament length displayed considerable variation (CV = 30.24%), ranging from 0.11 to 0.80 cm. The ovary dimensions were relatively consistent, with length and width showing CVs values of 18.43% and 15.43%, respectively.



Plants of *Theobroma cacao* in the Developed Germplasm Bank from the Loreto Region

#### FIGURE 2

Overview of the established *Theobroma cacao* L. germplasm bank at the Experimental Field "El Dorado," Loreto, Peru. (A) Entrance signage indicating the bank's establishment in 2017 with 140 accessions; (B) Details of mature cacao trees with developing pods, demonstrating successful establishment; (C) Panoramic view showing a systematic 3 x 3 m grid arrangement of cacao trees, illustrating the spatial organization of the 10,000 m<sup>2</sup> collection.

The fruit characteristics exhibited the highest overall variation among all traits measured. Fruit weight showed remarkable variation (CV = 48.50%), ranging from 59.00 to 1,655.00 g with a mean of  $470.22 \pm 228.67$  g. The fruit length and diameter were more consistent, with CVs of 18.82% and 15.82%, respectively. The pericarp weight also showed substantial variation (CV = 50.82%), ranging from 44.00 to 1,315.00 g. Pericarp thickness and furrow depth showed moderate variation, with CVs values of 25.24% and 22.94%, respectively.

Seed characteristics demonstrated notable variations, particularly in fresh weight measurements. Unpeeled fresh weight

ranged from 8.00 to 335.00 g (CV = 49.04%), while peeled fresh weight varied from 0.40 to 3.00 g (CV = 30.75%). The number of seeds per fruit averaged  $34.51 \pm 11.30$ , ranging from 5 to 57 seeds. Seed dimensions showed moderate variation, with length, width, and thickness having CVs values of 13.63%, 18.68%, and 14.57%, respectively.

The pod index analysis revealed substantial variation among the accessions, with values ranging from 19.49 to 1,666.67 and a mean of  $107.14 \pm 155.17$ . The high coefficient of variation (CV = 144.82%) and positive skewness (7.26) indicate a wide dispersion and right-skewed distribution of pod index values in the germplasm

**TABLE 1** Descriptive statistics of quantitative phenotypic descriptors measured in *Theobroma cacao* L. accessions of the germplasm bank from the Loreto Region, including vegetative, floral, fruit, and seed characteristics.

| Quantitative Phenotypic Descriptor | Mean $\pm$ SD       | Min   | Max      | Skewness | Kurtosis | CV (%) |
|------------------------------------|---------------------|-------|----------|----------|----------|--------|
| Leaf length (cm)                   | 35.84 $\pm$ 5.45    | 23.30 | 54.20    | 0.46     | 0.18     | 15.21  |
| Leaf width (cm)                    | 12.46 $\pm$ 1.93    | 7.60  | 19.20    | 0.58     | 0.86     | 15.52  |
| Petiole length (cm)                | 2.41 $\pm$ 0.78     | 1.10  | 9.00     | 3.56     | 21.33    | 32.52  |
| Pedicle length (cm)                | 1.38 $\pm$ 0.37     | 0.50  | 2.64     | 0.29     | 0.02     | 27.09  |
| Sepal length (cm)                  | 0.70 $\pm$ 0.11     | 0.30  | 1.40     | 0.85     | 4.04     | 15.75  |
| Sepal width (cm)                   | 0.23 $\pm$ 0.04     | 0.13  | 0.37     | 0.65     | -0.10    | 16.66  |
| Petal length (cm)                  | 0.33 $\pm$ 0.05     | 0.18  | 0.75     | 3.02     | 20.02    | 15.79  |
| Petal width (cm)                   | 0.20 $\pm$ 0.02     | 0.10  | 0.35     | 0.40     | 11.29    | 12.31  |
| Filament length (cm)               | 0.21 $\pm$ 0.06     | 0.11  | 0.80     | 4.81     | 34.07    | 30.24  |
| Staminode length (cm)              | 0.58 $\pm$ 0.09     | 0.16  | 0.85     | -0.65    | 3.56     | 15.73  |
| Style length (cm)                  | 0.19 $\pm$ 0.04     | 0.10  | 0.30     | -0.02    | -0.13    | 22.53  |
| Ovary length (cm)                  | 0.14 $\pm$ 0.03     | 0.08  | 0.20     | 0.19     | -0.21    | 18.43  |
| Ovary width (cm)                   | 0.11 $\pm$ 0.02     | 0.07  | 0.20     | 1.29     | 1.65     | 15.43  |
| Fruit weight (g)                   | 470.22 $\pm$ 228.07 | 59.00 | 1,655.00 | 1.74     | 5.78     | 48.50  |
| Fruit length (cm)                  | 16.21 $\pm$ 3.05    | 7.20  | 25.20    | 0.35     | 0.56     | 18.82  |
| Fruit diameter (cm)                | 8.38 $\pm$ 1.32     | 4.86  | 12.50    | 0.23     | 0.22     | 15.82  |
| Pericarp weight (g)                | 360.83 $\pm$ 183.37 | 44.00 | 1,315.00 | 1.76     | 5.54     | 50.82  |
| Pericarp thickness (cm)            | 1.61 $\pm$ 0.41     | 0.53  | 3.00     | 0.16     | -0.04    | 25.24  |
| Furrow depth (cm)                  | 0.97 $\pm$ 0.32     | 0.30  | 2.50     | 1.19     | 2.29     | 32.94  |
| Seed number                        | 34.51 $\pm$ 11.30   | 5.00  | 57.00    | -0.51    | -0.28    | 32.74  |
| Unpeeled fresh weight (g)          | 92.09 $\pm$ 45.16   | 8.00  | 335.00   | 1.21     | 3.61     | 49.04  |
| Peeled fresh weight (g)            | 1.36 $\pm$ 0.42     | 0.40  | 3.00     | 0.59     | 1.20     | 30.75  |
| Seed length (cm)                   | 2.30 $\pm$ 0.31     | 1.53  | 3.53     | 0.48     | 0.64     | 13.63  |
| Seed width (cm)                    | 1.16 $\pm$ 0.22     | 0.68  | 2.54     | 2.53     | 12.84    | 18.68  |
| Seed thickness (cm)                | 0.81 $\pm$ 0.12     | 0.47  | 1.14     | -0.04    | 0.05     | 14.57  |
| Pod index*                         | 107.14 $\pm$ 155.17 | 19.49 | 1,666.67 | 7.26     | 63.51    | 144.82 |

\*Was calculated using the following equation (Iwaro et al., 2003):  $Pod\ index = \frac{1000}{Average\ bean\ number \times Average\ bean\ weight\ (g)}$ .

collection. This extreme variation, further emphasized by the high kurtosis value (63.51), suggests the presence of both highly efficient and less efficient accessions regarding the relationship between pod fresh weight and dried bean yield.

Correlation analysis among the quantitative traits revealed complex relationships between vegetative, floral, and reproductive characteristics (Supplementary Table S5). The correlations were classified into very strong ( $\rho \geq 0.8$ ), strong ( $0.6 \leq \rho < 0.8$ ), moderate ( $0.4 \leq \rho < 0.6$ ), weak ( $0.2 \leq \rho < 0.4$ ), and very weak ( $\rho < 0.2$ ) relationships.

The vegetative traits showed significant positive correlations with each other. Leaf length and width demonstrated a strong positive correlation ( $\rho = 0.64$ ,  $p < 0.05$ ), whereas pedicle length was moderately correlated with both petiole length ( $\rho = 0.49$ ,  $p < 0.05$ )

and leaf measurements. These relationships suggest coordinated growth patterns in the vegetative structures.

Floral characteristics exhibited several significant interconnections. Notable correlations were observed between sepal dimensions and other floral traits, with sepal length showing strong positive correlations with petal ( $\rho = 0.55$ ,  $p < 0.05$ ) and staminode ( $\rho = 0.42$ ,  $p < 0.05$ ) lengths. Petal measurements were consistently correlated with other floral structures, particularly with the relationship between petal width and filament length ( $\rho = 0.41$ ,  $p < 0.05$ ).

The strongest correlations were observed among the fruit-related traits. Fruit weight was strongly positively correlated with pericarp weight ( $\rho = 0.77$ ,  $p < 0.05$ ) and unpeeled fresh weight ( $\rho = 0.73$ ,  $p < 0.05$ ). The fruit dimensions (length and diameter) were strongly correlated with each other ( $\rho = 0.56$ ,

$p < 0.05$ ) and moderately correlated with most seed characteristics. Interestingly, the pod index showed significant negative correlations with several traits, particularly with fruit weight ( $\rho = -0.69$ ,  $p < 0.05$ ) and pericarp weight ( $\rho = -0.60$ ,  $p < 0.05$ ).

Seed characteristics showed moderate to strong correlations. Seed dimensions were positively correlated, with the strongest relationship observed between seed length and width ( $\rho = 0.68$ ,  $p < 0.05$ ). The seed number was significantly positively correlated with fruit weight ( $\rho = 0.62$ ,  $p < 0.05$ ) and pericarp weight ( $\rho = 0.55$ ,  $p < 0.05$ ), suggesting that larger fruits typically contain more seeds.

### 3.2.2 Characterization based on qualitative phenotypic descriptors

Qualitative phenotypic characterization of *T. cacao* accessions from the Loreto Region germplasm bank comprised 11 descriptors across floral, fruit, and seed structures (Figure 3; Supplementary Table S8). These descriptors included anthocyanin pigmentation in three floral organs (filaments, ovaries, and sepals), six fruit characteristics (color at maturity and immaturity, shape, apex form, basal constriction, and rugosity), and two seed morphological traits (transversal and longitudinal shapes). Analysis of these traits revealed distinct patterns of variation and trait frequencies that characterize the morphological diversity of this germplasm collection.

Floral characteristics of the germplasm collection showed distinctive patterns of anthocyanin pigmentation across different structures. In the filaments, there was a relatively balanced distribution between absent (54.73%) and present (45.27%) anthocyanin pigmentation. However, ovary pigmentation was predominantly absent, with 98.26% of accessions showing no anthocyanin coloration, and only 1.74% presenting this trait. Similarly, sepal pigmentation was largely absent (83.33%), with only 16.67% of the accessions displaying anthocyanin.

Fruit color exhibited marked variation between mature and immature stages. Ripe fruits were predominantly yellow (95.77%), with very low frequencies of orange (2.24%) and green (1.99%) coloration. In contrast, unripe fruits showed a strong prevalence of green coloration (85.07%), followed by pigmented variants (13.18%) and a small proportion of red fruits (1.75%). The diversity in fruit morphology was evident across several characteristics. Fruit shape was dominated by oblong forms (64.16%), followed by elliptical (27.17%), with lower frequencies of obovate (4.73%), spherical (2.72%), and ovate (1.22%) shapes.

The fruit apex and basal characteristics displayed considerable variation within the collection. The fruit apex form was predominantly obtuse (43.03%) or attenuated (37.81%), with lower frequencies of the apezonate (14.68%) and acute (4.48%) forms. Regarding basal constriction, intermediate forms were the most common (54.73%), followed by absent (33.83%), and strong (11.44%) constrictions. The fruit surface texture analysis revealed a clear predominance of intermediate rugosity (76.12%), with lower proportions of rugged (15.67%) and smooth (6.47%) surfaces.

Seed morphology demonstrated distinct patterns in both transversal and longitudinal dimensions. Intermediate shapes were highly prevalent in the transversal section (70.64%), followed by flattened (22.89%), and rounded (6.47%) forms. The

longitudinal seed shapes showed a more balanced distribution, with irregular forms being the most frequent (43.03%), followed by oblong (30.35%), ovate (21.40%), and elliptical (5.22%) shapes. This diversity in seed morphology suggests significant genetic variation in the germplasm collection.

### 3.2.3 Phenotypic groups and geographic distribution pattern of *T. cacao* accessions

The phenotypic diversity of *T. cacao* accessions in the Loreto Region germplasm bank was comprehensively analyzed using Uniform Manifold Approximation and Projection (UMAP), revealing distinct patterns of morphological variation across the collected germplasm. Three-dimensional UMAP analysis successfully resolved five distinct phenotypic groups while preserving both the local and global structures in high-dimensional morphological data. The three UMAP dimensions collectively explained a substantial proportion of the total phenotypic variance, with UMAP 1 and UMAP 2 accounting for 33.3% and 37.1% of the variation, respectively, whereas UMAP 3 contributed an additional 29.6% (Figure 4A; Supplementary Table S6). The three-dimensional projection revealed intricate spatial relationships among the phenotypic groups, with phenotypic groups 2 and 4 showing notable proximity in the central region of the UMAP space, whereas phenotypic groups 1 and 5 occupied more peripheral positions (Video S1). The observed pattern of cluster distribution and overlap suggests the existence of both discrete phenotypic groups and transitional morphotypes, reflecting the complex nature of the phenotypic variation in this germplasm bank. Notably, while the phenotypic groups showed clear separation in certain regions of the UMAP space, the presence of overlapping boundaries between adjacent phenotypic groups, particularly evident at the interface between phenotypic groups 2, 3, and 4, indicated potential phenotypic continuity between these groups (Figure 4A; Video S1). This pattern is consistent with the inherent diversity of cacao populations in the Loreto Region and suggests the presence of distinct morphological variants, while acknowledging the existence of intermediate forms.

The geographic distribution of the five phenotypic groups identified through the UMAP analysis revealed distinct spatial patterns across the river systems and political provinces of the Loreto Region (Figure 4B). Phenotypic group 2 (blue) exhibited a notable concentration along the northeastern boundaries, particularly in the Putumayo and Mariscal Ramón Castilla provinces, while maintaining sporadic presence in central Maynas. Phenotypic group 1 (red) displayed the most widespread distribution, with representation across all provinces and a higher density along riverine corridors in Loreto-Nauta and central Maynas. Phenotypic groups 3 and 4 (green and pink, respectively) demonstrated more localized distributions, primarily clustered in the central portions of the Maynas Province. Phenotypic group 5 (orange) showed an intermediate distribution pattern, with its presence in both the central and eastern portions of the region, frequently occurring in areas where other phenotypic groups were also found. Spatial analysis revealed several zones of phenotypic overlap, particularly along major river systems where

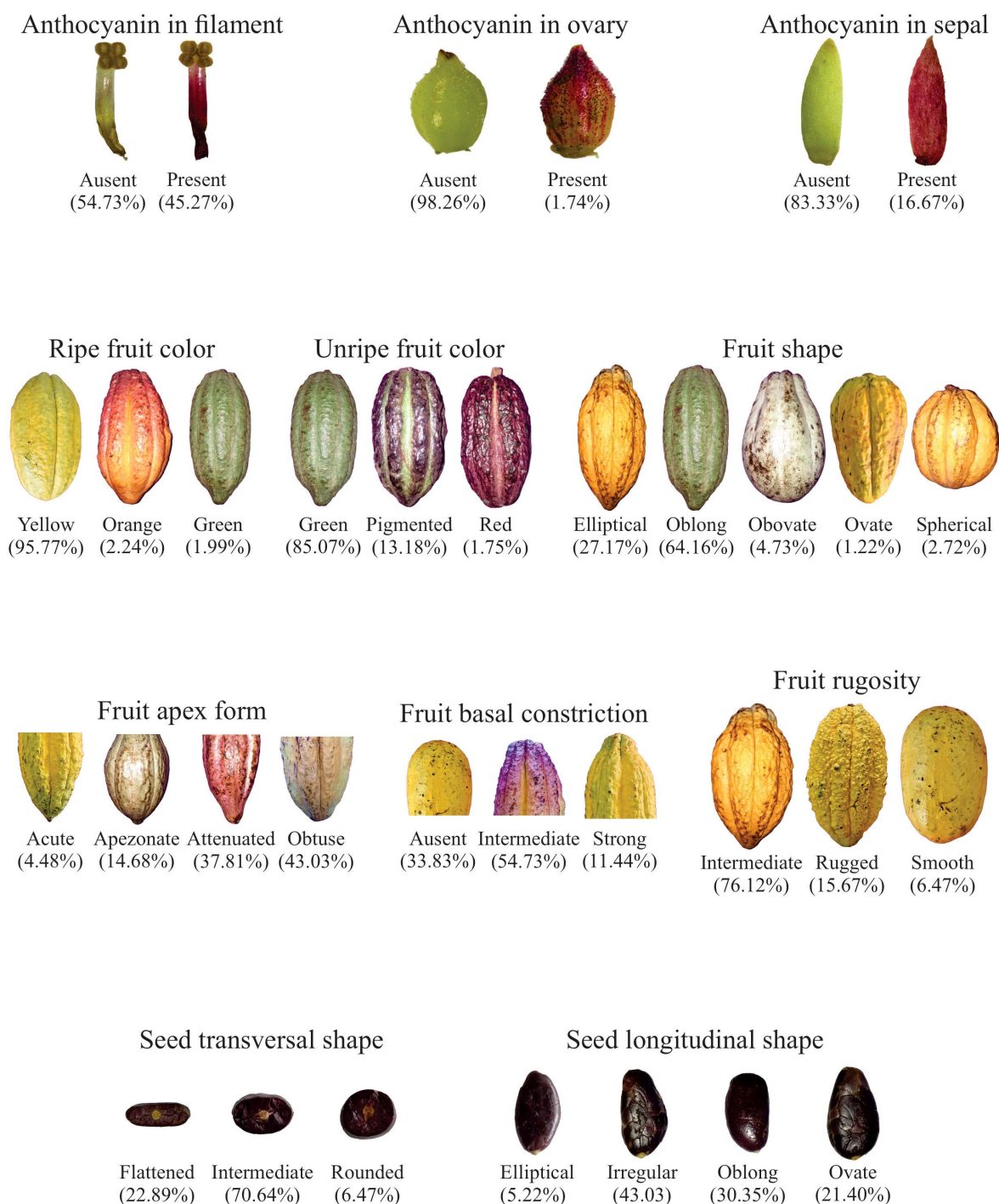


FIGURE 3

Qualitative phenotypic descriptors observed in *Theobroma cacao* L. accessions of the germplasm bank from the Loreto Region. The figure shows the variation and frequency distribution (%) of 11 morphological descriptors: anthocyanin pigmentation in floral structures (filament, ovary, and sepal), fruit characteristics (ripe fruit color, unripe fruit color, fruit shape, fruit apex form, fruit basal constriction, and fruit rugosity), and seed morphology (transversal and longitudinal shapes). Each trait category was accompanied by representative photographs and their corresponding frequency percentages on the germplasm bank.

multiple groups coexist in close proximity. While UMAP analysis demonstrated clear phenotypic differentiation among groups, their spatial arrangement showed no strict geographic boundaries, with multiple phenotypic groups often present within the same localities.

Additionally, the relationship between phenotypic differentiation (Gower distance) and geographic distance was examined using the Mantel test analysis. The results revealed a statistically significant ( $p = 0.0017$ ) but weak correlation between the phenotypic (Gower

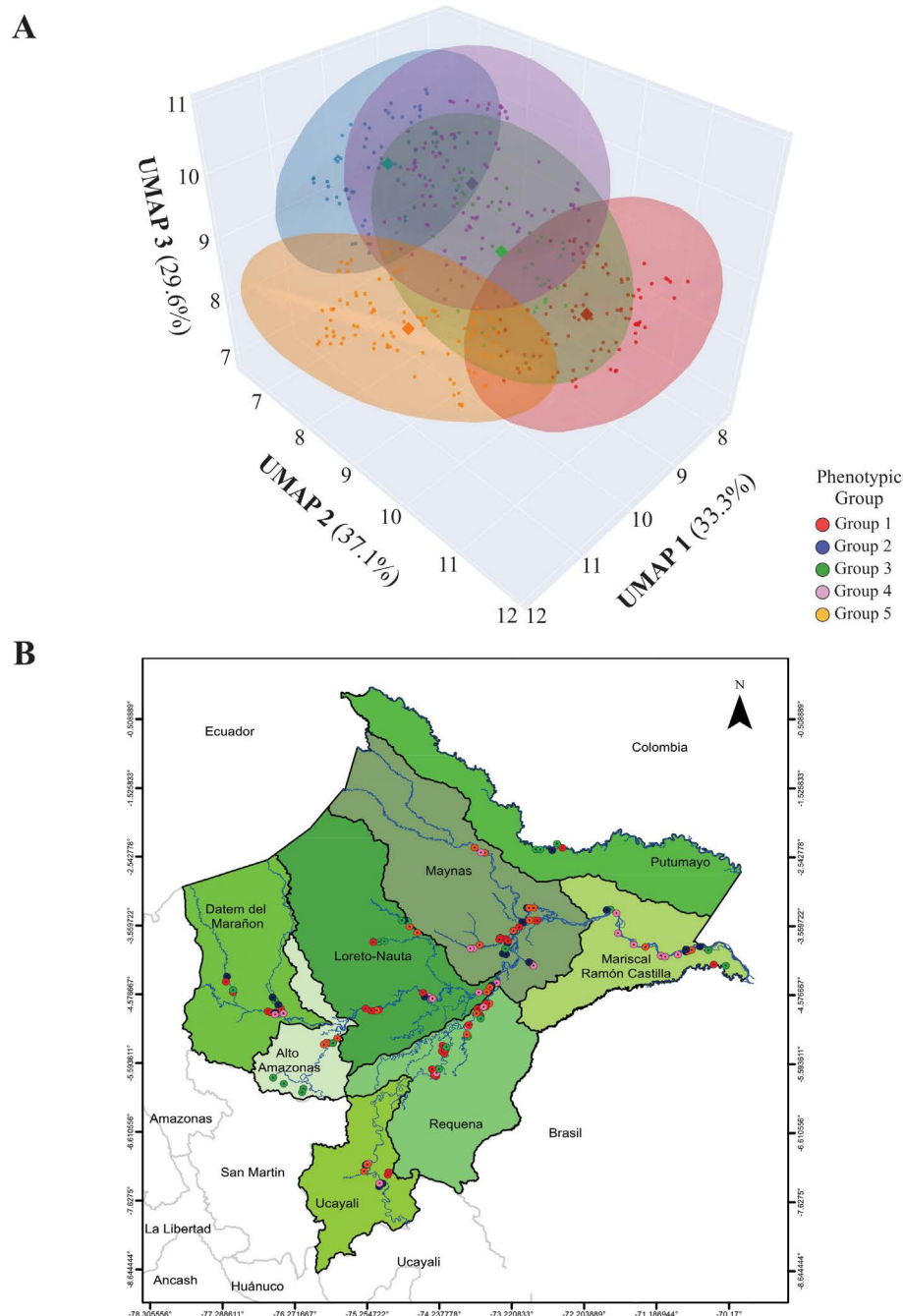


FIGURE 4

Phenotypic diversity analysis and geographic distribution of *Theobroma cacao* L. accessions from the Loreto Region germplasm bank. **(A)** Three-dimensional Uniform Manifold Approximation and Projection (UMAP) visualization showing clustering of accessions into five phenotypic groups based on quantitative and qualitative traits. The percentage of explained variance is shown for each UMAP dimension (UMAP 1:33.3%; UMAP 2:37.1%). **(B)** Geographical distribution of phenotypic groups across the Loreto Region, Peru. Each point represents an individual accession, color-coded according to their corresponding phenotypic group (group 1, red; group 2, blue; group 3, green; group 4, purple; and group 5, orange).

distance) and geographic distances ( $r = 0.0661$ ), with spatial separation accounting for 0.66% of the observed phenotypic variation (Supplementary Figure S8). A hexagonal heatmap visualization demonstrated that the highest density of pairwise comparisons occurred at geographic distances of 200–300 km and Gower distances of 0.15–0.25, represented by yellow-green regions. Throughout the distance gradient, we observed no clear linear trend,

and pairwise comparisons remained broadly distributed across both the geographic and phenotypic distance ranges. These findings are consistent with the spatial distribution patterns observed in the Loreto Region, where multiple phenotypic groups coexist across different geographic locations.

Statistical analysis of quantitative phenotypic descriptors in *T. cacao* accessions revealed distinct patterns of variation among

phenotypic groups, necessitating both parametric and non-parametric analytical approaches based on trait distributions (Supplementary Table S7). Normality tests identified three traits that were suitable for parametric analysis: fruit diameter, pericarp thickness, and seed thickness. One-way ANOVA of these normally distributed traits revealed highly significant differences between phenotypic groups ( $p < 0.0001$ ), with fruit diameter showing the highest level of differentiation ( $F = 159.53$ ), followed by pericarp thickness ( $F = 65.37$ ) and seed thickness ( $F = 19.44$ ).

Kruskal-Wallis analysis of non-normally distributed traits demonstrated varying levels of differentiation among characteristics (Supplementary Table S7). Reproductive traits exhibited significant differences ( $p < 0.0001$ ), fruit weight ( $H = 292.80$ ), pericarp weight ( $H = 284.11$ ), and fruit length ( $H = 164.35$ ). Seed characteristics also displayed significant variation, as evidenced by the seed width ( $H = 141.42$ ), seed length ( $H = 139.54$ ), and pod index ( $H = 178.01$ ). Floral traits showed moderate levels of differentiation, with ovary length ( $H = 133.20$ ), style length ( $H = 121.19$ ), and staminode length ( $H = 102.66$ ) differing significantly among the groups. In contrast, there were no significant differences in vegetative characteristics among the phenotypic groups, including leaf length ( $H = 5.89$ ,  $p = 0.2077$ ), leaf width ( $H = 5.17$ ,  $p = 0.2703$ ), petiole length ( $H = 8.02$ ,  $p = 0.0908$ ), and filament length ( $H = 4.39$ ,  $p = 0.3564$ ).

The analysis of qualitative phenotypic descriptors revealed varying levels of differentiation among phenotypic groups (Supplementary Table S8). Eight of the 11 descriptors showed statistically significant differences. Among floral characteristics, filament ( $\chi^2 = 10.96$ ,  $p = 0.0270$ ) and sepal ( $\chi^2 = 24.10$ ,  $p = 0.0001$ ) anthocyanin pigmentation differed significantly across groups, whereas ovary pigmentation did not ( $\chi^2 = 5.64$ ,  $p = 0.2276$ ).

Fruit characteristics displayed significant variation among the phenotypic groups in five out of the six traits (Supplementary Table S8). Both ripe ( $\chi^2 = 25.64$ ,  $p = 0.0012$ ) and unripe fruit colors ( $\chi^2 = 35.44$ ,  $p < 0.0001$ ) showed highly significant differences, along with fruit shape ( $\chi^2 = 33.28$ ,  $p = 0.0068$ ), fruit apex form ( $\chi^2 = 28.30$ ,  $p = 0.0050$ ), and fruit rugosity ( $\chi^2 = 17.13$ ,  $p = 0.0287$ ). There was no significant difference in fruit basal constriction ( $\chi^2 = 13.45$ ,  $p = 0.0973$ ). For seed characteristics, the transversal shape differed significantly among groups ( $\chi^2 = 22.76$ ,  $p = 0.0037$ ), while the longitudinal shape did not ( $\chi^2 = 13.89$ ,  $p = 0.3079$ ).

Pairwise comparisons between the phenotypic groups revealed specific patterns of trait differentiation (Supplementary Table S8). Phenotypic group 1 differed from phenotypic group 3 in fruit shape ( $p = 0.0033$ ), from phenotypic group 4 in sepal anthocyanin ( $p = 0.0001$ ) and fruit apex form ( $p = 0.0001$ ), and from phenotypic group 5 in filament anthocyanin ( $p = 0.0305$ ), ripe fruit color ( $p = 0.0274$ ), and unripe fruit color ( $p = 0.0002$ ). Phenotypic groups 3 and 4 showed differences in most traits including anthocyanin pigmentation, fruit characteristics, and seed morphology. Phenotypic groups 4 and 5 differed in filament anthocyanin content ( $p = 0.0070$ ), sepal anthocyanin content ( $p = 0.0010$ ), unripe fruit color ( $p = 0.0001$ ), fruit apex form ( $p = 0.0005$ ), and seed transversal shape ( $p = 0.0079$ ).

### 3.2.4 Phenotypic diversity analysis using Shannon-Weaver Index

The Shannon-Weaver Diversity Index ( $H'$ ) analysis revealed distinct patterns of variation across qualitative descriptors in the *T. cacao* germplasm bank (Figure 5A). Among all traits, seed longitudinal shape exhibited the highest diversity ( $H' \approx 1.1$ , range: 1.0-1.2), followed by fruit apex form ( $H' \approx 1.1$ , range: 0.85-1.15) and seed transversal shape ( $H' \approx 0.65$ , range: 0.4-0.8). Fruit characteristics showed intermediate diversity levels, with fruit shape ( $H' \approx 0.8$ ), fruit rugosity ( $H' \approx 0.7$ ), and fruit basal constriction ( $H' \approx 0.7$ ) displaying similar values. Fruit color traits showed a marked difference between stages, with unripe fruit color ( $H' \approx 0.4$ ) exhibiting higher diversity than ripe fruit color ( $H' \approx 0.2$ ).

Floral anthocyanin pigmentation traits consistently showed the lowest diversity indices among all characteristics examined. Specifically, anthocyanin in filaments ( $H' \approx 0.2$ ), anthocyanin in ovaries ( $H' \approx 0.1$ ), and anthocyanin in sepals ( $H' \approx 0.2$ ) displayed notably low diversity values, with minimal variation in their ranges.

Analysis of the diversity distribution across phenotypic groups revealed distinct patterns (Figure 5B). The groups showed varying levels of median diversity, with phenotypic group 1 exhibiting the highest median value ( $H' \approx 0.6$ ), followed by phenotypic group 5 (orange,  $H' \approx 0.62$ ), phenotypic group 4 (purple,  $H' \approx 0.58$ ), phenotypic group 2 (blue,  $H' \approx 0.52$ ), and phenotypic group 3 (green,  $H' \approx 0.45$ ), which showed the lowest median diversity. The complete collection ("All plants," gray) displayed the highest overall diversity ( $H' \approx 0.7$ ) and the broadest range of values (0.4-0.95).

Analysis of phenotypic relationships and diversity patterns in the *T. cacao* germplasm bank revealed complex associations between traits and distinct diversity profiles across phenotypic groups (Supplementary Figure S9A). Correlation analysis of the qualitative traits revealed several strong relationships. Notably, seed transversal shape showed strong positive correlations with anthocyanin pigmentation in both filaments ( $r = 0.88$ ) and ovaries ( $r = 0.79$ ), suggesting coordinated expression of these traits. Fruit characteristics exhibited interesting patterns, with fruit shape showing a strong negative correlation with fruit rugosity ( $r = -0.94$ ) and a strong positive correlation with ripe fruit color ( $r = 0.95$ ). Fruit basal constriction demonstrated moderate to strong correlations with several traits, including a negative correlation with anthocyanin content in the ovary ( $r = -0.84$ ) and a positive correlation with unripe fruit color ( $r = 0.88$ ). These correlations suggest potential developmental or genetic links between these morphological traits.

The distribution of phenotypic diversity across the phenotypic groups (Supplementary Figure S9B) showed varying patterns of trait expression. The seed longitudinal shape consistently exhibited high diversity across all phenotypic groups ( $H' > 1.0$ ), with the highest value in phenotypic group 3 ( $H' = 1.342$ ). Fruit characteristics showed moderate to high diversity, particularly fruit apex form, which maintained high diversity values across all phenotypic groups ( $H'$  ranging from 0.843 to 1.337). In contrast, anthocyanin pigmentation traits generally showed low diversity values, especially in ovary pigmentation ( $H' < 0.15$  across most groups).

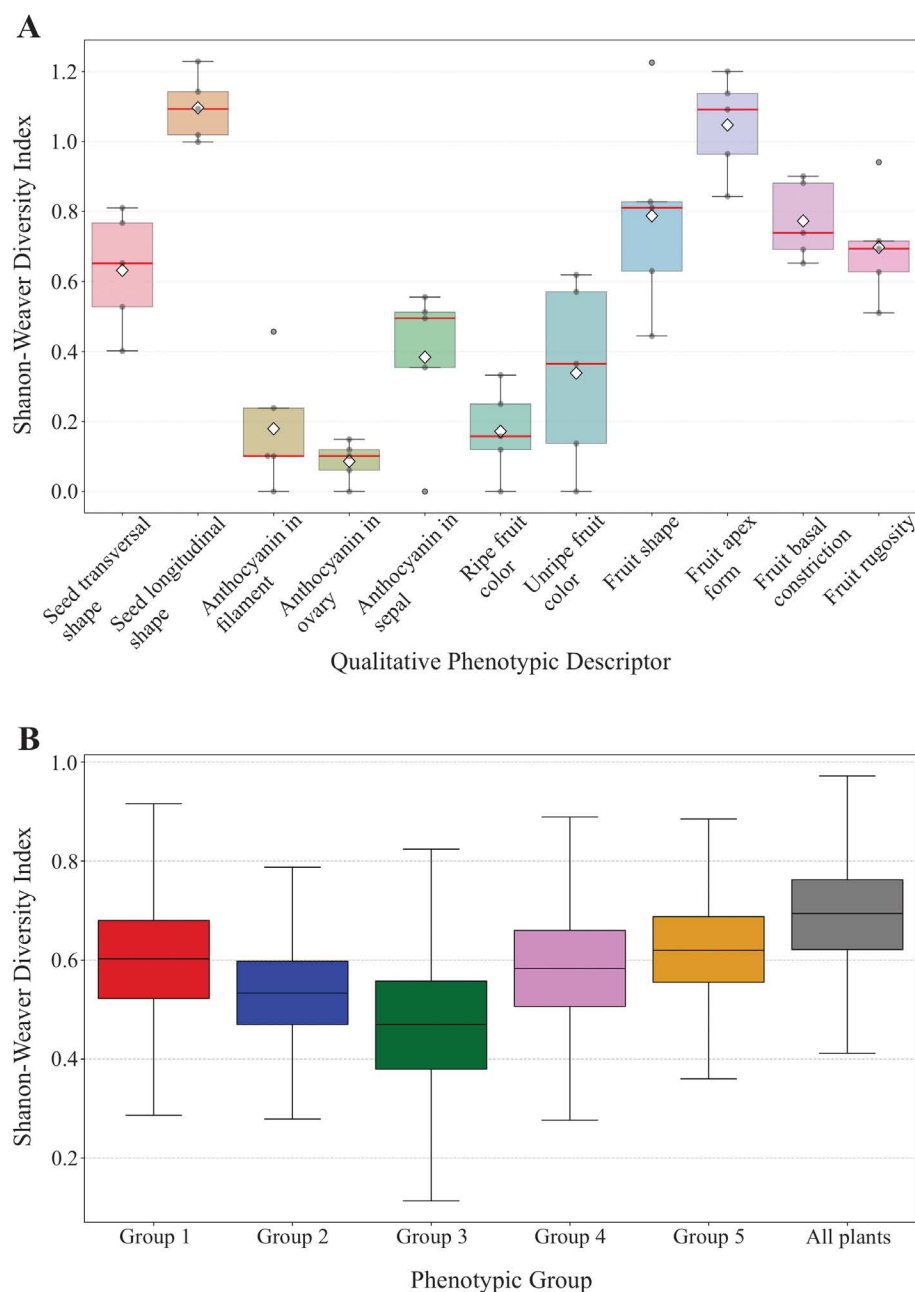


FIGURE 5

Shannon-Weaver Diversity Index analysis of *T. cacao* accessions from the Loreto Region germplasm bank. **(A)** Distribution of Shannon-Weaver Diversity Index values across qualitative phenotypic descriptors, showing the variation in diversity levels among different morphological traits. Box plots indicate the median (horizontal line), mean (diamond), quartiles (box), and outliers (dots). **(B)** Comparison of Shannon-Weaver Diversity Index values among the five phenotypic groups and the overall collection (All plants), illustrating the relative phenotypic diversity within each group.

Analysis of group-specific patterns revealed that phenotypic group 1 showed the highest global diversity index ( $H' = 0.602$ ) among the individual groups, while phenotypic group 3 showed the lowest ( $H' = 0.470$ ). The overall collection ("All plants") demonstrated the highest global diversity ( $H' = 0.692$ ), indicating that phenotypic grouping effectively captured complementary aspects of morphological variation. Notably, phenotypic group 5 showed particularly high diversity in fruit shape ( $H' = 1.225$ ), while phenotypic group 3 exhibited distinctive patterns in fruit rugosity ( $H' = 0.941$ ) compared to the other groups.

## 4 Discussion

### 4.1 Germplasm bank development

Establishing a native cacao germplasm bank in the Loreto region of the Peruvian Amazon represents a crucial initiative for global cacao conservation. This region is one of the primary centers of *T. cacao* origin and harbors exceptional genetic variability that is essential for breeding programs targeting disease resistance, yield improvement, and climate adaptation (Motamayor et al., 2008;

Thomas et al., 2012). The genetic diversity preserved in this bank holds dual importance: ensuring sustainable cacao production while safeguarding unique local varieties cultivated by indigenous communities for generations. These genetic resources are particularly valuable, given their potential adaptations to local environmental conditions, making them crucial for breeding programs focused on sustainability and resilience (Thomas et al., 2012; Zhang and Motilal, 2016a).

The successful establishment of the germplasm bank underscores the importance of integrating optimized grafting techniques and robust rootstock selection to conserve Amazonian cacao diversity. Achieving a 100% survival rate marks a significant advancement in the *ex situ* conservation of tropical tree species. This exceptional outcome aligns with the successful methodologies documented at the International Cocoa Genebank in Trinidad (Bekele and Bekele, 1996; Bekele et al., 2006) and can be attributed to two key factors: strategic selection of rootstock material and implementation of standardized grafting techniques. Such high establishment success is particularly noteworthy given the challenges typically associated with tropical tree species conservation, and provides a model for future germplasm bank initiatives.

The applied grafting methodology demonstrated a strong alignment with successful practices across major cacao-producing regions. Recent advances in grafting techniques have shown remarkable success rates, with *in vitro* micrografting achieving 95% success in apical grafts and 80% success in side grafts using axillary buds (Miguelé-Sierra et al., 2017). Critical factors influencing success include optimal timing, with the best results observed in three-month-old rootstock (N'zi et al., 2023), and careful clone selection, exemplified by the 82.6% survival rate of the Trinitario ICS40 clone (Tchatchoua et al., 2023). The integration of natural plant growth regulators further enhances grafting outcomes, particularly in promoting vegetative development (Sari and Utami, 2024).

The selection of IMC 67 as rootstock material proved instrumental in our success, although its use presents both opportunities and challenges for long-term germplasm conservation. Its key advantages include vigorous root systems, enhanced yield potential, and increased disease resistance (Martirosyan et al., 2023). Previous research has demonstrated that the dominant rootstock effects of IMC 67 can enhance nutrient use efficiency, while maintaining scion-determined bean quality characteristics (Schmidt et al., 2021). However, important limitations include its narrow genetic base, which may restrict adaptation across diverse edaphoclimatic conditions (Fernández-Paz et al., 2021; Galvis et al., 2023) and challenges in achieving consistent survival rates across different substrates (Schmidt et al., 2021). These limitations underscore the need for a comprehensive genetic and physiological characterization of IMC 67, particularly regarding rootstock-mediated variance in nutrient uptake and stress responses (Motilal et al., 2011; Montenegro et al., 2023; Ortiz-Álvarez et al., 2023).

## 4.2 Phenotypic characterization

Phenotypic characterization highlights the role of the Loreto Region as a genetic reservoir, with reproductive traits reflecting evolutionary pressure and human-driven selection. The extensive variation in fruit characteristics (e.g., weight and pericarp thickness) and pod index (Table 1) mirrors the diversity reported in the International Cocoa Genebank, Trinidad (Bekele et al., 2006, 2021), positioning it as a critical hotspot for cacao genetic resources. This variability likely reflects both natural evolutionary processes, such as adaptation to Amazonian microhabitats, and historical human-mediated selection during early domestication (Motamayor et al., 2008; Cornejo et al., 2018).

The pronounced differentiation in reproductive traits compared to vegetative traits (Supplementary Table S7) provides compelling evidence for differential selection pressures during cacao domestication. This pattern aligns with established evolutionary models of crop domestication, wherein traits directly linked to yield components and product quality, particularly fruit size, pericarp characteristics, and seed number, experience substantially stronger selective pressure than vegetative features (Motamayor et al., 2002; Cornejo et al., 2018; McElroy et al., 2018; Zarrillo et al., 2018). The quantitative analysis reveals remarkably high coefficients of variation in reproductive traits, with fruit weight and pericarp weight exhibiting CV values exceeding 48% (Table 1). These elevated variation levels strongly suggest targeted selection, a phenomenon similarly documented in other Neotropical perennial crops undergoing domestication processes (Clement et al., 2010; Iriarte et al., 2020; Osorio-Guarín et al., 2020; Lanaud et al., 2024). Conversely, comparatively constrained variation in vegetative characteristics, such as leaf dimensions and petiole length, likely reflects stabilizing selection for optimal structural architecture that ensures photosynthetic efficiency while maintaining mechanical resilience in heterogeneous Amazonian environments (Poorter et al., 2012). In the Loreto region, indigenous communities have practiced distinct selection criteria, as documented by Vásquez-Ocmín et al. (2018) who found that riverine communities along the Ucayali and Marañón rivers have traditionally selected for large fruits with high pulp content for fermented beverage production, while upland communities prioritized disease resistance and bean quality. Archaeological evidence from pre-Columbian sites in the region further suggests selection for specific fruit forms dating back to at least 1500 years ago (Motamayor et al., 2002; Zarrillo et al., 2018), contributing to the morphological diversity observed in our germplasm bank.

Ethnobotanical research in the Loreto region has documented sophisticated indigenous selection practices that have further shaped this morphological diversity. It has been demonstrated that riverine communities along the Peruvian Amazon watersheds have historically prioritized large-fruited phenotypes with abundant pulp for traditional fermented beverage production, creating distinct selection trajectories from upland communities that emphasized disease resistance and bean quality attributes (Motamayor et al., 2002;

Clement et al., 2010; Vásquez-Ocmín et al., 2018; Zarrillo et al., 2018). This differentiated selection across ecological niches has contributed to the morphological compartmentalization observed in our germplasm bank. Archaeological investigations at pre-Columbian sites throughout the region provide further temporal context, with evidence for selection of specific fruit morphotypes dating back approximately 1,500 years (Zarrillo et al., 2018; Lanaud et al., 2024). These findings collectively suggest that the exceptional phenotypic diversity documented in the present study reflects the cumulative influence of centuries of both natural selection and targeted human intervention across diverse Amazonian microenvironments.

The multivariate analysis of phenotypic diversity revealed complex spatial patterns that challenge traditional isolation-by-distance models of genetic structuring in Amazonian cacao populations. UMAP analysis successfully resolved five distinct phenotypic groups (Figure 4A), providing evidence of significant morphological differentiation within the germplasm collection despite geographical proximity of some accessions. While the spatial distribution of these phenotypic groups along the region's river systems (Figure 4B) initially suggests riparian-mediated dispersal patterns, statistical analysis reveals a more nuanced reality. The remarkably weak correlation between phenotypic and geographic distances (Mantel  $r = 0.066$ ,  $p = 0.0017$ ; Supplementary Figure S8) indicates that geographical proximity explains less than 1% of the observed phenotypic variation. This finding points to a complex interplay of dispersal mechanisms operating across the Amazonian landscape. Several factors likely contribute to this pattern: limited-range pollination by Ceratopogonidae midges (50–200 meters) creating localized gene flow (Groeneveld et al., 2010), seasonal flood-mediated seed dispersal through hydrochory (Thomas et al., 2012), and the interconnected tributary network facilitating both natural and human-mediated germplasm exchange between geographically distant but hydrologically connected communities (Solorzano et al., 2012). These river systems have historically served as conduits for human migration and crop dissemination throughout the Amazon (Thomas et al., 2012; Zhang and Motilal, 2016b; Colli-Silva et al., 2024), creating opportunities for admixture between wild and cultivated populations that drive phenotypic novelty and adaptive diversity (Bidot Martínez et al., 2015a; Nieves-Orduña et al., 2021; Fouet et al., 2022). This complex evolutionary history has produced cacao populations with remarkable genetic variability and environmental resilience—characteristics essential for sustainable cultivation across diverse environments (Zhang and Motilal, 2016b; Friedman, 2020; Apshara and Sane, 2025).

The remarkable diversity in seed and fruit morphology observed in this germplasm collection offers significant practical advantages for cacao breeding programs targeting quality improvement and disease resistance. Shannon-Weaver diversity analysis revealed exceptionally high indices for seed longitudinal shape ( $H' \approx 1.1$ ) and fruit apex form ( $H' \approx 1.1$ ; Figure 5A), traits with direct implications for post-harvest processing and disease management. Seed shape significantly influences fermentation dynamics, with irregular and oblong morphotypes demonstrating more uniform fermentation patterns compared to ovate or elliptical

seeds (Kongor et al., 2016). This natural variation provides breeders with valuable genetic resources for selecting genotypes with optimal fermentation characteristics that enhance flavor development. Similarly, fruit apex morphology plays a critical role in disease susceptibility through its influence on moisture management. Pods with attenuated apex forms exhibit reduced water accumulation and retention compared to those with obtuse or rounded shapes, directly affecting susceptibility to black pod disease caused by *Phytophthora* species (Daymond and Hadley, 2008). Research by (Hoopen et al., 2012) demonstrates that pod morphology directly influences surface wetness duration, with prolonged moisture conditions facilitating pathogen propagule germination and subsequent infection. This relationship between morphological characteristics and disease resistance mechanisms presents valuable selection targets for breeding programs aimed at developing resilient varieties without compromising yield components (Marita et al., 2001; Bekele and Phillips-Mora, 2019; Nieves-Orduña et al., 2024). These findings highlight how the phenotypic diversity documented in our germplasm collection provides tangible opportunities for addressing key production challenges through targeted breeding interventions.

Notably, the lower diversity in floral anthocyanin pigmentation (Figure 5A) likely reflects more complex evolutionary dynamics than simple domestication. This pattern could be explained by multiple factors: the conservation of anthocyanin pathway genes, even in unpigmented lineage species (Ho and Smith, 2016), the role of gene expression rather than structural mutations in pigmentation loss (Rauscher, 2008), or the interaction between floral traits and pollinator selection pressures (Trunschke et al., 2021). The conservation of anthocyanin-related traits suggests evolutionary flexibility, which could be valuable for future adaptation and breeding efforts.

Strong correlations between reproductive traits, such as fruit weight, pericarp weight, and seed number (Supplementary Table S5), suggest pleiotropic genetic control or developmental integration, a common phenomenon in crop plants (Conner et al., 2011). These linkages highlight the potential for simultaneous improvement of multiple yield-related traits through targeted breeding. Conversely, the negative correlation between the pod index and fruit weight (Supplementary Table S5) underscores the trade-offs that must be managed in breeding programs (Bekele and Phillips-Mora, 2019).

The exceptional phenotypic diversity documented in this germplasm collection establishes a strategic foundation for accelerating cacao improvement through integration with advanced biotechnological approaches. Recent breakthroughs in CRISPR-Cas9 gene editing technology demonstrate particular promise for leveraging this diversity, as evidenced by successful applications in differentiating fine and bulk cocoa varieties (Scharf et al., 2020; La-Rostami et al., 2022), enhancing disease resistance (Fister et al., 2018), and modifying key agronomic traits in tree species using genome editing (Pak et al., 2022; Pal and Pal, 2024). When combined with multi-omics platforms, these technologies can significantly accelerate trait introgression and genetic improvement (Karumamkandathil et al., 2022; Chaturvedi et al.,

2024; Kulesza et al., 2024). To maximize the utility of this germplasm collection, future research should prioritize genome-wide association studies (GWAS) that establish robust connections between phenotypic variation and underlying molecular markers, an approach that has already yielded significant advancements in flavor profile enhancement and yield improvement in cacao (Osorio-Guarín et al., 2020; Bekele et al., 2022; Colonges et al., 2022). While implementing these genomic approaches presents logistical challenges in the Amazonian context, including limited access to high-throughput phenotyping infrastructure and bioinformatic resources, alternative methodologies such as targeted candidate gene sequencing and reduced representation approaches (RAD-seq or GBS) offer cost-effective solutions for initial genetic characterization (Ricaño-Rodríguez et al., 2019; Ramirez-Ramirez et al., 2024). Establishing strategic collaborative networks between regional institutions and international research centers will be essential for overcoming these technical and resource limitations while ensuring research benefits are directed toward local breeding programs and smallholder farmers. The integration of these germplasm resources with emerging technologies in high-throughput phenotyping and genomic prediction models represents a powerful pathway toward developing climate-resilient, high-yielding cacao varieties with enhanced quality characteristics (Niazian and Niedbala, 2020; Dijk et al., 2021; Yan and Wang, 2023; Eftekhari et al., 2024).

## 5 Conclusions

This study presents the successful development and comprehensive characterization of a significant cacao germplasm bank from the Loreto region of the Peruvian Amazon. The exceptional establishment success and subsequent 100% long-term survival observed across multiple seasons, including during the severe 2019 and 2024 drought periods, demonstrate the resilience of the selected rootstock-scion combinations. This robustness under environmental stress highlights the adapted nature of these native germplasm resources and their potential value for climate resilience breeding programs. The exceptional variation observed in economically important traits, particularly pod characteristics and yield components, positions this germplasm bank as a valuable resource for breeding programs that address contemporary challenges in cacao production.

The identification of five distinct phenotypic groups, coupled with high Shannon-Weaver diversity indices, indicates the preservation of significant genetic diversity from the center of origin of cacao. The stronger differentiation in reproductive traits compared with vegetative characteristics provides insights into the evolutionary history of the species and suggests promising avenues for trait improvement.

This germplasm bank offers unprecedented opportunities for multiomics exploration of native cacao diversity. Future studies should integrate genomic, epigenomic, transcriptomic, proteomic, and metabolomic analyses to develop comprehensive molecular

profiles for these unique genotypes. Such multiomics approaches will reveal the molecular basis of important traits and identify the regulatory networks that control their expression. The generation of high-quality reference genomes from diverse accessions will facilitate comparative genomic analyses and the identification of structural variations associated with adaptive traits.

The extensive phenotypic diversity of the novel germplasm bank, combined with emerging biotechnologies, presents novel opportunities for accelerated cacao improvement. The application of CRISPR-Cas9 and base editing technologies can enable precise genetic modifications for trait enhancement, whereas emerging techniques such as prime editing and epigenome editing offer additional approaches for germplasm improvement. These modern breeding tools, supported by multi-omics data, will facilitate the development of climate-resilient high-yielding varieties with enhanced disease resistance and quality traits.

The comprehensive phenotypic characterization presented herein provides a robust foundation for these advanced studies. Continuous evaluation using both traditional and emerging technologies will maximize the contribution of this valuable genetic resource to sustainable cacao production. The integration of phenotypic data with multi-omics profiles and modern breeding approaches has positioned this germplasm bank as a crucial resource for securing the future of global cacao production in the face of mounting environmental challenges.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#). Further inquiries can be directed to the corresponding authors.

## Ethics statement

Written informed consent was obtained from the individual(s)/minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

## Author contributions

SI: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. AS: Investigation, Methodology, Software, Writing – original draft. JR: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. MC: Conceptualization, Investigation, Validation, Writing – review & editing. CP: Data curation, Investigation, Methodology, Writing – review & editing. JC: Formal analysis, Investigation, Supervision, Validation, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by a grant (no. 013-2015-INIA-PNIA-DE) from the National Program for Agricultural Innovation (PNIA) through the Instituto Nacional de Innovación Agraria (INIA). Publication of this article was funded by the Universidad Nacional de la Amazonía Peruana (UNAP) under Rectoral Resolution No. 0631-2025-UNAP.

## Acknowledgments

We gratefully acknowledge Lolo Romel Lumba Vásquez for his contributions to the sample collection and grafting activities. We also thank Oberluis Panduro Murayari for the sample collection, while Angel Ricardo Vizcarra Vela and Robinson Murayari Tamani provided additional support with grafting. We dedicate this work to the memory of the Eng. Agustín Gonzales Coral of the Peruvian Amazon Research Institute (IIAP), who served as a valued member of the project's technical-scientific team.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. The authors declare that Generative AI (Claude) has been used only for grammar and spelling checks in this manuscript. No generative AI was used for content generation or scientific interpretation.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcsc.2025.1576239/full#supplementary-material>

### SUPPLEMENTARY FIGURE 1

Location and aerial view of the *Theobroma cacao* L. germplasm bank at the Experimental Field "El Dorado" in Loreto, Peru. (A) Geographic location of the study site within Peru, showing its position along the Iquitos-Nauta highway; (B) Satellite image showing the access road to the germplasm bank within the Instituto Nacional de Innovación Agraria (INIA) facilities; (C) Detailed aerial view of the 10,000 m<sup>2</sup> germplasm bank area (red polygon) at coordinates 03° 57'01" S, 73°24'59" W; (D) Ground-level photographs showing the entrance signage and established cacao plants within the germplasm bank.

### SUPPLEMENTARY FIGURE 2

Establishment of temporary shade species for the *T. cacao* L. germplasm bank at the Experimental Field "El Dorado", Loreto, Peru. The images show the systematic planting arrangement of banana (*Musa x paradisiaca* L.) at 3 × 3 m spacing and ice-cream bean (*Inga edulis* Mart.) at 12 × 12 m intervals during January–August 2016. (A–C) Different views showing the early establishment phase of the shading species. (D, E) Development of shade trees after planting, with identification markers visible, demonstrating successful establishment of the temporary shade system.

### SUPPLEMENTARY FIGURE 3

The sequential process of rootstock development using IMC 67 (Iquitos Mixed Calabacillo) cacao clone at the Experimental Field "El Dorado", Loreto, Peru. (A) Mature cocoa fruits of IMC 67 selected for seed extraction; (B) Initial seed germination setup using banana leaves as germination beds; (C) Successfully germinated seeds showing radicle emergence after 48 hours; (D) Technical staff preparing growing substrate (2:1:1 ratio of agricultural soil, wood organic matter, and poultry manure) and filling polyethylene bags (15 × 30 cm); (E) Five-month-old rootstock seedlings under 80% shade conditions (Raschel® mesh), showing uniform development with standardized morphological characteristics (stem diameter: 8.77 ± 0.88 mm; height: 78.74 ± 4.22 cm; leaf number: 25.28 ± 1.97).

### SUPPLEMENTARY FIGURE 4

Plant material collection process of native *T. cacao* L. in the Loreto region, Peru (February–August 2017). (A) Field collection team during expedition; (B) Technician using climbing techniques to access the middle third of the cacao tree canopy; (C) Selection and harvesting of healthy scions using sterilized pruning equipment; (D) Sequential preparation of collected scions showing selection of healthy material, removal of excess foliage, and wrapping in moistened newspaper; (E) Transportation system using portable coolers to maintain scion viability during field expeditions, with proper labeling.

### SUPPLEMENTARY FIGURE 5

Grafting techniques and establishment of the *T. cacao* L. germplasm bank at the Experimental Field "El Dorado", Loreto, Peru. (A, B) Demonstration of grafting methods: side grafting showing the 3–4 cm downward-slanting cut and flap creation (left), and top-cleft grafting showing the vertical slit and wedge insertion (right); (C–F) Successfully grafted plants showing early development stages using both grafting techniques; (G) Final establishment of the germplasm bank with 140 accessions (700 plants total) arranged in a systematic 3 × 3 m grid pattern within a 10,000 m<sup>2</sup> area.

### SUPPLEMENTARY FIGURE 6

Frequency distributions of quantitative phenotypic descriptors in *T. cacao* L. accessions from the Loreto Region germplasm bank. Histograms (green bars) and fitted normal distribution curves (red lines) are shown for 26 morphological traits including vegetative (leaf, petiole, and pedicel measurements), floral (sepal, petal, filament, staminode, style, and ovary dimensions), fruit (weight, length, diameter, pericarp characteristics, and furrow depth), and seed characteristics (number, fresh weight, dimensions, and pod index). For each trait, the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) are provided. The x-axis represents the measured values in their respective units, and the y-axis shows the frequency of observations. The overlaid normal distribution curves illustrate the degree of deviation from normality for each trait.

### SUPPLEMENTARY FIGURE 7

Cluster analysis validation for phenotypic grouping of *T. cacao* accessions from the Loreto Region germplasm bank. (A) Elbow plot showing the relationship between the number of clusters ( $k$ ) and Within-Cluster Sum of Squares (WCSS). The red point indicates the optimal number of clusters ( $k = 5$ ) determined by the Elbow method, where additional clusters do not substantially reduce WCSS; (B) Silhouette plot displaying the clustering

quality for the five identified phenotypic groups. Each horizontal bar represents an individual accession, with colors corresponding to different clusters. The width of the silhouette indicates the degree of sample membership within its assigned cluster. The vertical red dashed line represents the average silhouette score (0.319) across all clusters, indicating reasonable separation between clusters. Higher silhouette scores (closer to 1) indicate better cluster assignment, while scores closer to 0 indicate potential overlap between clusters.

#### SUPPLEMENTARY FIGURE 8

Correlation between phenotypic (Gower) and geographic distances among native cacao accessions of the germplasm bank from the Loreto Region, Peru. The hexagonal heatmap shows the density of pairwise comparisons, with colors indicating the frequency of observations (Count). The weak but significant positive correlation (Mantel  $r = 0.0661$ ,  $p = 0.0017$ ) suggests minimal spatial structure in phenotypic variation across the study area.

#### SUPPLEMENTARY FIGURE 9

Correlation analysis and phenotypic diversity patterns in *T. cacao* accessions. (A) Correlation matrix showing the relationships between qualitative phenotypic descriptors. Red colors indicate positive correlations, blue colors indicate negative correlations and color intensity represents correlation strength; (B) Heat map of Shannon-Weaver Diversity Index values across phenotypic groups and traits, including global diversity index. The color scale indicates the diversity level from low (blue) to high (red).

#### SUPPLEMENTARY VIDEO 1

Interactive three-dimensional UMAP visualization of phenotypic diversity in native cacao accessions of the germplasm bank from the Loreto Region. The interactive plot shows clustering patterns of 402 accessions into five phenotypic groups based on 36 phenotypic descriptors. UMAP dimensions explain 33.3% (UMAP1), 37.1% (UMAP2), and 29.6% (UMAP3) of total variance. Each point represents an individual accession, color-coded by phenotypic group (Group 1: red, Group 2: blue, Group 3: green, Group 4: purple, Group 5: orange). Ellipsoids represent 95% confidence intervals around cluster centroids, characterizing representative phenotypic traits for each group. HTML format enables interactive rotation and zoom for detailed exploration of cluster relationships.

#### SUPPLEMENTARY TABLE 1

A comprehensive dataset of passport information and phenotypic characterization of 134 native *T. cacao* L. accessions conserved in the Germplasm Bank from the Loreto Region.

#### SUPPLEMENTARY TABLE 2

Morphological descriptors used for the phenotypic characterization of native *T. cacao* L. accessions in the Germplasm Bank from the Loreto Region: quantitative and qualitative phenotypic descriptors for leaf, flower, fruit, and seed organs, with their respective measurement units and categorical states.

#### SUPPLEMENTARY TABLE 3

Normality tests and statistical analysis approach for quantitative phenotypic descriptors in *T. cacao* accessions from the Loreto Region germplasm bank, showing Shapiro-Wilk, Kolmogorov-Smirnov, and Anderson-Darling test results for the overall collection and individual phenotypic groups.

#### SUPPLEMENTARY TABLE 4

Descriptive statistics of quantitative phenotypic descriptors in *T. cacao* accessions from the Loreto Region germplasm bank, showing overall collection and phenotypic group-specific measurements.

#### SUPPLEMENTARY TABLE 5

Spearman's rank correlation coefficients ( $\rho$ ) among quantitative phenotypic descriptors in *T. cacao* accessions from the Loreto Region germplasm bank.

#### SUPPLEMENTARY TABLE 6

Multivariate analysis of phenotypic descriptors in native cacao accessions from the Loreto Region. The Excel file includes a Gower distance matrix showing phenotypic similarity among accessions, One-way ANOVA test results for phenotypic descriptors across groups, correlation coefficients between phenotypic descriptors and UMAP dimensions, and the percentage contribution of each descriptor to UMAP dimensions.

#### SUPPLEMENTARY TABLE 7

Statistical analysis of quantitative phenotypic variation in *T. cacao* L. accessions: ANOVA with Tukey's HSD for normally distributed descriptors and Kruskal-Wallis with Dunn's test for non-normally distributed descriptors from the Loreto Region germplasm bank.

#### SUPPLEMENTARY TABLE 8

Distribution and statistical analysis of qualitative phenotypic descriptors in all samples and across the phenotypic groups in *T. cacao* accessions from the Loreto Region germplasm bank, showing frequency distribution results (%), Chi-square test results, and Pairwise comparisons between phenotypic groups.

## References

- Aikpokpodion, P. O. (2010). Variation in agro-morphological characteristics of cacao, *Theobroma cacao* L., in farmers' fields in Nigeria. *N. Z. J. Crop Hortic. Sci.* 38, 157–170. doi: 10.1080/0028825X.2010.488786
- Aikpokpodion, P. O., Motamayor, J. C., Adetimirin, V. O., Adu-Ampomah, Y., Ingelbrecht, I., Eskes, A. B., et al. (2009). Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. *Tree Genet. Genomes* 5, 699–711. doi: 10.1007/s11295-009-0221-1
- Aphara, S. E., and Sane, A. (2025). "Genetic diversity of cocoa (*Theobroma cacao* L.) and sustainable utilization," in *Genetic Diversity of Fruits and Nuts*, vol. 18. Ed. H. N. Murthy (CRC Press, Boca Raton).
- Bekele, F., and Bekele, I. (1996). A sampling of the phenetic diversity of cacao in the international cocoa gene bank of Trinidad. *Crop Sci.* 36, 57–64. doi: 10.2135/cropsci1996.0011183X003600010010x
- Bekele, F. L., Bekele, I., Butler, D. R., and Bidaisee, G. G. (2006). Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm from the International Cocoa Genebank, Trinidad. *Genet. Resour. Crop Evol.* 53, 933–948. doi: 10.1007/10722-004-6692-x
- Bekele, F. L., Bidaisee, G. G., Allegre, M., Argout, X., Fouet, O., Boccara, M., et al. (2022). Genome-wide association studies and genomic selection assays made in a large sample of cacao (*Theobroma cacao* L.) germplasm reveal significant marker-trait associations and good predictive value for improving yield potential. *PLoS One* 17, e0260907. doi: 10.1371/journal.pone.0260907
- Bekele, F., Bidaisee, G., and Saravanakumar, D. (2021). Examining phenotypic diversity and economic value of cacao (*Theobroma cacao* L.) conserved at the International Cocoa Genebank, Trinidad to support improvement in cocoa yield globally. *Trop. Agric.* 97, 82–93. doi: 10.1007/978-3-030-23265-8\_12
- Bekele, F. L., Bidaisee, G. G., Singh, H., and Saravanakumar, D. (2019). Morphological characterisation and evaluation of cacao (*Theobroma cacao* L.) in Trinidad to facilitate utilisation of Trinitario cacao globally. *Genet. Resour. Crop Evol.* 67, 621–643. doi: 10.1007/s10722-019-00793-7
- Bekele, F., and Phillips-Mora, W. (2019). "Cacao (*Theobroma cacao* L.) breeding," in *Advances in Plant Breeding Strategies: Industrial and Food Crops*. Eds. J. M. Al-Khayri, S. M. Jain and D. V. Johnson (Springer International Publishing, Cham), 409–487. doi: 10.1007/978-3-030-23265-8\_12
- Bidot Martínez, I., Riera Nelson, M., Flamand, M.-C., and Bertin, P. (2015a). Genetic diversity and population structure of anciently introduced Cuban cacao *Theobroma cacao* plants. *Genet. Resour. Crop Evol.* 62, 67–84. doi: 10.1007/s10722-014-0136-z
- Bidot Martínez, I., Valdés de la Cruz, M., Riera Nelson, M., and Bertin, P. (2015b). Morphological characterization of traditional cacao (*Theobroma cacao* L.) plants in Cuba. *Genet. Resour. Crop Evol.* 64, 73–99. doi: 10.1007/s10722-015-0333-4
- Ceccarelli, V., Fremout, T., Chavez, E., Argüello, D., Loo Solórzano, R. G., Sotomayor Cantos, I. A., et al. (2024). Vulnerability to climate change of cultivated and wild cacao in Ecuador. *Clim. Change* 177, 1–22. doi: 10.1007/s10584-024-03756-9
- Ceccarelli, V., Lastra, S., Loo Solórzano, R. G., Chacón, W. W., Nolasco, M., Sotomayor Cantos, I. A., et al. (2022). Conservation and use of genetic resources of cacao (*Theobroma cacao* L.) by gene banks and nurseries in six Latin American countries. *Genet. Resour. Crop Evol.* 69, 1283–1302. doi: 10.1007/s10722-021-01304-3
- Chaturvedi, P., Pierides, I., Zhang, S., Schwarzerova, J., Ghatak, A., and Weckwerth, W. (2024). "Multiomics for crop improvement," in *Frontier Technologies for Crop Improvement*. Eds. M. K. Pandey, A. Bentley, H. Desmae, M. Roorkiwal and R. K. Varshney (Springer Nature, Singapore), 107–141. doi: 10.1007/978-981-99-4673-0\_6

- Clement, C., De Cristo-Araújo, M., Coppens D'Eeckenbrugge, G., Alves Pereira, A., Picanço-Rodrigues, D., Clement, C. R., et al. (2010). Origin and domestication of native amazonian crops. *Diversity* 2, 72–106. doi: 10.3390/d2010072
- Colli-Silva, M., Richardson, J. E., Figueira, A., and Pirani, J. R. (2024). Human influence on the distribution of cacao: insights from remote sensing and biogeography. *Biodivers. Conserv.* 33, 1009–1025. doi: 10.1007/s10531-023-02777-7
- Colonges, K., Jimenez, J.-C., Saltos, A., Seguíne, E., Looor Solorzano, R. G., Fouet, O., et al. (2022). Integration of GWAS, metabolomics, and sensorial analyses to reveal novel metabolic pathways involved in cocoa fruity aroma GWAS of fruity aroma in Theobroma cacao. *Plant Physiol. Biochem.* 171, 213–225. doi: 10.1016/j.plaphy.2021.11.006
- Conner, J. K., Karoly, K., Stewart, C., Koelling, V. A., Sahli, H. F., and Shaw, F. H. (2011). Rapid independent trait evolution despite a strong pleiotropic genetic correlation. *Am. Nat.* 178, 429–441. doi: 10.1086/661907
- Cornejo, O. E., Yee, M. C., Dominguez, V., Andrews, M., Sockell, A., Strandberg, E., et al. (2018). Population genomic analyses of the chocolate tree, Theobroma cacao L., provide insights into its domestication process. *Commun. Biol.* 1, 1–12. doi: 10.1038/s42003-018-0168-6
- Crouzillat, D., Lerceteau, E., Petiard, V., Morera, J., Rodriguez, H., Walker, D., et al. (1996). Theobroma cacao L.: a genetic linkage map and quantitative trait loci analysis. *Theor. Appl. Genet.* 93, 205–214. doi: 10.1007/BF00225747
- Daymond, A. J., and Hadley, P. (2008). Differential effects of temperature on fruit development and bean quality of contrasting genotypes of cacao (Theobroma cacao). *Ann. Appl. Biol.* 153, 175–185. doi: 10.1111/j.1744-7348.2008.00246.x
- Díaz-Valderrama, J. R., Leiva-Espinoza, S. T., and Aime, M. C. (2020). The history of cacao and its diseases in the Americas. *Phytopathology* 110, 1604–1619. doi: 10.1094/PHYTO-05-20-0178-RVW
- Dijk, A. D. J., Kootstra, G., Kruijer, W., and de Ridder, D. (2021). Machine learning in plant science and plant breeding. *iScience* 24, 101890. doi: 10.1016/j.isci.2020.101890
- Dourojeanni, M. (2021). *Loreto sostenible al 2021* (Lima: Derecho, Ambiente y Recursos Naturales). Available online at: [https://www.dar.org.pe/archivos/publicacion/Loreto2021\\_completo2.pdf](https://www.dar.org.pe/archivos/publicacion/Loreto2021_completo2.pdf). Primera (Accessed January 10, 2025).
- Eftekhari, M., Ma, C., and Orlov, Y. L. (2024). Editorial: Applications of artificial intelligence, machine learning, and deep learning in plant breeding. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1420938
- Engels, J. M. M. (1983). A systematic description of cacao clones. II. The discriminative value of qualitative characteristics and the practical compatibility of the discriminative value of quantitative and qualitative descriptors. *Euphytica* 32, 387–396. doi: 10.1007/BF00021447
- Engels, J. M. M. (1986). *The systematic description of cacao clones and its significance for taxonomy and plant breeding* (Wageningen: Agricultural University Wageningen).
- Fernández-Paz, J., Cortés, A. J., Hernández-Varela, C. A., Mejía-de-Tafur, M. S., Rodríguez-Medina, C., and Baligar, V. C. (2021). Rootstock-mediated genetic variance in cadmium uptake by juvenile cacao (Theobroma cacao L.) genotypes, and its effect on growth and physiology. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.777842
- Fister, A. S., Landherr, L., Maximova, S. N., and Gultinan, M. J. (2018). Transient expression of CRISPR/Cas9 machinery targeting TcNPR3 enhances defense response in Theobroma cacao. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00268
- Fouet, O., Looor Solorzano, R. G., Rhoné, B., Subia, C., Calderón, D., Fernández, F., et al. (2022). Collection of native L. accessions from the Ecuadorian Amazon highlights a hotspot of cocoa diversity. *Plants People Planet* 4, 605–617. doi: 10.1002/ppp3.10282
- Friedman, J. (2020). The evolution of annual and perennial plant life histories: ecological correlates and genetic mechanisms. *Annu. Rev. Ecol. Syst.* 51, 461–481. doi: 10.1146/annurev-ecolsys-110218-024638
- Galvis, D. A., Jaimes-Suárez, Y. Y., Rojas Molina, J., Ruiz, R., León-Moreno, C. E., and Carvalho, F. E. L. (2023). Unveiling cacao rootstock-genotypes with potential use in the mitigation of cadmium bioaccumulation. *Plants* 12, 2941. doi: 10.3390/plants12162941
- Georges, M. E., Melo, C. A. F., de Souza, M. M., and Corrêa, R. X. (2023). Cacao genotypes cultivated in agroforestry systems in Bahia have wide genetic variability in morpho-agronomic characters. *Ciênc. E Agrotecnologia* 47, e004923. doi: 10.1590/1413-7054202347004923
- Gopaulchan, D., Motilal, L. A., Bekele, F. L., Clause, S., Ariko, J. O., Ejang, H. P., et al. (2019). Morphological and genetic diversity of cacao (Theobroma cacao L.) in Uganda. *Physiol. Mol. Biol. Plants* 25, 361–375. doi: 10.1007/s12298-018-0632-2
- Groeneveld, J. H., Tschardtke, T., Moser, G., and Clough, Y. (2010). Experimental evidence for stronger cacao yield limitation by pollination than by plant resources. *Perspect. Plant Ecol. Evol. Syst.* 12, 183–191. doi: 10.1016/j.ppees.2010.02.005
- Ho, W. W., and Smith, S. D. (2016). Molecular evolution of anthocyanin pigmentation genes following losses of flower color. *BMC Evol. Biol.* 16, 98. doi: 10.1186/s12862-016-0675-3
- Hoopen, G. M., Deberdt, P., Mbenoun, M., and Cilas, C. (2012). Modelling cacao pod growth: implications for disease control. *Ann. Appl. Biol.* 160, 260–272. doi: 10.1111/j.1744-7348.2012.00539.x
- Iriarte, J., Elliott, S., Maezumi, S. Y., Alves, D., Gonda, R., Robinson, M., et al. (2020). The origins of Amazonian landscapes: Plant cultivation, domestication and the spread of food production in tropical South America. *Quat. Sci. Rev.* 248, 106582. doi: 10.1016/j.quascirev.2020.106582
- Iwano, A. D., Bekele, F. L., and Butler, D. R. (2003). Evaluation and utilisation of cacao (Theobroma cacao L.) germplasm at the International Cocoa Genebank, Trinidad. *Euphytica* 130, 207–221. doi: 10.1023/A:1022855131534
- Karumamkandathil, R., Uthup, T. K., and Jacob, J. (2022). “Application of omics technologies in Rubber, Cocoa, and Betel nut,” in *Omics in Horticultural Crops*. Eds. G. R. Rout and K. V. Peter (Academic Press, London, England), 501–526. doi: 10.1016/B978-0-323-89905-5.00028-8
- Kongor, J. E., Hinnah, M., de Walle, D. V., Afoakwa, E. O., Boeckx, P., and Dewettinck, K. (2016). Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile — A review. *Food Res. Int.* 82, 44–52. doi: 10.1016/j.foodres.2016.01.012
- Kongor, J. E., Owusu, M., and Oduro-Yeboah, C. (2024). Cocoa production in the 2020s: challenges and solutions. *CABI Agric. Biosci.* 5, 1–28. doi: 10.1186/s43170-024-00310-6
- Kristanto, Y., Tarigan, S., June, T., Sulistyantara, B., and Wijayanti, P. (2024). Indirect use value of improved soil health as natural capital that supports essential ecosystem services: A case study of cacao agroforestry. *Agric. Econ.* 70, 137–154. doi: 10.17221/281/2023-AGRICECON
- Kulesza, E., Thomas, P., Prewitt, S. F., Shalit-Kaneh, A., Wafula, E., Knollenberg, B., et al. (2024). The cacao gene atlas: a transcriptome developmental atlas reveals highly tissue-specific and dynamically-regulated gene networks in Theobroma cacao L. *BMC Plant Biol.* 24, 601. doi: 10.1186/s12870-024-05171-9
- Lanaud, C., Vignes, H., Utge, J., Valette, G., Rhoné, B., Caputi, M. G., et al. (2024). A revisited history of cacao domestication in pre-Columbian times revealed by archaeogenomic approaches. *Sci. Rep.* 14, 2972–2989. doi: 10.1038/s41598-024-53010-6
- La-Rostami, F., Wax, N., Druschka, M., Adams, E., Albert, C., and Fischer, M. (2022). *In vitro* CRISPR-cpf1 assay for differentiation of fine and bulk cocoa (Theobroma cacao L.). *J. Agric. Food Chem.* 70, 8819–8826. doi: 10.1021/acs.jafc.2c02537
- Marita, J. M., Nienhuis, J., Pires, J. L., and Aitken, W. M. (2001). Analysis of genetic diversity in Theobroma cacao with emphasis on witches' broom disease resistance. *Crop Sci.* 41, 1305–1316. doi: 10.2135/cropsci2001.4141305x
- Martirosyan, G. S., Vardanian, I., Tadevosyan, L., Avagyan, A., Adjemyan, G., and Harutunyan, Z. E. (2023). “Evaluation of the genebank germplasm on suitability for use as rootstock in green agriculture,” in *Emerging Issues in Agricultural Sciences*. Ed. A. M. Al-Naggar (B P International, United Kingdom), 31–43. doi: 10.9734/bpi/eias/v6
- McElroy, M. S., Navarro, A. J. R., Mustiga, G., Stack, C., Gezan, S., Peña, G., et al. (2018). Prediction of Cacao (Theobroma cacao) Resistance to Moniliophthora spp. Diseases via Genome-Wide Association Analysis and Genomic Selection. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00343
- Migueluez-Sierra, Y., Hernández-Rodríguez, A., Acebo-Guerrero, Y., Baucher, M., and El Jaziri, M. (2017). *In vitro* micrografting of apical and axillary buds of cacao. *J. Hortic. Sci. Biotechnol.* 92, 25–30. doi: 10.1080/14620316.2016.1215231
- Montenegro, J., Morante Carriel, J., Acosta-Farías, M., Jaimez, R., Carranza Patiño, M., Bru-Martínez, R., et al. (2023). Molecular response of cocoa (Theobroma cacao) to water deficit conditions. *J. Anim. Plant Sci.* 33, 1314–1321. doi: 10.36899/JAPS.2023.6.0671
- Morán, P., Castro, R., Chirinos, D. T., García, L. C., Castro, J., and Kondo, T. (2024). “Sustainable pest management of cacao in the Neotropics: Challenges and opportunities,” in *Sustainable Cacao Cultivation in Latin America*, vol. 23. (London: Routledge).
- Motamayor, J. C., Lachenaud, P., Mota, J. W. da S. e, Looor, R., Kuhn, D. N., Brown, J. S., et al. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (Theobroma cacao L.). *PLoS One* 3, e3311. doi: 10.1371/journal.pone.0003311
- Motamayor, J. C., Risterucci, A. M., Lopez, P. A., Ortiz, C. F., Moreno, A., and Lanaud, C. (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89, 380–386. doi: 10.1038/sj.hdy.6800156
- Motilal, L. A., Zhang, D., Umaharan, P., Mischke, S., Pinney, S., and Meinhardt, L. W. (2011). Microsatellite fingerprinting in the International Cocoa Genebank, Trinidad: accession and plot homogeneity information for germplasm management | Plant Genetic Resources. *Plant Genet. Resour.* 9, 430–438. doi: 10.1017/S147926211100058X
- Moundanga, S. M., Petit, J., Ndangui, C. B., Scher, J., and Nzikou, J.-M. (2024). Impact of cocoa variety on merchant quality and physicochemical characteristics of raw cocoa beans and roasted cocoa mass. *Discov. Food* 4, 1–15. doi: 10.1007/s44187-024-00188-3
- Munjuga, M., Kariuki, W., Njoroge, J. B. M., Ofori, D., and Jamnadass, R. (2013). Effect of rootstock type, scion source and grafting methods on the healing of Allantblackia stuhlmannii grafts under two nursery conditions. *Afr. J. Hortic. Sci.* 7, 1–10.
- N'zi, J.-C., Koné, I., M'bo, K. A. A., Koné, S., and Kouamé, C. (2023). Successful grafting elite cocoa clones (Theobroma cacao L.) as a function of the age of rootstock. *Heliyon* 9, e18732. doi: 10.1016/j.heliyon.2023.e18732
- Niazian, M., and Niedbala, G. (2020). Machine learning for plant breeding and biotechnology. *Agriculture* 10, 436. doi: 10.3390/agriculture10100436
- Nieves-Orduña, H. E., Krutovsky, K. V., and Gailing, O. (2023). Geographic distribution, conservation, and genomic resources of cacao Theobroma cacao L. *Crop Sci.* 63, 1750–1778. doi: 10.1002/csc2.20959

- Nieves-Orduña, H. E., Müller, M., Krutovsky, K. V., and Gailing, O. (2021). Geographic Patterns of Genetic Variation among Cacao (*Theobroma cacao* L.) Populations Based on Chloroplast Markers. *Diversity* 13, 249. doi: 10.3390/d13060249
- Nieves-Orduña, H. E., Müller, M., Krutovsky, K. V., and Gailing, O. (2024). Genotyping of cacao (*Theobroma cacao* L.) germplasm resources with SNP markers linked to agronomic traits reveals signs of selection. *Tree Genet. Genomes* 20, 1–18. doi: 10.1007/s11295-024-01646-w
- Nousias, O., Zheng, J., Li, T., Meinhardt, L. W., Bailey, B., Gutierrez, O., et al. (2024). Three *de novo* assembled wild cacao genomes from the Upper Amazon. *Sci. Data* 11, 369. doi: 10.1038/s41597-024-03215-1
- Oliva-Cruz, M., Goñas, M., García, L. M., Rabanal-Oyarce, R., Alvarado-Chuqui, C., Escobedo-Ocampo, P., et al. (2021). Phenotypic characterization of fine-aroma cocoa from northeastern Peru. *Int. J. Agron.* 2021, 1–12. doi: 10.1155/2021/2909909
- Ortiz-Álvarez, A., Magnitskiy, S., Silva-Arero, E. A., Rodríguez-Medina, C., Argout, X., and Castaño-Marin, Á.M. (2023). Cadmium Accumulation in Cacao Plants (*Theobroma cacao* L.) under Drought Stress. *Agronomy* 13, 2490. doi: 10.3390/agr13102490
- Osorio-Guarín, J. A., Berdugo-Cely, J. A., Coronado-Silva, R. A., Baez, E., Jaimes, Y., and Yockteng, R. (2020). Genome-Wide Association Study Reveals Novel Candidate Genes Associated with Productivity and Disease Resistance to Moniliophthora spp. in Cacao (*Theobroma cacao* L.). *G3 GenesGenomesGenetics* 10, 1713–1725. doi: 10.1534/g3.120.401153
- Pak, S., Li, C., Pak, S., and Li, C. (2022). Progress and challenges in applying CRISPR/Cas techniques to the genome editing of trees. *For. Res.* 2, 1–14. doi: 10.48130/FR-2022-0006
- Pal, P., and Pal, S. (2024). “CRISPR genome editing of woody trees: Current status and future prospects,” in *CRISPRized Horticulture Crops*. Eds. K. A. Abd-Elsalam, A. Ahmad and B. Zhang (Academic Press, London, England), 401–418. doi: 10.1016/B978-0-443-13229-2.00001-6
- Poorter, H., Niklas, K. J., Reich, P. B., Oleksyn, J., Poot, P., and Mommer, L. (2012). Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control. *New Phytol.* 193, 30–50. doi: 10.1111/j.1469-8137.2011.03952.x
- Ramírez-Ramírez, A. R., Mirzaei, K., Menéndez-Grenot, M., Clapé-Borges, P., Espinosa-López, G., Bidot-Martínez, I., et al. (2024). Using ddRADseq to assess the genetic diversity of in-farm and gene bank cacao resources in the Baracoa region, eastern Cuba, for use and conservation purposes. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1367632
- Rauscher, M. D. (2008). Evolutionary transitions in floral color. *Int. J. Plant Sci.* 169, 7–21. doi: 10.1086/523358
- Ricaño-Rodríguez, J., Hipólito-Romero, E., Ramos-Prado, J. M., and Cocoltzi-Vásquez, E. (2019). Genotyping-by-Sequencing of native varieties of *Theobroma cacao* (Malvaceae) from the States of Tabasco and Chiapas, Mexico. *Bot. Sci.* 97, 381–397. doi: 10.17129/botsci.2258
- Sari, W. K., and Utami, N. P. (2024). Successful shoot tip grafting of cacao (*Theobroma cacao* L.) due to the application of plant growth regulators on various concentrations | Sari | Kultivasi. *J. Kultiv.* 23, 35–42. doi: 10.24198/kultivasi.v23i1.46246
- Scharf, A., Lang, C., and Fischer, M. (2020). Genetic authentication: Differentiation of fine and bulk cocoa (*Theobroma cacao* L.) by a new CRISPR/Cas9-based *in vitro* method. *Food Control* 114, 107219. doi: 10.1016/j.foodcont.2020.107219
- Schmidt, J. E., DuVal, A., Puig, A., Tempeleu, A., and Crow, T. (2021). Interactive and dynamic effects of rootstock and rhizobiome on scion nutrition in cacao seedlings. *Front. Agron.* 3. doi: 10.3389/fagro.2021.754646
- Solorzano, R. G. L., Fouet, O., Lemainque, A., Pavet, S., Boccara, M., Argout, X., et al. (2012). Insight into the Wild Origin, Migration and Domestication History of the Fine Flavour Nacional *Theobroma cacao* L. Variety from Ecuador. *PLoS One* 7, e48438. doi: 10.1371/journal.pone.0048438
- Tchatchoua, T. D., Essola, E. E. J., Caspa, R. G., and Donalson, B. B. (2023). Vegetative propagation of cocoa (*Theobroma cacao*) by grafting: Aptitude of grafting on four clones. *J. Hort. For.* 15, 51–57. doi: 10.5897/JHF2023.0711
- Thomas, E., Zonneveld, M., Loo, J., Hodgkin, T., Galluzzi, G., and Etten, J. v. (2012). Present spatial diversity patterns of *theobroma cacao* L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. *PLoS One* 7, e47676. doi: 10.1371/journal.pone.0047676
- Trunschke, J., Lunau, K., Pyke, G. H., Ren, Z.-X., and Wang, H. (2021). Flower color evolution and the evidence of pollinator-mediated selection. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.617851
- Vásquez-Ocmín, P., Cojean, S., Rengifo, E., Suyyagh-Albouz, S., Amasifuen Guerra, C. A., Pomel, S., et al. (2018). Antiprotozoal activity of medicinal plants used by Iquitos-Nauta road communities in Loreto (Peru). *J. Ethnopharmacol.* 210, 372–385. doi: 10.1016/j.jep.2017.08.039
- Yan, J., and Wang, X. (2023). Machine learning bridges omics sciences and plant breeding. *Trends Plant Sci.* 28, 199–210. doi: 10.1016/j.tplants.2022.08.018
- Zarrillo, S., Gaikwad, N., Lanaud, C., Powis, T., Viot, C., Lesur, I., et al. (2018). The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. *Nat. Ecol. Evol.* 2, 1879–1888. doi: 10.1038/s41559-018-0697-x
- Zhang, D., Arevalo-Gardini, E., Mischke, S., Zúñiga-Cernades, L., Barreto-Chavez, A., and Del Aguila, J. A. (2006). Genetic diversity and structure of managed and semi-natural populations of cocoa (*Theobroma cacao*) in the Huallaga and Ucayali Valleys of Peru. *Ann. Bot.* 98, 647–655. doi: 10.1093/aob/mcl146
- Zhang, D., and Motilal, L. (2016a). “Origin, dispersal, and current global distribution of cacao genetic diversity,” in *Cacao Diseases: A History of Old Enemies and New Encounters*, vol. 633. Eds. B. A. Bailey and L. W. Meinhardt (Springer Cham, Switzerland). doi: 10.1007/978-3-319-24789-2\_1
- Zhang, D., and Motilal, L. (2016b). “Origin, dispersal, and current global distribution of cacao genetic diversity,” in *Cacao Diseases: A History of Old Enemies and New Encounters*. Eds. B. A. Bailey and L. W. Meinhardt (Springer International Publishing, Cham), 3–31. doi: 10.1007/978-3-319-24789-2\_1
- Zion Market Research (2025). Global chocolate market size, share, value and forecast 2032. *Zion Mark. Res.* Available at: <https://www.zionmarketresearch.com/report/chocolate-market> (February 4, 2025).