



## OPEN ACCESS

## EDITED BY

Shujaul Mulk Khan,  
Quaid-i-Azam University, Pakistan

## REVIEWED BY

Saraj Bahadur,  
Hainan University, China  
Abdullah Abdullah,  
Quaid-i-Azam University, Pakistan

## \*CORRESPONDENCE

Sixto A. Imán

✉ [siman@inia.gob.pe](mailto:siman@inia.gob.pe)

Angelo F. Samanamud

✉ [asamanamud@inia.gob.pe](mailto:asamanamud@inia.gob.pe)

Juan C. Castro

✉ [juan.castro@unapiquitos.edu.pe](mailto:juan.castro@unapiquitos.edu.pe)

RECEIVED 06 May 2025

ACCEPTED 14 July 2025

PUBLISHED 08 August 2025

## CITATION

Imán SA, Samanamud AF, Ramirez JF, Cobos M and Castro JC (2025) Phenotypic characterization of wild *Myrciaria dubia* (Kunth) McVaugh *ex situ* germplasm bank for breeding, conservation, and sustainable development in the Peruvian Amazon. *Front. Conserv. Sci.* 6:1623515. doi: 10.3389/fcosc.2025.1623515

## COPYRIGHT

© 2025 Imán, Samanamud, Ramirez, Cobos and Castro. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Phenotypic characterization of wild *Myrciaria dubia* (Kunth) McVaugh *ex situ* germplasm bank for breeding, conservation, and sustainable development in the Peruvian Amazon

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> and Juan C. Castro<sup>4,5\*</sup>

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

**Introduction:** *Myrciaria dubia* “camu-camu” is an economically important Amazonian fruit shrub known for its exceptionally high vitamin C content. Despite its commercial value, comprehensive phenotypic characterization of available genetic resources remains limited, hindering breeding programs and conservation strategies. This study aimed to characterize the phenotypic diversity of wild accessions maintained in a 36-year-old *ex situ* germplasm bank, one of the most comprehensive *M. dubia* collections globally, to provide baseline data for genetic improvement and conservation.

**Methods:** We evaluated 43 wild accessions systematically collected from eight major hydrographic basins in the Loreto region using a stratified sampling approach to capture maximum ecological diversity and maintained in an *ex situ* germplasm bank established in 1988 at the National Institute of Agrarian Innovation in Loreto, Peru. Twenty-three quantitative and six qualitative morphological descriptors were assessed using standardized protocols specifically developed for *M. dubia*, emphasizing commercially important descriptors including fruit weight, pulp content, and seed characteristics. Statistical analyses encompassed univariate variability assessment, bivariate correlations, and multivariate classification through hierarchical clustering and principal component analysis.

**Results:** Phenotypic characterization revealed moderate overall variability with coefficient of variation averaging 17.4%, with reproductive descriptors showing greater variation than vegetative traits. Fruit and seed descriptors exhibited the highest variability exceeding 20%, while qualitative descriptors showed limited diversity with Shannon Index of 0.823. Multivariate analysis identified four distinct phenetic groups with no significant correlation to geographic origin (Mantel test,

$p=0.4034$ ). Principal component analysis revealed fruit-related descriptors as primary drivers of phenotypic differentiation, explaining 57.1% of observed variability. Three accessions from phenetic group 3 (PER1000416, PER1000423, and PER1000411) demonstrated superior trait combinations: fruit weight exceeding 13 g, pulp content above 75%, and reduced seed count below 2.5 seeds per fruit.

**Conclusion:** The moderate phenotypic variability observed reflects the natural distribution patterns and limited domestication history of the species. This comprehensive characterization provides essential baseline data and a foundation for targeted breeding programs, conservation strategies, and sustainable production systems supporting development while preserving the genetic diversity in the Peruvian Amazon.

#### KEYWORDS

ascorbic acid, biological variation, breeding, domestication, economic development, multivariate analysis, phenotype

## 1 Introduction

Systematic characterization and conservation of plant genetic resources represent fundamental strategies for addressing global challenges in food security, climate change adaptation, and sustainable agriculture. This is especially critical for native fruit species with untapped nutritional and economic potential that can simultaneously support biodiversity conservation and rural livelihoods in tropical regions. *Myrciaria dubia* (Kunth) McVaugh “camu-camu”, an Amazonian shrub native to floodplain ecosystems, exemplifies such potential through its remarkable nutritional profile (Paiva and Das Chagas, 2016; Castro et al., 2018). With L-ascorbic acid levels ranging from 1,500 to 3,000 mg/100 g of pulp, far exceeding traditional citrus fruits, this species represents one most concentrated sources of this essential vitamin of nature (Villachica, 1996; Imán et al., 2011; Verde and Ríos, 2018; Ferreira, 2020).

Native to riverine ecosystems throughout the Amazon basin, *M. dubia* reaches its highest population density along the Ucayali and Amazon rivers in Peru (Lim, 2012; Verde and Ríos, 2018). The species demonstrates remarkable ecological plasticity, thriving in hot, humid tropical environments with temperatures above 20°C and annual rainfall exceeding 1,200 mm, from sea level to 300 m elevation (Lim, 2012). This adaptability, combined with its nutritional properties, has driven increasing commercial interest. The Loreto region currently produces over 90% of Peru’s national production, yielding 14,226.32 tons in 2024, with consistent growth trends (MIDAGRI, 2025). These production figures underscore the critical need for improved varieties to sustain industry expansion.

*Ex situ* conservation through germplasm banks plays a crucial role in preserving biodiversity and enabling systematic characterization for crop improvement. These collections are

particularly valuable for species like *M. dubia* that offer significant economic opportunities while supporting conservation objectives (Dulloo et al., 2013). In the Peruvian Amazon, where sustainable development remains a priority, developing improved varieties from characterized germplasms could significantly contribute to regional development (Penn, 2006). Evidence from comparable programs reveals compelling benefits: cultivation of high-value native species generates two to three times the income of traditional crops while requiring 40% less land area, with household incomes reaching US\$3,000-5,000 per hectare annually. Furthermore, *M. dubia* agroforestry systems maintain 60-70% of native biodiversity compared to monocultures, demonstrating how economic development can align with conservation goals (Penn, 2006; Newton et al., 2013; Schroth and Ruf, 2013; Araujo et al., 2024).

The nutritional value of *M. dubia* extends beyond its primary claim to fame, as the fruit contains a diverse array of bioactive compounds. Research has identified multiple phenolic compounds, including catechin and epicatechin, various ellagic acid derivatives, and anthocyanins. The phytochemical profile also encompasses flavonols such as rutin and its derivatives, alongside flavanones including naringenin and eriodictyol derivatives (Chirinos et al., 2010). This complex phytochemical profile contributes to the fruit’s antioxidant capacity and potential application in functional foods, natural supplements, and nutraceuticals. However, comprehensive phenotypic characterization of existing germplasm collections has not kept pace with commercial interest, creating a knowledge gap that constrains breeding and conservation efforts.

Previous studies have examined various aspects of *M. dubia* biology, from fruit yield and biochemical composition to genetic variability in both wild and cultivated populations. Notable investigations by Oliva et al. (2005); Puente Ganz (2008); Freitas

et al. (2016); Nunes et al. (2017); Castro et al. (2018), and Sampaio de Paiva (2019) have provided valuable insights into specific traits. Nevertheless, none have comprehensively characterized a long-term *ex situ* collection spanning multiple decades, which is particularly problematic given the increasing commercial interest and the ongoing threats to natural habitats through deforestation and climate change. This gap severely limits the development of improved varieties and effective conservation strategies.

The National Institute of Agrarian Innovation (INIA) in Iquitos, Peru, maintains one of the oldest and most comprehensive *ex situ* collections of *M. dubia* germplasm globally, with 43 accessions representing eight major hydrographic basins in the Loreto region. These accessions were systematically collected using a stratified sampling approach designed to capture maximum ecological diversity (Imán and Aldana, 2007). Established in 1988, this 36-year-old collection represents a crucial resource for understanding the genetic diversity of this species in its center of origin while preserving germplasm that may be vulnerable to habitat loss in the wild. Despite its value for marker-trait associations and breeding decisions, this collection has never undergone systematic phenotypic characterization.

The present study addresses this critical gap by conducting detailed phenotypic characterization of the *ex situ* *M. dubia* germplasm bank. Our objectives were to: (1) assess the phenotypic variability of wild *M. dubia* accessions by comprehensively evaluating qualitative and quantitative descriptors; (2) identify phenetic groups within the germplasm collection using multivariate statistical approaches; (3) determine the relationship between phenotypic variation and geographic origin; and (4) identify promising accessions with superior descriptors for future breeding programs that can contribute to sustainable development in the Peruvian Amazon.

Based on our understanding of the natural history and limited domestication of *M. dubia*, we hypothesized that: (1) wild *M. dubia* *ex situ* germplasm bank would exhibit moderate phenotypic variability, with greater variation in fruit-related descriptors than in vegetative descriptors; (2) phenotypic variation would not show a strong correlation with geographic origin due to the riparian habitat and water-mediated seed dispersal which promotes gene flow across river systems; and (3) distinct phenotypic groups could be identified that harbor promising accessions with superior combinations of commercially valuable descriptors suitable for developing improved varieties. The pronounced variability in reproductive descriptors likely reflects adaptive responses to the dynamic flooding regime of Amazonian rivers, where selection favors larger, more buoyant fruits during high-water periods (December–May) for enhanced dispersal, while simultaneously maintaining seed size plasticity to exploit variable microhabitats during the dry season. This dual selective pressure, combined with outcrossing mating systems and limited gene flow barriers within river systems, has maintained high heterozygosity while preventing strong regional differentiation. These findings will advance scientific understanding of *M. dubia* genetic diversity, accelerate variety development, and strengthen conservation strategies for this valuable Amazonian resource.

## 2 Materials and methods

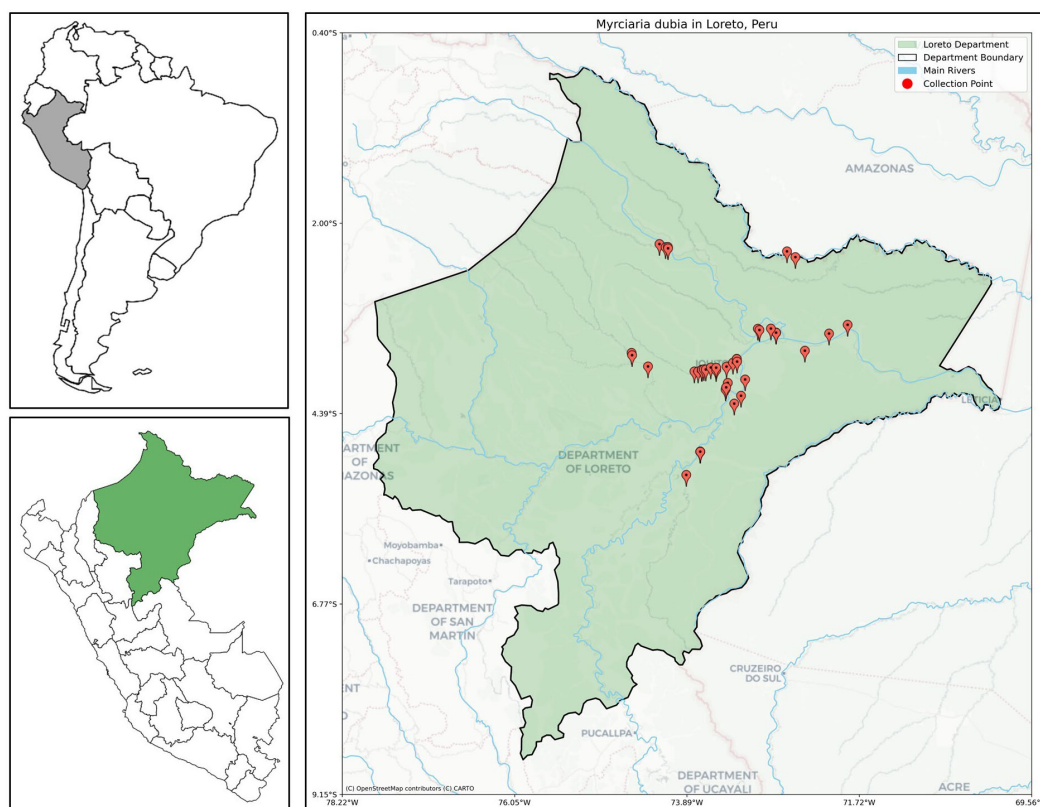
### 2.1 Origin and establishment of the *ex situ* germplasm bank

This study examined 43 accessions of *M. dubia* maintained in an *ex situ* germplasm bank at the Experimental Field “El Dorado” (03°57' 17"S, 73°24'55"W, 112 m.a.s.l), located at km 25.5 of the Iquitos-Nauta highway in Loreto, Peru. The site experiences typical Amazonian conditions: tropical humid climate with a mean annual temperature of 26.83°C, relative humidity of 89.5%, and average monthly rainfall of 272.5 mm. Established in 1988, this 36-year-old germplasm collection captures natural genetic diversity of eight major hydrographic basins in the Loreto region: Amazonas, Curaray, Itaya, Nanay, Napo, Putumayo, Tigre, and Ucayali (Figure 1). These basins were strategically selected to represent both the accessibility constraints of remote Amazonian locations and the distribution patterns of major *M. dubia* populations. Within these basins, the 43 wild accessions were systematically collected using a stratified sampling approach designed to capture maximum ecological diversity, targeting populations from contrasting microhabitats including várzea (seasonally flooded forests), terra firme margins, oxbow lakes, and tributary confluences, thereby ensuring representation of the full spectrum of environmental conditions that shape the species' adaptive variation across its natural range.

Owing to the lack of effective vegetative propagation techniques at that time, a seed-based germplasm bank approach was implemented following protocols for forest seed handling (Willian, 1985). Healthy, mature plants were identified in 43 natural populations from which fully ripe fruits were harvested. The seeds were extracted, washed, air-dried, and germinated in shaded beds. After two months, the resulting seedlings were transferred to unshaded seedbeds with 10 cm spacing between plants and rows. Seedlings were transplanted at a height of 50 cm to field positions with 4×4 m spacing. This seed-based approach was necessary because of the lack of effective vegetative propagation methods. A germination rate of 100% was achieved through optimal seed handling and immediate sowing after extraction, which is consistent with the recalcitrant seed behavior of *M. dubia*. Each accession comprises 10 individuals maintained under uniform management conditions. Vegetative (agamic) propagation methods such as air layering and central cleft grafting have been developed and are now employed to produce uniform, high-quality planting materials. The complete passport data are provided in [Supplementary Table S1](#).

### 2.2 Sampling design for phenotypic characterization

The timing of phenotypic evaluation is critical to ensure accurate descriptor measurements and data reliability. After 35 years of establishment, the germplasm collection reached full maturity with stable phenotypic expression, enabling comprehensive characterization during the 2023–2024 fruiting season. The evaluation period from December 2023 to March



**FIGURE 1**  
 Geographic distribution of 43 *M. dubia* wild accessions collected from eight major hydrographic basins in the Peruvian Amazon. Red markers indicate collection sites distributed across the Loreto region, encompassing the Amazonas, Curaray, Itaya, Nanay, Napo, Putumayo, Tigre, and Ucayali river systems. Insets show Peru’s position in South America (top left) and the Loreto region location within Peru (bottom left).

2024 was strategically selected to coincide with peak flowering and fruiting when morphological and agronomic descriptors achieved their most representative expression under the local climatic conditions of the Peruvian Amazon.

A systematic sampling approach was implemented to balance statistical rigor with practical constraints of germplasm evaluation. The sampling design followed established protocols for tropical fruit germplasm characterization (Hernández-Delgado et al., 2018; Morales et al., 2019), selecting five individuals from ten plants maintained per accession for detailed evaluation. The 50% sampling intensity was determined through power analysis and validated against practical constraints including: (1) spatial heterogeneity within accession blocks requiring stratified selection, (2) potential edge effects minimized by selecting from central positions, (3) phenological synchrony challenges addressed through multiple observation periods, and (4) resource optimization for comprehensive descriptor measurements. This approach balanced statistical rigor with logistical feasibility while ensuring adequate representation of within-accession variability.

Quality control measures were integrated throughout the sampling process to ensure data reliability and minimize potential bias. Plant selection within each accession followed predetermined

criteria designed to minimize edge effects and microenvironmental variation, with individuals chosen from central positions within accession blocks representing the full range of observed phenotypic expression. The selection criteria prioritized plants exhibiting typical morphological characteristics, good health status, and consistent vigor, with standardized parameters applied uniformly across all accessions. The extended establishment period of the collection represents a particular strength of this study, as three decades of cultivation under consistent management conditions likely stabilized phenotypic expression, allowing a clearer manifestation of genetic potential while reducing environmental noise in the dataset.

### 2.3 Quantitative and qualitative descriptor measurement procedures

Data collection encompassed the leaves, flowers, fruits, and seeds of the selected plants across all 43 accessions, following standardized descriptors for *M. dubia* (Imán et al., 2022). We evaluated twenty-three quantitative descriptors and six qualitative descriptors (Supplementary Table S2). For each plant, there were

10 fully expanded leaves from the middle canopy third, 10 flowers at anthesis, 10 mature fruits at commercial ripeness, and all seeds from the sampled fruits.

Measurements followed standardized protocols with calibrated instruments: linear dimensions (leaf length/width, petiole length, flower parts, fruit size, and seed size) using rulers and digital Vernier calipers (Mitutoyo® 6" × 150 mm); mass (fruit weight, pulp weight, and seed weight) using analytical balances (precision ± 0.01 g); sugar content (°Brix) using a digital refractometer (HANNA® HI96801 calibrated with purified water); and pH using a digital potentiometer (Bonajay® PH-009 III, calibrated with buffers at pH 4.0 and 7.0). All the instruments were calibrated before each measurement session. To ensure data reliability, all measurements were conducted by a trained four-person team following standardized operating procedures with inter-observer calibration sessions. Cross-validation protocols included 10% re-measurement by different observers, achieving >95% concordance for quantitative traits and 100% agreement for qualitative descriptors, thereby minimizing measurement bias across the multi-month evaluation period. For all quantitative data, the measurements were averaged across ten samples per plant to obtain representative values. Qualitative data (leaf shape, apex shape, base shape, base angle, leaf margin, and the presence of anthocyanins) were recorded as modal categories. [Supplementary Figure S1](#) illustrates the key morphological features measured during characterization, including leaf length, petiole length, style length, flower characteristics, presence of anthocyanin pigmentation, fruit dimensions, sugar content, and seed parameters.

## 2.4 Integrated multi-level statistical analysis for germplasm diversity assessment

We employed a three-tiered statistical framework that systematically progressed through increasing levels of complexity ([Supplementary Figure S2](#)). The analysis began with univariate methods to examine individual descriptor distributions and assess variability patterns. These initial findings informed our bivariate assessments, which revealed associations and correlations between descriptors. Finally, multivariate techniques identified underlying patterns and group structures within the germplasm collection. This hierarchical approach ensured that each analytical level built upon insights from the previous stage, ultimately providing a comprehensive understanding of diversity patterns across multiple dimensions. All statistical procedures were performed using Python (version 3.11.6) selected for its extensive scientific libraries enabling reproducible workflows and handling of large multidimensional datasets, while the Julius computational platform (<https://julius.ai/>) provided cloud-based processing power essential for computationally intensive multivariate analyses. This combination ensured both methodological transparency and computational efficiency for analyzing 23 quantitative and 6 qualitative descriptors across 215 individual plants.

### 2.4.1 Univariate analysis

For quantitative data, we performed descriptive statistical analysis using summary measures and calculated the Coefficient of Variation ( $CV = SD/mean \times 100$ ) for each descriptor to assess variability levels (Pearson, 1896). The overall variability of the germplasm was determined by averaging the CVs across all descriptors. CV values were categorized as low (<10%), moderate (10-20%), or high (>20%). While most variables exhibited interpretable patterns of dispersion, not all met the assumption of normality, which was considered when selecting the subsequent analytical approaches.

For qualitative data, we calculated the frequencies of each category and estimated Shannon's Diversity Index (H) for each descriptor (Shannon, 1948) for all samples and at the accession levels ([Supplementary Table S3](#)). The overall qualitative variability was obtained by averaging the Shannon indices across all qualitative descriptors.

### 2.4.2 Bivariate analysis

For quantitative data, we used Spearman's Rho test (Spearman, 1904) to measure the strength and direction of the associations between descriptors. This non-parametric method was preferred over Pearson's correlation because not all descriptors followed a normal distribution, making Spearman's correlation more suitable for capturing monotonic relationships without requiring normality. This analysis helped to identify strongly correlated descriptors ( $\rho > 60\%$ ), allowing the identification of descriptors that could be excluded from future analyses to reduce redundancy.

For qualitative data, Cramer's V test (Cramér, 1946) was employed to measure association strength. Descriptors with strong associations (Cramer's V > 50%) were identified to avoid redundancy in the subsequent analyses.

### 2.4.3 Multivariate analysis

To determine variability patterns and estimate phenetic groups and their relationships, we constructed a phenogram using Ward's hierarchical method (Ward, 1963), as modified by Murtagh and Legendre (2014), employing the Rogers-Tanimoto distance measure (Rogers and Tanimoto, 1960), appropriate for mixed quantitative-qualitative data types, to determine the degree of relationship between the accessions. The optimal number of phenetic groups ([Supplementary Figure S3](#)) was determined using the elbow method (Thorndike, 1953) and validated using the cophenetic correlation coefficient ([Supplementary Table D4](#)).

Principal Component Analysis (PCA) was performed on the standardized data to determine how many eigenvalues could explain the variability in the dataset, identify the contribution of various descriptors to each principal component (Hotelling, 1933), and determine how the eigenvectors contribute to the variability of each component. Because PCA is also a deterministic method, no random initialization is required. Components with eigenvalues >1 were retained following the Kaiser criterion (Kaiser, 1960) and were

TABLE 1 Summary statistics of 23 quantitative descriptors measured in 215 plants from 43 accessions of *M. dubia ex situ* germplasm bank.

Descriptor	Average	SD	Min	Max	Median	Skewness	Kurtosis	CV %	
Leaf length (cm)	8.48	0.90	6.60	10.64	8.40	0.01	-0.64	10.61	Moderate
Leaf width (cm)	3.57	0.37	2.85	4.50	3.50	0.21	-0.77	10.31	Moderate
Petiole length (cm)	0.65	0.13	0.40	0.99	0.65	0.40	-0.42	20.82	High
Sepal width (cm)	0.26	0.04	0.17	0.36	0.25	0.37	-0.18	14.20	Moderate
Sepal length (cm)	0.21	0.03	0.14	0.30	0.21	0.45	-0.06	16.52	Moderate
Petal width (cm)	0.37	0.04	0.26	0.45	0.37	-0.29	-0.42	11.30	Moderate
Petal length (cm)	0.42	0.04	0.30	0.50	0.42	-0.39	-0.19	10.35	Moderate
Pedicle length (cm)	0.31	0.07	0.15	0.48	0.30	0.07	-0.20	21.42	High
Stamen length (cm)	0.95	0.17	0.55	1.40	0.93	0.19	-0.32	18.03	Moderate
Style length (cm)	1.14	0.13	0.83	1.41	1.15	-0.14	-0.43	11.33	Moderate
Fruit length (cm)	2.48	0.49	1.10	3.85	2.51	-0.21	-0.14	19.92	Moderate
Fruit width (cm)	2.65	0.53	1.15	3.90	2.67	-0.07	-0.42	19.93	Moderate
Fruit weight (g)	10.35	2.15	5.20	15.89	10.17	0.11	-0.28	20.82	High
Shell weight (g)	1.07	0.21	0.60	1.56	1.07	0.04	-0.95	20.07	High
Pulp weight (g)	7.80	1.61	3.75	11.73	7.83	0.03	-0.20	20.63	High
Sugar content (°Brix)	6.51	0.88	3.90	9.40	6.50	0.26	0.41	13.55	Moderate
Hydrog. potential (pH)	2.86	0.09	2.65	3.11	2.87	-0.08	-0.68	3.24	Low
Seeds number	2.73	0.65	1.00	4.00	3.00	0.11	-0.42	23.70	High
Seed length (cm)	1.40	0.31	0.57	2.13	1.38	0.07	-0.78	22.13	High
Seed width (cm)	1.08	0.24	0.39	1.72	1.07	0.20	-0.29	22.14	High
Seed thickness (cm)	0.33	0.08	0.17	0.50	0.33	0.14	-0.76	22.44	High
Seed weight (g)	0.56	0.13	0.29	0.89	0.55	0.16	-0.75	23.07	High
Total seeds weight (g)	1.48	0.35	0.67	2.40	1.49	0.17	-0.30	23.76	High

Standard deviation (SD), coefficient of variation (CV).

confirmed by Scree plot examination (Supplementary Figure S4). A 3D Biplot graph was created to visualize the relationship between descriptors with respect to the accessions across the three most important principal components (Gabriel, 1971), considering the phenetic groups formed.

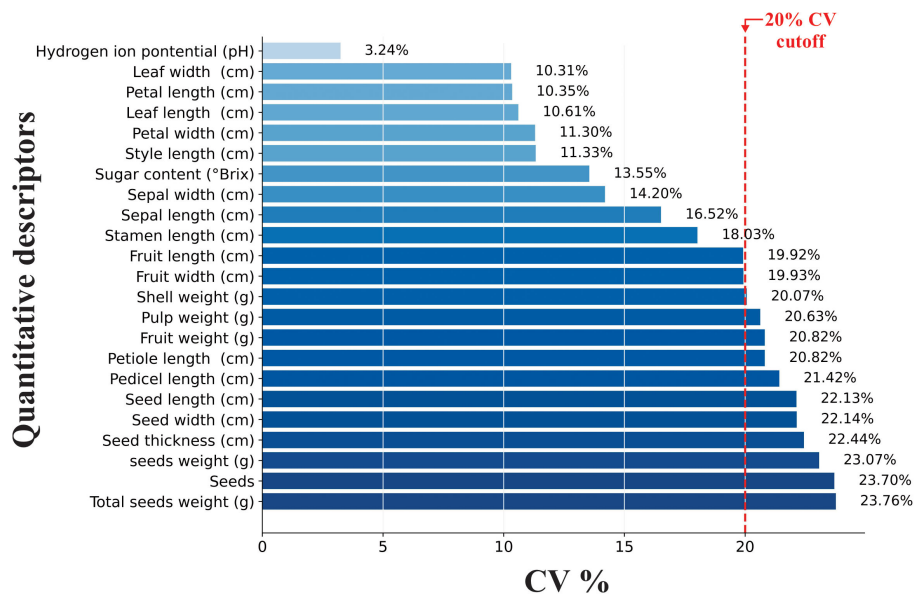
Multivariate Analysis of Variance (MANOVA) was used to test group differentiation significance, with *post-hoc* Hotelling's  $T^2$  (Hotelling (1931)) tests for pairwise comparisons. In addition to determine significant differences between phenetic groups were used multiple test criteria for robustness, including Wilks' Lambda (Wilks, 1938), Pillai's Trace (Pillai, 1955), the Lawley-Hotelling trace (Lawley, 1938), and Roy's Largest Root (Roy, 1958), which focuses on the strongest linear combination separating the groups. These results were corroborated and complemented by Fisher's Analysis of Variance (ANOVA) (Fisher, 1925) and Scott and Knott's (1974) comparative tests to identify specific differences in descriptors of interest between phenetic groups.

The Mantel test (Mantel, 1967) was used to determine the relationship between the phenotypic and geographic distance (Supplementary Figure S5), and the distribution of clusters was mapped across the Loreto Region. Finally, we performed a detailed descriptive analysis of the accessions in the clusters that were most closely related to desirable fruit descriptors.

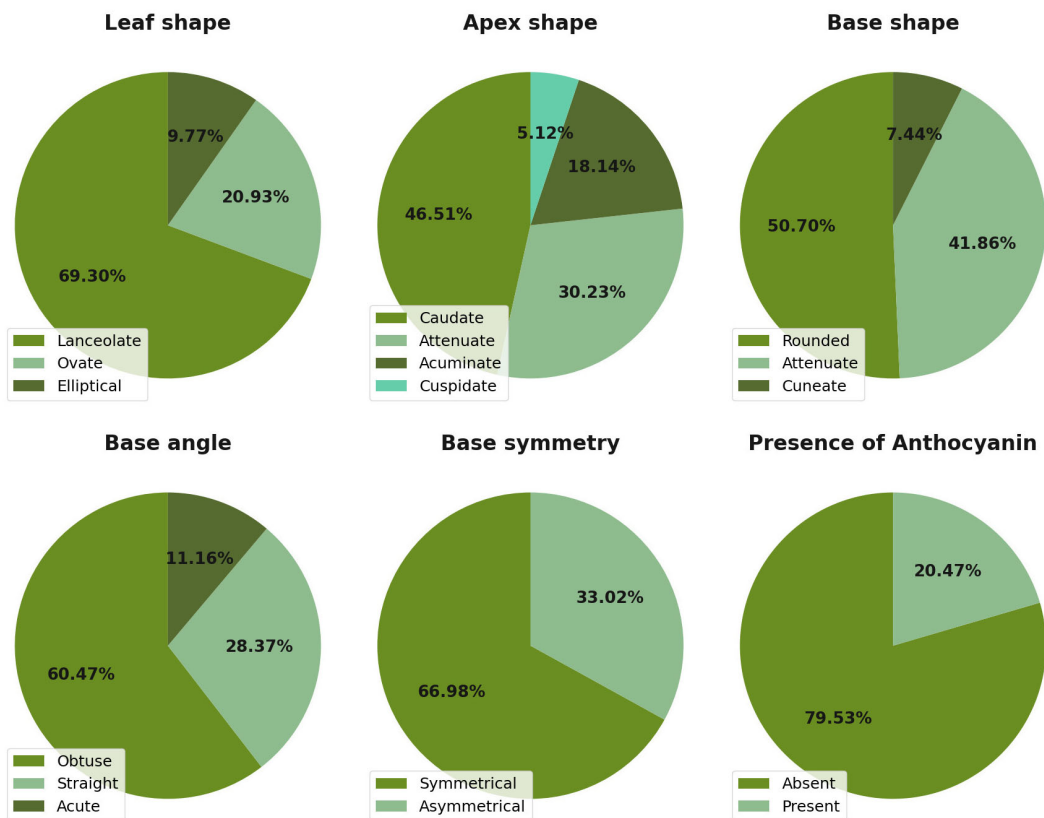
## 3 Results

### 3.1 Variability of quantitative descriptors

The *M. dubia* germplasm bank exhibited distinct patterns of phenotypic variability that aligned with our hypothesis regarding differential variation in descriptors. The descriptive statistics revealed a moderate overall variability (mean CV = 17.4%) across all descriptors, with descriptors falling into three distinct categories



**FIGURE 2** Hierarchical variability pattern of 23 quantitative morphological descriptors in *M. dubia* germplasm bank, arranged by increasing coefficient of variation. Dark blue bars represent high-variability traits (CV>20%), light blue bars represent moderate variability (10-20%), and the lightest bar represents low variability (<10%). The red dashed line indicates the 20% CV threshold.



**FIGURE 3** Frequency distribution of six qualitative morphological descriptors revealing limited polymorphism in the *M. dubia ex situ* germplasm bank (n = 215 plants). Bar charts display the percentage of plants exhibiting each category for leaf shape, apex shape, base shape, base angle, base symmetry, and anthocyanin presence.

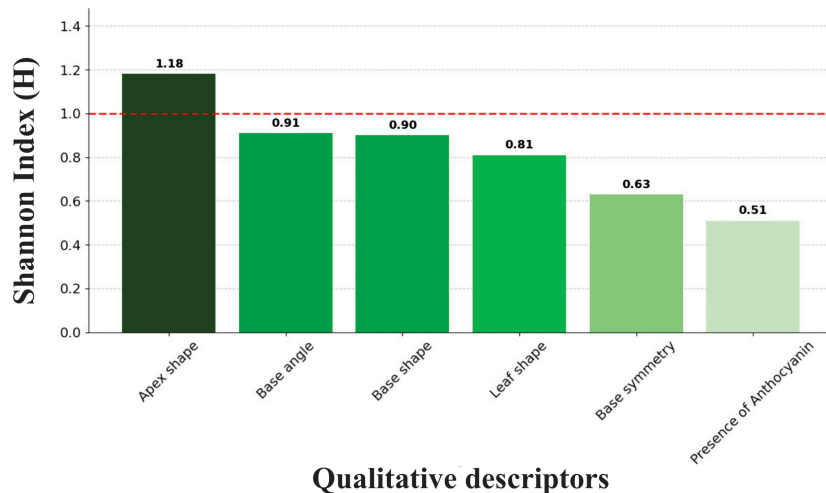


FIGURE 4

Shannon diversity index (H) ranking for six qualitative descriptors demonstrates limited morphological polymorphism in the *M. dubia ex situ* germplasm bank. Bars are arranged in descending order of diversity. The red dashed line indicates the moderate diversity threshold (H=1.0).

(Table 1, Figure 2): low variability (CV = 3.24%), moderate variability (CV from 10.31 to 19.93%), and high variability (CV from 20.07 to 23.76%).

A clear hierarchical pattern of variability emerged across the 23 quantitative descriptors, with reproductive structures demonstrating substantially greater variation than vegetative traits (Figure 2). Seed-related descriptors exhibited the highest variability levels, all exceeding the 20% coefficient of variation threshold, with total seed weight showing maximum variation at 23.76%. Fruit mass components displayed similarly high variability, including fruit weight (20.82%) and pulp weight (20.63%). Among vegetative traits, petiole length emerged as a notable exception with high variability (20.82%), while among floral descriptors, pedicel length showed exceptional variation (21.42%), distinguishing these two structural traits from their respective categories.

In contrast, the remaining vegetative and floral descriptors demonstrated more conservative variation patterns. Leaf dimensions maintained moderate variability around 10%, indicating morphological stability in foliar traits. Floral measurements, excluding pedicel length, remained within the moderate range, suggesting relatively stable flower morphology across the germplasm collection. Biochemical parameters revealed the most striking contrast in variability patterns: while sugar content showed moderate variation (13.55%), pH exhibited remarkable stability across all accessions with the lowest coefficient of variation of any measured trait (3.24%). This hierarchical distribution of variability from high in reproductive structures to low in biochemical traits characterized the phenotypic diversity pattern within the germplasm bank.

These patterns of differential variability between descriptor categories confirm our first hypothesis that the wild *M. dubia* germplasm exhibits moderate overall phenotypic variability with greater reproductive structure variation than vegetative descriptors.

### 3.2 Distribution patterns of qualitative descriptors

Qualitative descriptor assessment across the germplasm bank revealed limited polymorphism, with most traits showing restricted variation. Leaf shape exhibited the highest degree of morphological uniformity, with lanceolate forms comprising 69.3% of the 215 assessed plants (Figure 3). The remaining accessions displayed ovate (20.93%) and elliptical (9.77%) leaf shapes. Additional qualitative traits demonstrated similarly limited diversity: base angle was predominantly obtuse (60.47%), base shape was evenly distributed between rounded (50.70%) and attenuated (41.86%) forms, and base symmetry was primarily symmetrical (66.98%).

Apex shape represented the most variable qualitative descriptor, displaying a more even distribution across four morphological categories. Caudate forms were most frequent at 46.51% of accessions, followed by attenuate (30.23%), acuminate (18.14%), and cuspidate (5.12%) variants. Anthocyanin pigmentation occurred in 79.53% of all assessed plants, while 20.47% lacked this trait. These frequency distributions across the six qualitative descriptors characterized the morphological diversity patterns within the *ex situ* collection, with apex shape alone showing balanced representation across multiple trait states.

### 3.3 Diversity analysis of qualitative descriptors

Shannon diversity analysis of qualitative descriptors revealed limited polymorphism across the germplasm bank. The overall mean diversity index of  $H = 0.823$  fell below the moderate diversity threshold of 1.0, with individual descriptors displaying a hierarchical pattern of variability (Figure 4). The diversity values

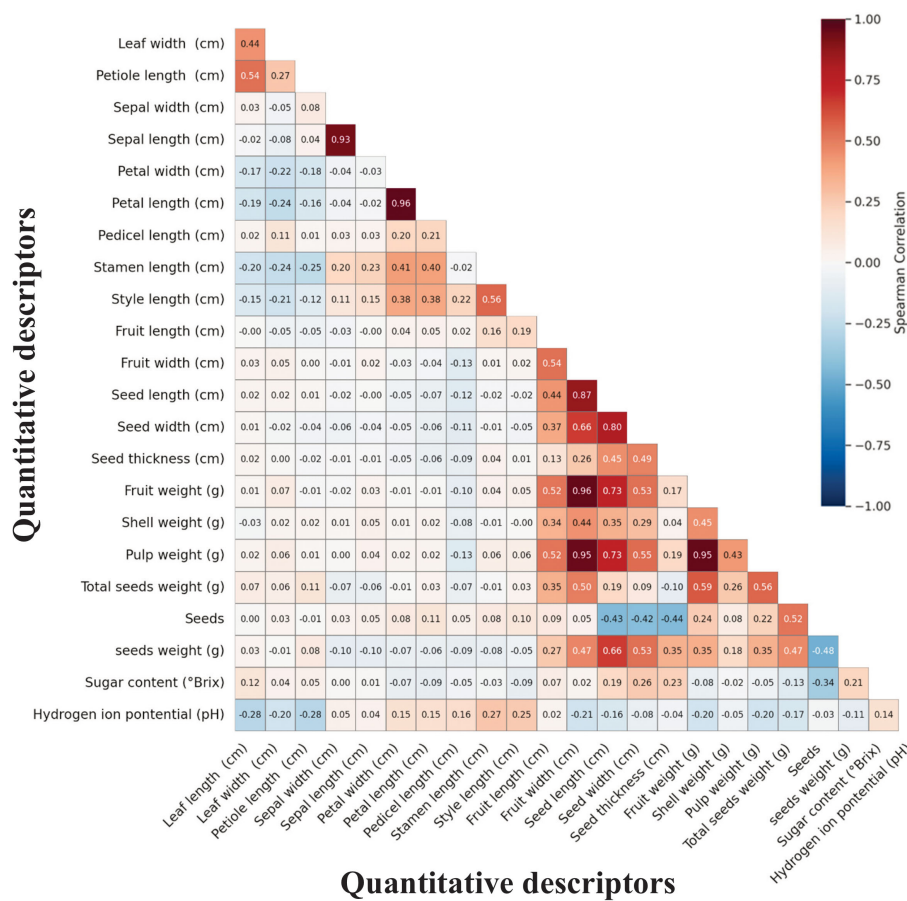


FIGURE 5 Spearman correlation matrix for 23 quantitative descriptors in the *M. dubia* germplasm bank. Descriptors are arranged by hierarchical clustering. Color scale ranges from dark blue (strong negative correlation) through white (no correlation) to dark red (strong positive correlation).

ranged from 0.51 to 1.18, demonstrating substantial variation in polymorphism levels among different morphological traits. Only the apex shape surpassed the moderate diversity threshold with  $H = 1.18$ , distinguishing it as the most variable qualitative descriptor in the germplasm bank.

The remaining five descriptors exhibited progressively lower diversity values, forming a clear gradient of morphological variability. Base angle ( $H = 0.91$ ) and base shape ( $H = 0.90$ ) approached but did not reach the moderate diversity threshold, followed by leaf shape with  $H = 0.81$ . The lowest diversity values were recorded for base symmetry ( $H = 0.63$ ) and presence of anthocyanins ( $H = 0.51$ ), with the latter showing the most limited variation among all assessed qualitative traits. This hierarchical distribution of diversity indices characterized the morphological variability structure within the collection.

Complementing the collection-level patterns, within-accession analysis demonstrated considerable heterogeneity for most descriptors. Apex shape maintained its distinction as the most variable trait, with 97.7% of accessions (42 of 43) containing multiple morphological variants (Supplementary Figure S6; Supplementary Table S3). Similar within-accession polymorphism was observed for other foliar descriptors, with approximately 75%

of accessions exhibiting variation in base shape, base angle, and leaf shape. In contrast, anthocyanin presence showed within-accession variation in only 44.2% of accessions, consistent with its low collection-level diversity.

### 3.4 Correlation patterns among quantitative descriptors

Spearman correlation analysis revealed distinct patterns of trait associations across the 23 quantitative descriptors, with correlations ranging from strongly negative to strongly positive relationships (Figure 5). The correlation matrix displayed clear clustering of functionally related traits, with the strongest associations occurring within morphological categories rather than between them. Three primary correlation patterns emerged: tightly correlated fruit mass components, inverse relationships between seed number and seed size, and largely independent variation among vegetative, floral, and reproductive structures.

Fruit mass components formed the most cohesive correlation cluster in the dataset. Fruit weight, pulp weight, and fruit width demonstrated exceptionally strong positive associations with

correlation coefficients exceeding 0.90, indicating these traits vary in concert. Shell weight and total seed weight showed moderate to strong positive correlations with fruit mass traits, ranging from 0.55 to 0.73. In contrast, biochemical parameters displayed minimal associations with morphological traits, with pH showing particularly weak correlations across all descriptors.

A notable inverse relationship characterized seed traits, revealing a fundamental trade-off in reproductive allocation. Seed number exhibited consistent negative correlations with all individual seed dimensions, with coefficients ranging from -0.42 to -0.48. Conversely, individual seed measurements showed strong positive intercorrelations among themselves, with values between 0.67 and 0.80 for length, width, and thickness combinations. This pattern extended to individual seed weight, which correlated positively with seed dimensions but negatively with seed count.

Vegetative and floral descriptors demonstrated largely independent variation from reproductive traits. Within floral structures, petal and sepal dimensions formed tightly correlated pairs, with coefficients reaching 0.96 for petal measurements and 0.93 for sepal traits. Vegetative descriptors showed weaker intercorrelations, with leaf dimensions displaying only moderate associations. Cross-category correlations between vegetative and reproductive structures remained consistently low, with most coefficients below 0.30, indicating these trait categories vary independently across the germplasm collection.

### 3.5 Association patterns among qualitative descriptors

Association analysis among qualitative descriptors revealed structured relationships that inform efficient phenotyping strategies. Base angle and base shape showed the strongest interdependence ( $V = 0.667$ ,  $p < 0.001$ ), indicating these traits are largely redundant for characterization purposes (Figure 6). This strong association suggests measuring one trait could predict the other with reasonable accuracy, streamlining data collection protocols.

Conversely, several descriptor pairs showed complete independence, providing additional dimensions for accession characterization. The absence of association between leaf shape and base symmetry ( $V = 0.000$ ) or between base symmetry and anthocyanin presence indicates these traits vary independently, potentially reflecting different genetic control mechanisms or adaptive responses. Understanding these association patterns enables more efficient germplasm characterization by focusing on non-redundant traits that maximize discriminatory power.

### 3.6 Cluster analysis and phenetic group identification

Hierarchical cluster analysis successfully classified the germplasm collection into four distinct phenetic groups, providing a framework for understanding morphological diversity patterns (Figure 7). The optimal number of phenetic groups within

the *M. dubia ex situ* germplasm bank was determined through the elbow method (Supplementary Figure S3). At this classification level, the sum of the squared distances within clusters was 634, with the curve showing a notable change in slope that designated four as the most efficient number of clusters for the classification of accessions (Supplementary Table S5).

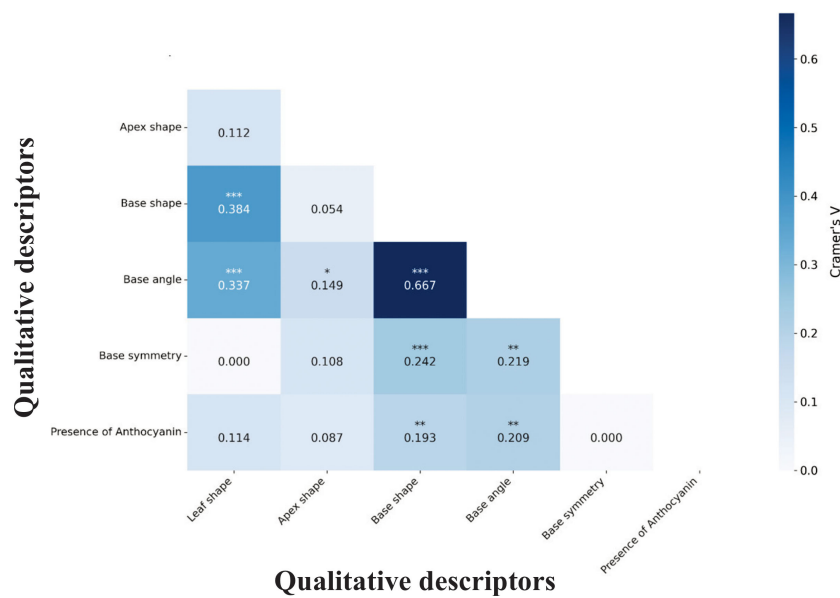
The phenogram constructed using Ward's modified hierarchical method and Rogers-Tanimoto distance measure (Figure 7; Supplementary Table S4) confirmed the four-group classification structure. The dendrogram exhibited clear differentiation among phenetic groups when sectioned at the 0.55 distance threshold (red dashed line). The cophenetic correlation coefficient ( $r = 0.677$ ; Supplementary Table S4) confirms the robustness of our clustering approach, indicating a good representation of the original distance matrix by the clustering structure, and supporting the validity of the phenetic relationships portrayed in the dendrogram.

The 43 *M. dubia* accessions were unevenly distributed across the four phenetic groups (Supplementary Table S5). Phenetic groups 1 and 4 each contained 11 accessions, whereas groups 2 and 4 comprised 10 accessions. In phenetic group 1, accessions PER1000421–PER1000409 formed a distinct branch that was separated at approximately 0.35 distance units. Group 2 included accessions PER1000388 through PER1000425, which diverged from the remaining groups by approximately 0.45 distance units. Group 3 contained accessions PER1000405–PER1000423, branching at approximately 0.40 distance units. Finally, group 4 included accessions PER1000396 through PER1000408, which formed the most distinct branch, separated by a 0.48 distance level from the other clusters.

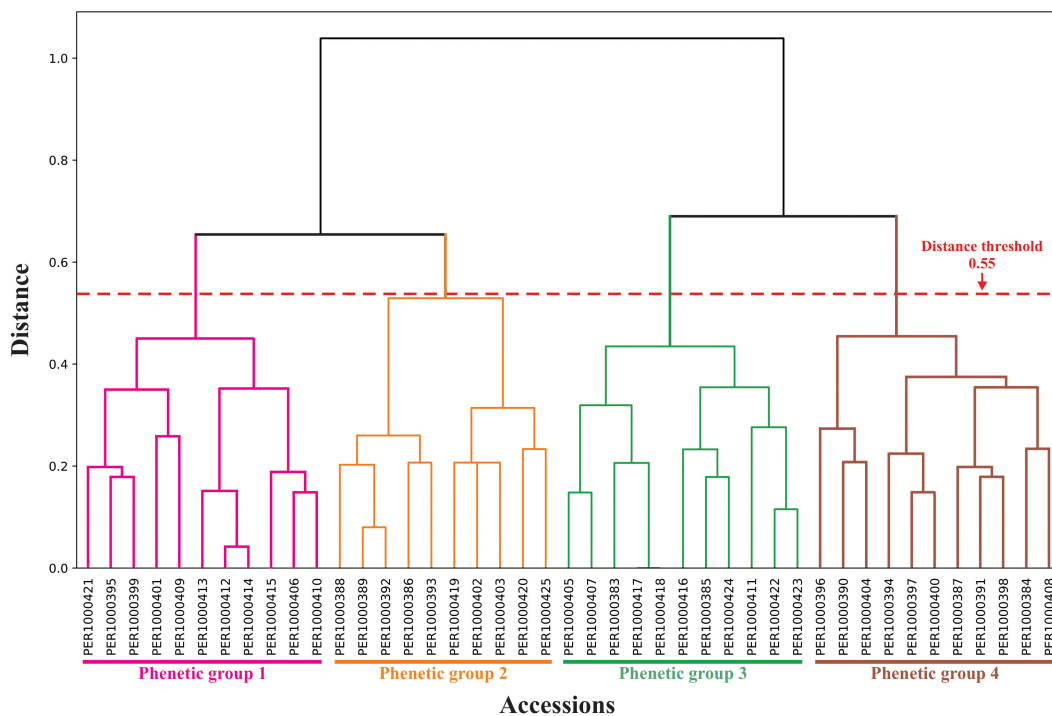
The dendrogram revealed distinct patterns of internal cohesion within each phenetic group. Group 1 showed the highest within-group heterogeneity, with multiple sub-clusters branching at distances ranging from 0.15 to 0.45. Group 3 displayed intermediate cohesion with most accessions clustering below 0.35 distance units, suggesting moderate phenotypic uniformity. Group 2 exhibited a bifurcated structure with two main sub-clusters separating at approximately 0.32 distance. Group 4 demonstrated the tightest internal cohesion, with most branching occurring below 0.30 distance units, indicating greater phenotypic similarity among its members. Notably, no single accession appeared as an outlier, confirming that all 43 accessions were successfully integrated into coherent phenetic groups at the established threshold. This absence of outliers supports the comprehensive nature of the morphological variation captured within the four groups and validates the classification system.

### 3.7 Contributions of quantitative descriptors to principal components

Principal component analysis (PCA) of the 23 quantitative descriptors revealed a structured pattern of phenotypic variation, with the first seven components explaining 82.9% of total variance (Table 2; Supplementary Figure S4). The analysis identified distinct



**FIGURE 6** Cramer's V association matrix for six qualitative descriptors in the *M. dubia* germplasm bank. Values indicate association strength between descriptor pairs. Statistical significance levels: \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ .



**FIGURE 7** Hierarchical cluster dendrogram of 43 *M. dubia* accessions based on morphological characterization data. Clustering performed using Ward's modified method with Rogers-Tanimoto distance. The horizontal red dashed line indicates the distance threshold of 0.55 used to define cluster groups. Branch colors correspond to four phenetic groups: Group 1 (pink), Group 2 (orange), Group 3 (green), and Group 4 (brown). Each terminal branch represents one of the 43 accessions from the *ex situ* germplasm bank.

TABLE 2 Principal component analysis of 23 quantitative descriptors in accessions of the *M. dubia ex situ* germplasm bank: eigenvalues, eigenvectors, and explained variance.

Eigenvector	Eigenvalue	PC1	PC2	PC3	PC4	PC5	PC6	PC7
	5.987	4.032	3.115	2.078	1.744	1.119	0.981	
Leaf length	-0.006	0.300	-0.040	0.096	0.477	0.045	-0.046	
Leaf width	-0.069	0.298	-0.074	0.208	0.108	-0.369	-0.150	
Petiole length	0.031	0.302	-0.039	0.041	0.461	0.079	-0.157	
Sepal width	0.058	-0.046	-0.233	-0.558	0.212	-0.121	-0.083	
Sepal length	0.044	-0.104	-0.232	-0.548	0.234	-0.106	-0.061	
Petal width	0.021	-0.389	-0.031	0.263	0.184	0.064	-0.346	
Petal length	0.029	-0.382	-0.015	0.274	0.187	0.031	-0.360	
Pediceal length	0.075	-0.083	-0.222	0.187	0.128	-0.718	0.024	
Stamen length	0.022	-0.373	-0.073	-0.125	0.214	0.266	0.080	
Style length	0.082	-0.355	-0.164	0.053	0.115	0.043	0.174	
Fruit length	-0.166	-0.120	-0.059	0.087	0.217	-0.076	0.686	
Fruit width	-0.388	-0.025	-0.130	-0.026	-0.014	-0.032	-0.093	
Seed length	-0.373	-0.056	0.137	-0.102	0.041	-0.117	-0.102	
Seed width	-0.344	-0.083	0.223	-0.033	0.018	-0.086	-0.062	
Seed thickness	-0.200	-0.099	0.354	-0.152	0.062	-0.024	-0.099	
Fruit weight	-0.357	-0.009	-0.244	0.001	-0.056	-0.020	-0.098	
Shell weight	-0.270	-0.097	-0.053	0.014	-0.111	-0.151	0.246	
Pulp weight	-0.366	-0.024	-0.205	0.004	-0.035	0.066	-0.097	
Total seed weight	-0.233	0.052	-0.334	0.189	0.065	0.235	0.065	
Seeds number	0.034	0.037	-0.506	0.171	-0.141	0.183	0.013	
Seed weight	-0.287	-0.017	0.257	0.032	0.187	0.045	0.028	
Brix degrees	0.045	0.038	0.185	0.145	0.418	0.120	0.253	
Hydrogenion potential	0.160	-0.311	0.122	0.080	-0.020	-0.276	0.050	
% Explained variance	26.0%	17.5%	13.5%	09.0%	07.6%	04.9%	04.3%	
% Cumulative variance	26.0%	43.6%	57.1%	66.1%	73.7%	78%6	82.9%	

groupings of morphological descriptors that contributed differentially to accession differentiation within the germplasm bank. Three primary axes of variation emerged, each characterized by specific descriptor loadings that captured different aspects of morphological diversity.

PC1, with an eigenvalue of 5.987, was predominantly characterized by negative loadings from fruit-related descriptors. The five variables with the strongest negative contributions to PC1 were fruit width (-0.388), seed length (-0.373), pulp weight (-0.366), fruit weight (-0.357), and seed width (-0.344). Hydrogen potential (pH) exhibited the strongest positive association with PC1 (0.160), although this contribution was relatively modest. These loading patterns indicated that PC1 primarily captured variations in fruit size and dimensional attributes.

PC2, with an eigenvalue of 4.032, showed distinctly different descriptor associations than PC1. The most substantial negative loadings on PC2 were from floral descriptors, specifically petal

width (-0.389), petal length (-0.382), stamen length (-0.373), style length (-0.355), and hydrogen potential (-0.311). Conversely, vegetative descriptors, including leaf length (0.300), petiole length (0.302), and leaf width (0.298), provided the strongest positive contributions to this component. This pattern indicated that PC2 primarily differentiated accessions based on the contrast between reproductive and vegetative structures.

PC3, with an eigenvalue of 3.115, captured variation patterns that were not accounted for by the first two components. The strongest negative contributions to PC3 were from seed number (-0.506), total seed weight (-0.334), fruit weight (-0.244), and pulp weight (-0.205). In contrast, seed thickness (0.354), weight (0.257), and width (0.223) showed the highest positive loadings. This component appears to reflect an inverse relationship between seed count and individual seed dimensions within the fruits.

The remaining four components (PC4-PC7) each explained between 4.3% and 9.0% of variance, with more specific trait associations. Sepal descriptors dominated PC4, while PC5 was characterized by vegetative traits and sugar content. PC6 showed strong association with pedicel length, and PC7 was primarily influenced by fruit length. Together, these seven components provided a comprehensive representation of the morphological variation present in the germplasm collection.

Quantitative loading patterns across these principal components revealed that fruit mass-related measurements contributed more substantially to phenotypic differentiation than spatial dimensions. Fruit width exhibited the highest absolute loading value on any single component (-0.388 on PC1), followed closely by seed length (-0.373 on PC1), highlighting its prominent role in distinguishing between accessions in the germplasm collection.

### 3.8 Multivariate characterization of phenetic groups and descriptor relationships in principal component space

A three-dimensional biplot visualization (Figure 8; Supplementary Video S1) provided a comprehensive representation of the relationships between morphological descriptors and accession distributions across the four phenetic groups identified in the *M. dubia* germplasm bank. This multivariate projection, encompassing the first three principal components that collectively explained 57.1% of the total phenotypic variation, revealed clear patterns of descriptor associations and group separation in multidimensional space. The spatial distribution of the 43 accessions, color-coded by phenetic group assignment, demonstrated distinct clustering patterns that corresponded to specific descriptor combinations, as indicated by the direction and magnitude of descriptor vectors. Figure 8 reveals the three-dimensional relationships among accessions, with phenetic group 3 accessions (red points) occupying a distinct region in the negative PC1 space, pulled in that direction by fruit width and pulp weight vectors. Their separation from other groups along this commercially valuable axis, combined with intermediate positions on PC2 and PC3, indicates these accessions have optimized fruit traits without compromising plant vigor or seed quality.

Examination of the descriptor vector patterns within the 3D coordinate system revealed structured associations between morphologically and functionally related descriptors. Fruit-related descriptors (fruit width, fruit weight, and pulp weight) showed a strong parallel orientation, projecting primarily along the negative PC1 axis (-0.388, -0.357, -0.366) and forming a tight cluster of vectors. Seed dimension descriptors (seed length, width, and thickness) were similarly oriented along the negative PC1 axis but diverged along PC3, with seed thickness showing a stronger positive association with PC3 (0.354) compared to seed number, which projected strongly negatively along PC3 (-0.506). Floral descriptors exhibited a distinct pattern, clustering primarily along the negative PC2 axis, with petal dimensions and stamen length forming a coherent group separate from the fruit and seed descriptors.

Phenetic group 1 (yellow points) occupied a region characterized by positive values along PC1 and showed considerable dispersion along both the PC2 and PC3 axes. This group demonstrated an inverse relationship with fruit-related descriptors, displaying generally smaller fruits with a lower weight (8.85 g) and reduced pulp content (6.68 g). Accessions in this group, including PER-399, PER-410, and PER-413, exhibited positive associations with hydrogen potential (pH) and Brix degrees, suggesting higher sugar content and acidity levels compared to the other groups. These accessions showed intermediate to small seed dimensions, but maintained relatively normal floral characteristics.

Phenetic group 2 (green points) displayed a more centralized distribution in multivariate space, with moderate dispersion predominantly along PC3. These accessions, including PER-389, PER-402, and PER-419, exhibited intermediate values for most fruit-related descriptors, which were neither strongly positive nor negative along PC1. The defining descriptors of this group were notably smaller sepal dimensions (sepal width 0.23 cm, sepal length, 0.19 cm). Despite having intermediate fruit weights (9.49 g), the proportional relationship between pulp weight (7.28 g) and total fruit weight suggested higher pulp efficiency, with the pulp percentage exceeding that of groups with larger absolute fruit dimensions.

Phenetic group 3 (red points) occupied the negative sector of PC1, with considerable spread along both PC2 and PC3 axes. Accessions in this group, including PER-411, PER-417, and PER-424, demonstrated strong positive associations with the commercially desirable descriptors. These included wider fruits (2.97 cm), higher fruit weights (11.51 g), greater pulp content (8.72 g), and notably larger seed dimensions (seed length 1.60 cm, seed width 1.20 cm, seed thickness 0.37 cm). Interestingly, this group showed an inverse relationship with leaf length (7.95 cm), which was significantly shorter than that of the other groups.

Phenetic group 4 (purple points) occupied a region characterized by negative values along PC1 and PC2, with moderate dispersion along PC3. This group, containing accessions PER-391, PER-396, and PER-404, exhibited the largest leaf dimensions (leaf length 8.89 cm, leaf width, 3.77 cm) among all groups. Similar to Group 3, these accessions produced larger and heavier fruits (11.46 g) with substantial pulp weights (8.47 g) but were distinguished by having significantly more seeds per fruit and greater total seed weight (1.64 g) than the other groups. A notable characteristic of this cluster was its consistently lower pH value (2.80), representing the most acidic fruit in the collection.

Comparative analysis revealed distinctive trade-offs between commercially important descriptors across the four phenetic groups. Groups 3 and 4 contained the largest and heaviest fruits, respectively. Group 3 exhibited superior seed descriptors (larger and thicker individual seeds), whereas group 4 produced more seeds per fruit. Although groups 1 and 2 displayed smaller absolute fruit dimensions, group 2 demonstrated the highest pulp-to-fruit weight ratio, suggesting differential efficiency in resource allocation across the germplasm bank. These multidimensional relationships were further elucidated through an interactive three-dimensional

3D Biplot - PCA of *Myrciaria dubia*

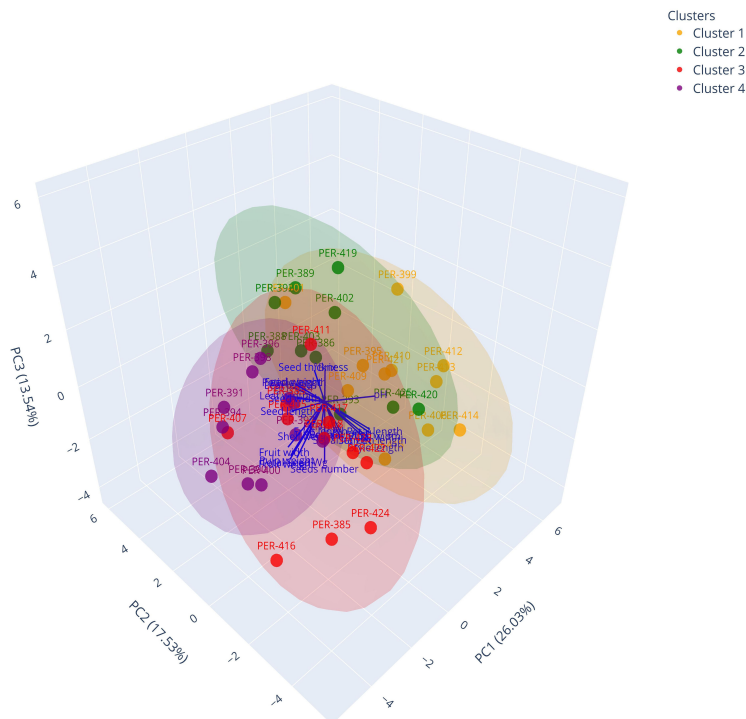


FIGURE 8

Three-dimensional PCA biplot of *M. dubia* accessions and morphological descriptors. Points represent individual accessions colored by phenetic group assignment. Arrows indicate descriptor loadings on the first three principal components. PC1, PC2, and PC3 axes show the percentage of variance explained.

representation provided in the [Supplementary Material \(Supplementary Video S1\)](#), allowing a more comprehensive visualization of phenotypic associations and accession distributions within the multivariate morphological space.

### 3.9 Statistical validation of phenetic groups differentiation

Multivariate Analysis of Variance (MANOVA) revealed highly significant differences among the four phenetic groups across all the measured quantitative variables (Table 3). All four multivariate test statistics demonstrated significant differentiation among the clusters ( $p < 0.0001$ ). The Wilks' Lambda test yielded a value of 0.01 ( $F = 2.9$ ,  $df = 69, 52$ ), the Pillai's trace was 2.34 ( $F = 2.91$ ,  $df = 69, 57$ ), the Lawley-Hotelling trace showed a value of 12.7 ( $F = 2.88$ ,  $df = 69, 47$ ), and Roy's largest root was 7.09 ( $F = 5.85$ ,  $df = 23, 19$ ). These consistently significant results across all four multivariate criteria confirmed robust differentiation among the phenetic groups.

*Post-hoc* analysis using Hotelling's  $T^2$  test identified three statistically distinct groups among the four phenetic clusters. Phenetic group 4 was classified independently as group A, whereas phenetic group 3 was assigned to group B, indicating significant multivariate differentiation between these two clusters. In contrast, phenetic groups 1 and 2 were both categorized as group C, demonstrating that these two clusters were not significantly different

from each other in the multivariate space despite being separated in the hierarchical cluster analysis. The clear statistical separation of phenetic groups 3 and 4 from each other and from groups 1 and 2 provided quantitative confirmation of the distinctiveness of the clusters identified through hierarchical classification.

### 3.10 Discriminating quantitative descriptors among phenetic groups

Analysis of Variance revealed significant differences among the four phenetic groups for 17 of the measured quantitative descriptors (Table 4). These differences were statistically significant, with ten descriptors showing highly significant differences ( $p < 0.001$ ), five showing very significant differences ( $p < 0.01$ ), and two showing significant differences ( $p < 0.05$ ). The Scott-Knott test distinguished specific groupings among the four phenetic groups for each descriptor.

For vegetative and flower descriptors, distinct patterns emerged across the phenetic groups. pH values showed groups 1, 2, and 3 forming a statistically homogeneous group (2.87-2.91) that differed significantly from group 4 (2.80). Leaf length measurements differentiated group 3 (7.95 cm) from the other three groups, which formed a statistically uniform group (8.47-8.89 cm). Leaf width separated groups 1 and 3 (3.43-3.46 cm) from groups 2 and 4 (3.63-3.77 cm). Phenetic groups 1 and 3 consistently formed a

TABLE 3 Multivariate Analysis of Variance (MANOVA) and Hotelling's  $T^2$  *post-hoc* test results for phenetic groups of *M. dubia* germplasm bank accessions.

Variation source	Statistical value	F-value	df (num)	df (den)	p-value	Method	Phenetic group	Hotelling's Test
Clusters	0.01	2.9	69	52	<0.0001	Wilks	4	A
	2.34	2.91	69	57	<0.0001	Pillai	3	B
	12.7	2.88	69	47	0.0001	Lawley-Hotelling	2	C
	7.09	5.85	23	19	0.0001	Roy	1	C

single statistical group with higher sepal width, sepal length, petal width, petal length, and stamen length, whereas phenetic groups 2 and 4 generally displayed lower values.

The fruit-related descriptors demonstrated particularly pronounced differences between phenetic groups. Fruit width measurements separated the clusters into three distinct groups: groups 3 and 4 with the highest values (2.88–2.97 cm), group 2 had intermediate values (2.46 cm), and group 1 had the lowest values (2.25 cm). Fruit weight followed a similar pattern, with groups 3 and 4 forming one statistical group with higher weights (11.46–11.51 g) that differed significantly from groups 1 and 2 (8.85–9.49 g). Shell weight, pulp weight, and total seed weight showed similar grouping patterns, with groups 3 and 4 consistently forming a statistically homogeneous group with higher values than those of groups 1 and 2.

The seed descriptors displayed more complex differentiation patterns across phenetic groups. Seed length showed the highest degree of discrimination, with each cluster forming a statistically distinct group, except for groups 1 and 2, which were grouped together. Seed width separated the clusters into three groups: group 3 with the highest width (1.20 cm), groups 2 and 4 with intermediate values (1.07–1.10 cm), and group 1 with the lowest values (0.95 cm). Seed thickness differentiated group 3 (0.37 cm) from the other three groups (0.32–0.34 cm), while seed weight separated groups 3 and 4 (0.57–0.61 g) from group 1 and 2 (0.50–0.54 g).

### 3.11 Geographic distribution of phenetic groups

The spatial distribution of the four phenetic groups across the Loreto region revealed a heterogeneous pattern, with no clear geographic structure (Figure 9). Phenetic group 1 (yellow) was predominantly found in areas near Iquitos, with several accessions clustered around the central part of the region. Phenetic group 2 (green) showed a dispersed distribution, with representatives found in the northwest, northeast, and central sections of the study area. Phenetic group 3 (red), which contained accessions with superior fruit characteristics, displayed a widespread distribution across the region, with representatives found along multiple river systems, including areas north of Iquitos and in central river networks. Phenetic group 4 (purple) was concentrated in two main areas: the

central zone around Iquitos and the southern portions of the Loreto department.

Statistical analysis using the Mantel test revealed no significant correlation between phenotypic variability and geographical distance ( $r = 0.0639$ ,  $p = 0.4034$ ), as shown in Figure 9 and Supplementary Figure S5, confirming our hypothesis that local environmental factors, rather than regional adaptations, drive the phenotypic patterns. The low correlation coefficient and non-significant p-value demonstrated that accessions with similar phenotypic characteristics were not necessarily geographically proximate. The random distribution pattern was consistent across all eight hydrographic basins in the germplasm collection, with representatives of multiple phenetic groups often found within the same river system. This random spatial arrangement was particularly evident in the central region, where accessions belonging to all four phenetic groups occurred in relatively close proximity, despite their phenotypic differences.

The distribution map also revealed that phenetic group membership transcended collection site boundaries, with accessions from the same phenetic group often originating from distant, disconnected water systems. This pattern was evident for all four phenetic groups, with no exclusive association between any single phenetic group and a particular geographic region or hydrographic basin within the Loreto Department. The most phenotypically distinct groups (3 and 4) showed overlapping geographical distributions with intermediate groups (1 and 2), further emphasizing the lack of geographic structuring in the observed phenotypic variation.

### 3.12 Identification of promising germplasm accessions

Statistical evaluation of commercially important fruit descriptors across the 12 selected accessions confirmed our third hypothesis that distinct phenotypic groups would harbor promising germplasm with superior trait combinations (Figure 10). The comprehensive analysis encompassed six key parameters: fruit weight, pulp weight, total seed weight, pulp percentage, seed number per accession, and seed count in high pulp accessions (>75% pulp). While parametric tests revealed no significant differences among accessions for fruit mass components ( $p > 0.05$ ), non-parametric analyses identified

TABLE 4 Analysis of Variance (ANOVA) and Scott-Knott test results for quantitative descriptors across four phenetic groups of the *M. dubia ex situ* germplasm bank.

Phenetic group	pH (***)	Leaf length (**)	Leaf width (**)	Sepal width (***)	Sepal length (***)	Petal width (*)	Petal length (*)	Stamen length (**)	
1	2.91 <sup>a</sup>	8.60 <sup>a</sup>	3.46 <sup>b</sup>	0.27 <sup>a</sup>	0.22 <sup>a</sup>	0.38 <sup>a</sup>	0.43 <sup>a</sup>	1.00 <sup>a</sup>	
2	2.88 <sup>a</sup>	8.47 <sup>a</sup>	3.63 <sup>a</sup>	0.23 <sup>b</sup>	0.19 <sup>b</sup>	0.37 <sup>b</sup>	0.41 <sup>b</sup>	0.89 <sup>b</sup>	
3	2.87 <sup>a</sup>	7.95 <sup>b</sup>	3.43 <sup>b</sup>	0.27 <sup>a</sup>	0.22 <sup>a</sup>	0.38 <sup>a</sup>	0.43 <sup>a</sup>	1.04 <sup>a</sup>	
4	2.80 <sup>b</sup>	8.89 <sup>a</sup>	3.77 <sup>a</sup>	0.26 <sup>a</sup>	0.21 <sup>a</sup>	0.35 <sup>b</sup>	0.40 <sup>b</sup>	0.88 <sup>b</sup>	
Phenetic group	Fruit width (***)	Fruit weight (***)	Shell weight (**)	Pulp weight (***)	Total seeds weight (***)	Seed length (***)	Seed width (***)	Seed thickness (***)	Seeds weight (**)
1	2.25 <sup>c</sup>	08.85 <sup>b</sup>	1.01 <sup>b</sup>	6.68 <sup>c</sup>	1.36 <sup>b</sup>	1.18 <sup>d</sup>	0.95 <sup>c</sup>	0.32 <sup>b</sup>	0.50 <sup>b</sup>
2	2.46 <sup>b</sup>	09.49 <sup>b</sup>	1.05 <sup>b</sup>	7.28 <sup>b</sup>	1.36 <sup>b</sup>	1.34 <sup>c</sup>	1.07 <sup>b</sup>	0.34 <sup>b</sup>	0.54 <sup>b</sup>
3	2.97 <sup>a</sup>	11.51 <sup>a</sup>	1.13 <sup>a</sup>	8.72 <sup>a</sup>	1.55 <sup>a</sup>	1.60 <sup>a</sup>	1.20 <sup>a</sup>	0.37 <sup>a</sup>	0.61 <sup>a</sup>
4	2.88 <sup>a</sup>	11.46 <sup>a</sup>	1.08 <sup>a</sup>	8.47 <sup>a</sup>	1.64 <sup>a</sup>	1.47 <sup>b</sup>	1.10 <sup>b</sup>	0.32 <sup>b</sup>	0.57 <sup>a</sup>

p-value: <0.05 (\*), <0.01 (\*\*), <0.001 (\*\*\*) . Different superscript letters within the same variable indicate statistically significant differences.

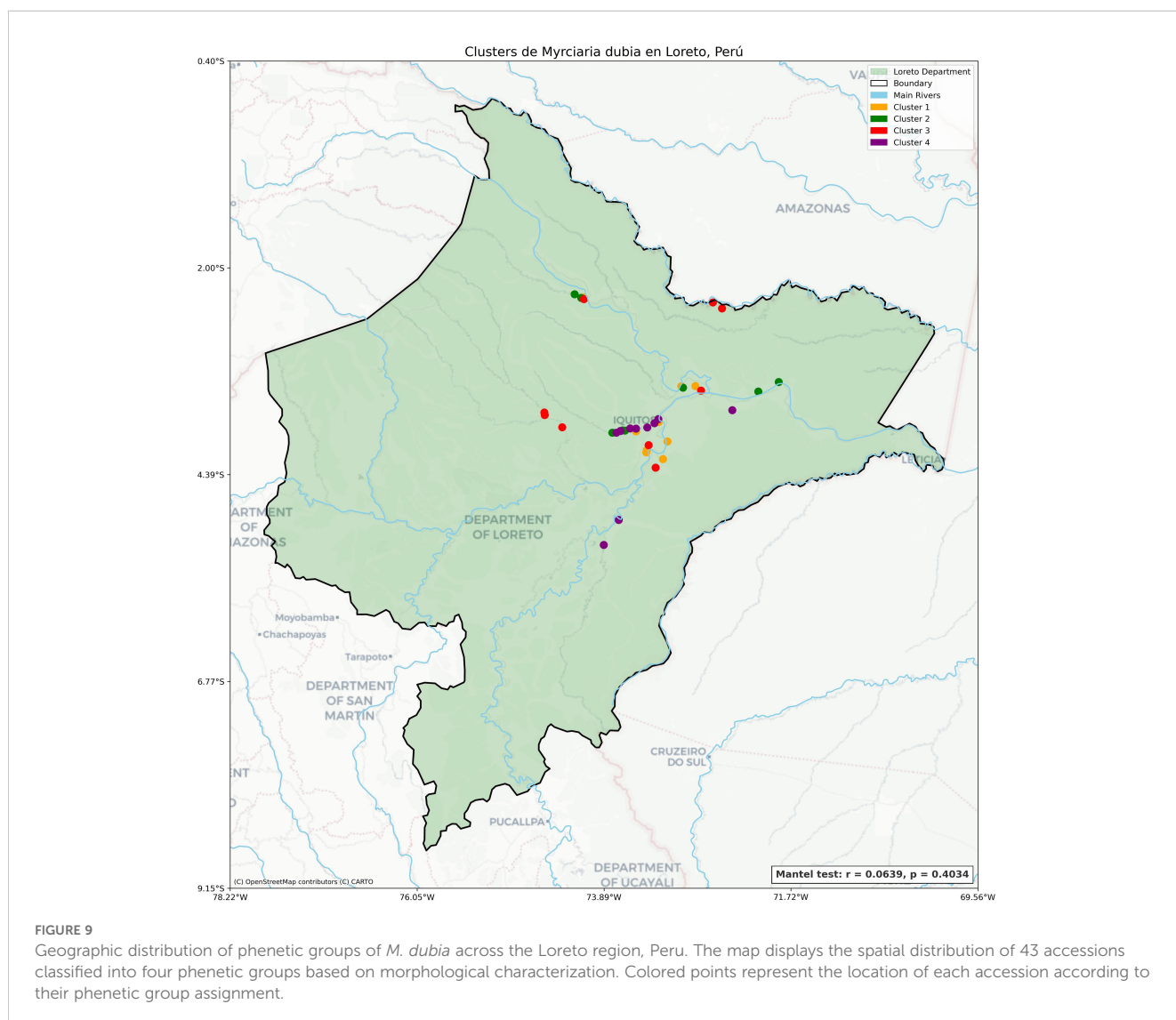
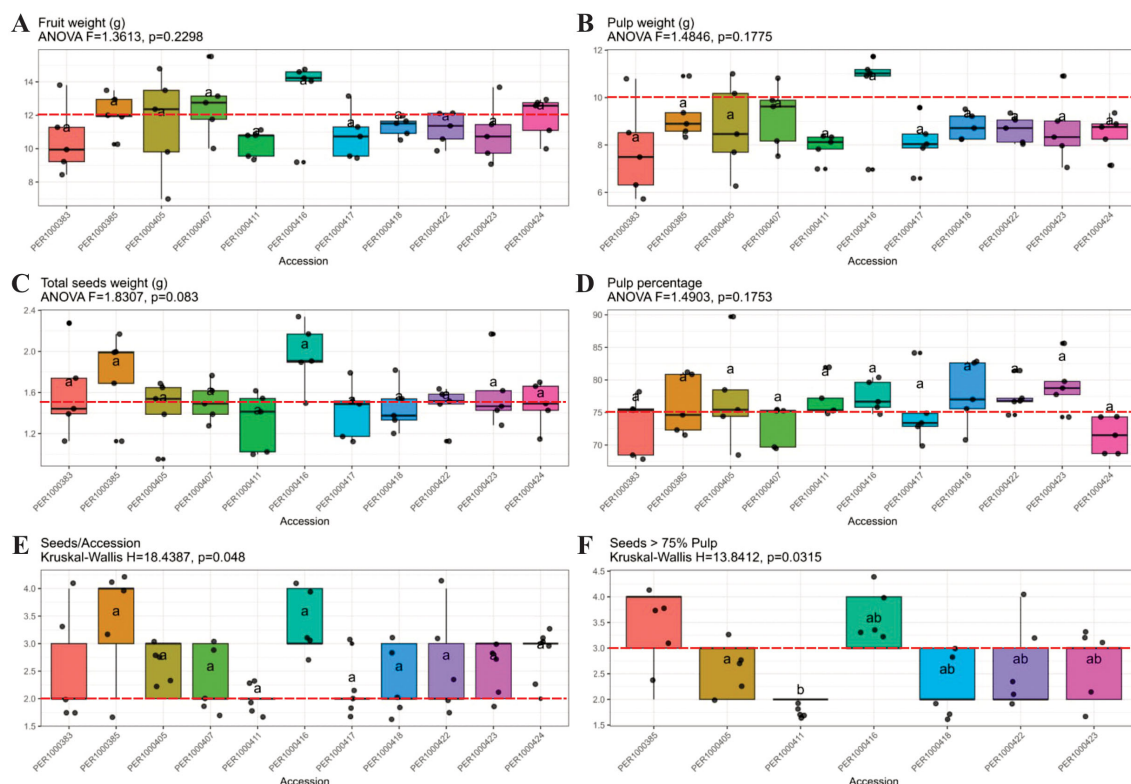


FIGURE 9 Geographic distribution of phenetic groups of *M. dubia* across the Loreto region, Peru. The map displays the spatial distribution of 43 accessions classified into four phenetic groups based on morphological characterization. Colored points represent the location of each accession according to their phenetic group assignment.



**FIGURE 10** Box plot analysis of six fruit quality descriptors for eleven *M. dubia* accessions selected from the germplasm bank. Panels show fruit weight (A), pulp weight (B), total seed weight (C), pulp percentage (D), seed number per accession (E), and seed number in fruits with >75% pulp content (F). Red dashed lines indicate commercial quality thresholds. Statistical test results appear above each panel (ANOVA for panels A–D; Kruskal–Wallis for panels E, F). Letters above boxes denote homogeneous groups at  $p < 0.05$ , where identical letters indicate no significant difference between accessions.

significant variation in seed-related traits with direct implications for commercial selection.

Fruit weight distribution showed considerable uniformity across evaluated accessions, with no significant differences detected (ANOVA  $F=1.3613$ ,  $p=0.2298$ ; Figure 10A). Values ranged from 8.5 to 14.5 g, with median weights clustering between 10–13 g. The red dashed line at 11 g represents the overall mean, revealing that accessions PER1000411 and PER1000416 displayed the highest median fruit weights (approximately 13 g), while PER1000383 showed the lowest (approximately 10 g). This statistical homogeneity suggests fruit weight alone is insufficient as a primary selection criterion.

Pulp weight patterns mirrored total fruit weight, showing no significant differences among accessions (ANOVA  $F=1.4846$ ,  $p=0.1775$ ; Figure 10B). Values ranged from 6 to 11 g, with median values between 7.5–10.5 g. Accessions PER1000411 and PER1000416 maintained superior performance with the highest median pulp weights (approximately 10 g), while PER1000383 exhibited the lowest. The strong correlation between fruit and pulp weight ( $\rho=0.95$ ) was evident in these parallel patterns.

Total seed weight showed marginally non-significant variation (ANOVA  $F=1.8307$ ,  $p=0.083$ ; Figure 10C), ranging from 1.0 to 2.3 g. Despite the lack of statistical significance, PER1000416 displayed notably higher total seed weight (median >2.0 g), while

PER1000407 and PER1000411 showed the lowest values (median <1.4 g), suggesting differential seed production strategies among accessions.

Pulp percentage, a critical commercial parameter, showed no significant differences (ANOVA  $F=1.4903$ ,  $p=0.1753$ ; Figure 10D), with values ranging from 68% to 80%. Accessions PER1000385, PER1000417, and PER1000418 displayed the highest median pulp percentages (78–80%), while PER1000424 showed the lowest (approximately 72%). The red dashed line at 75% indicates the threshold for high-quality fruit, with most accessions exceeding this benchmark.

Seed number per accession revealed the first significant differentiation (Kruskal–Wallis  $H=18.4387$ ,  $p=0.048$ ; Figure 10E). *Post-hoc* analysis clearly separated accessions into three groups: PER1000411 and PER1000407 with the lowest seed counts (group ‘a’, median 2.0–2.2 seeds), intermediate accessions (group ‘ab’), and PER1000385 and PER1000405 with the highest seed numbers (group ‘b’, median 3.5–4.0 seeds).

The most discriminating analysis emerged when examining seed number in high-pulp accessions (>75% pulp content). This subset analysis revealed highly significant differences (Kruskal–Wallis  $H=13.8412$ ,  $p=0.0315$ ; Figure 10F), effectively separating accessions into distinct commercial categories. PER1000411 demonstrated the optimal combination (group ‘ab’) with

consistently low seed numbers (2.0-2.2) while maintaining pulp percentage above 75%. PER1000385 showed the highest seed count (group 'a', 4.0 seeds) despite high pulp percentage, while PER1000416 occupied an intermediate position (group 'ab').

These results confirmed our hypothesis and identified three superior accessions from phenetic group 3 with complementary trait profiles. Accession PER1000411 emerged as the most promising overall, combining high fruit weight, substantial pulp weight, minimal seed number, and a superior pulp percentage. Accession PER1000416 excelled in absolute fruit and pulp weight but carried a moderate seed load. Accession PER1000417 balanced the moderate fruit size with an excellent pulp percentage and acceptable seed count. The concentration of these elite materials within phenetic group 3, despite the limited statistical differentiation in individual traits, validates the multivariate clustering approach for identifying synergistic trait combinations that would be overlooked by univariate selection alone.

## 4 Discussion

### 4.1 Patterns of phenotypic variability and evolutionary context

Phenotypic characterization of the *M. dubia ex situ* germplasm bank reveals distinct variability patterns that provide significant insights into the evolutionary history and domestication status of the species. The differential pattern between quantitative and qualitative descriptors characteristic of species with limited domestication history emerges clearly from our analysis (Table 1; Figures 2, 4). This distinction proves particularly informative because the restricted diversity of qualitative descriptors prevented clear morphological differentiation between accessions. In contrast, quantitative descriptors, especially those related to fruit mass, facilitated the identification of distinct phenetic groups. The absence of a correlation between these groups and geographic origin suggests that local environmental conditions, rather than regional adaptations, are the primary drivers shaping the observed phenotypic diversity patterns.

Understanding these variability patterns requires examination of the specific environmental pressures that have shaped trait evolution in *M. dubia*. The higher variability in reproductive descriptors compared to vegetative descriptors likely reflects specific environmental pressures in riparian habitats (Haghpanah et al., 2024; Seleiman et al., 2021). During the annual flooding cycle (December-May), *M. dubia* trees remain partially submerged for up to 4–5 months, creating selection for larger, more buoyant fruits that enhance water-mediated dispersal. The negative correlation between seed number and size suggests resource allocation trade-offs optimized for variable flooding intensities, fewer large seeds for long-distance dispersal versus numerous small seeds for local colonization. Additionally, the remarkably low pH variability indicates a strong stabilizing selection for fruit acidity, potentially maintaining vitamin C stability under humid conditions (Peters and Vasquez, 1987; Ferreira et al., 2010).

While our findings reveal novel aspects of *M. dubia* diversity, they also complement and extend previous research on this species. These findings align with previous research while offering a more comprehensive assessment of the variability patterns of *M. dubia* in relation to its evolutionary trajectory. Our study extends beyond the single-trait analyses of Imán et al. (2011) by providing a multivariate characterization of 23 traits. Our observations corroborate the findings of Bardales-Lozano et al. (2016), who reported higher variability in fruit morphological descriptors than vegetative descriptors in Brazilian populations, but our comprehensive dataset revealed that this pattern is consistent across the species range. Chagas et al. (2015) identified substantial variations in fruit descriptors across wild populations in the northern Amazon Basin, with fruit mass and pulp content displaying particularly high coefficients of variation. Our analysis uniquely demonstrated that this variation clusters into distinct phenetic groups rather than continuous gradients. This consistency across independent studies conducted in different regions of the Amazon Basin strengthens the conclusion that fruit-related variability represents an intrinsic characteristic of *M. dubia* populations, shaped by natural selection rather than human intervention.

The patterns observed in *M. dubia* become particularly meaningful when contrasted with variability patterns in domesticated crops. The variability pattern observed in *M. dubia* differs markedly from that typically documented in highly domesticated crops, providing evidence for its limited historical domestication. In contrast to *M. dubia*, intensively selected crops display pronounced morphological variation across multiple trait categories, resulting from human-mediated selection (Pickersgill et al., 1976; Meyer and Purugganan, 2013; Pickersgill, 2016). *M. dubia* exhibits a variability pattern that more closely resembles that of semi-domesticated or recently domesticated species (Clement et al., 2010). This limited domestication status has important implications for breeding: the high natural variability in fruit traits provides excellent selection opportunities without the genetic bottlenecks typical of domesticated crops, whereas the absence of domestication syndrome allows for rapid genetic gains through simple mass selection.

This limited domestication status creates specific opportunities for targeted improvement efforts. The absence of classic domestication syndrome features, such as the loss of seed dormancy, uniform fruit size, or reduced chemical defenses, indicates specific opportunities for improvement. Future domestication efforts should aim to (1) extend the harvest season through selection for asynchronous ripening, (2) develop thornless varieties for easier harvesting, (3) reduce seed number while maintaining fruit size, and (4) enhance post-harvest fruit stability. Understanding these targets allows breeders to accelerate domestication while preserving valuable wild characteristics, such as flood adaptation and high vitamin C content (Stetter, 2020; Iqbal et al., 2020).

Molecular genetic evidence remarkably aligns with our phenotypic observations. Recent genomic investigations using microsatellite markers have revealed higher diversity within

populations than between populations, indicating that *M. dubia* maintains considerable heterozygosity while exhibiting limited differentiation between geographically separated populations (Castro et al., 2024). This pattern aligns remarkably well with our phenotypic findings of low population-level differentiation, combined with moderate individual variability. Similar conclusions were reached by Šmíd et al. (2017), whose microsatellite analysis confirmed this genetic structure, and by Nunes et al. (2017), who used ISSR markers to demonstrate that the *M. dubia* population structure exhibits high within-population variation (73.8%) and comparatively low between-population variation (26.2%).

Such a genetic organization is characteristic of outcrossing species that have experienced minimal artificial selection pressure, which is a key indicator of early stage domestication. Genetic evidence presents a cohesive picture of a species maintaining high heterozygosity that contributes to its adaptive potential and phenotypic plasticity, while notably lacking the selective sweeps and reduced diversity commonly observed in genomic regions of crops subjected to intense artificial selection. This congruence between the genetic and phenotypic data strongly reinforces our hypothesis that *M. dubia* remains in the preliminary stages of domestication, with its current genetic and phenotypic characteristics predominantly shaped by natural evolutionary processes rather than sustained human-mediated selection.

However, some apparent discrepancies merit further explanation. Although neutral molecular markers show limited population differentiation (Šmíd et al., 2017), our phenotypic data revealed four distinct clusters. This reflects different evolutionary forces: neutral markers capture demographic history and gene flow, whereas phenotypic traits respond to local selection pressures. The phenetic groups we identified likely represent adaptive peaks for different microhabitat conditions within the relatively uniform genetic background maintained by gene flow.

These insights into population structure and genetic organization have direct implications for conservation and breeding strategies. The moderate genetic uniformity observed necessitates strategic conservation approaches: (1) prioritizing accessions from different phenetic groups to maintain allelic diversity, (2) implementing complementary *in situ* conservation in diverse microhabitats to capture ongoing adaptive evolution, and (3) utilizing the retained heterozygosity for rapid genetic gains through recurrent selection without the constraints of narrow genetic bases typical of highly domesticated crops. This conservation strategy should incorporate both static preservation in germplasm banks and dynamic conservation through on-farm management, allowing continued evolution in response to changing environmental conditions.

## 4.2 Significance of phenetic groups and association with commercial descriptors

Multivariate analysis provided valuable insights into the structure of the phenotypic diversity within the germplasm bank (Figures 7, 8; Supplementary Video S1). The identification of four

phenetic groups primarily differentiated by fruit-related descriptors has important implications for breeding strategies, as it indicates that selection of fruit descriptors can effectively target the primary axes of variation within the species.

Building on previous phenetic studies in *M. dubia*, our analysis reveals both consistencies and novel insights. Our results extend those of Chagas et al. (2015), who identified five groups when analyzing fruit descriptors in wild *M. dubia* populations in northern Brazil. While previous studies have identified fruit descriptors as the most discriminating variables in Peruvian cultivated populations (Pinedo, 2017; Šmíd et al., 2017), our comprehensive analysis uniquely demonstrates that these traits form coherent clusters suitable for targeted breeding. The consistency across these independent studies suggests that fruit descriptors are evolutionarily and agronomically significant in *M. dubia*, likely reflecting both natural selection for reproductive success, and incipient human selection for improved fruit quality.

The geographic distribution of these phenetic groups reveals important patterns about the forces shaping diversity in this species. The lack of correlation between phenetic groups and geographic origins (Supplementary Figure S3), indicates that phenotypic variation was not structured by regional adaptation. This finding differs from the conclusions of Bardales-Lozano et al. (2016), who documented a significant relationship between fruit descriptors and geographic distribution in Roraima, Brazil. In contrast, the present study aligns with broader research highlighting the complex spatial distribution patterns of Amazonian plant diversity influenced by multiple factors, including river dynamics and historical biogeography (de Oliveira et al., 2014; Maestri and Duarte, 2020; Thom et al., 2020). These findings suggest that ecological and biogeographical interactions in the Amazon are convoluted and are often shaped by habitat heterogeneity, which plays a crucial role in species richness and distribution (Palmer et al., 2010).

These complex ecological interactions are increasingly understood through advances in landscape genetics. For example, recent studies using these approaches have significantly enhanced our understanding of the complex interplay among genetic variation, environmental conditions, and phenotypic descriptors in Amazonian tree species. Luize et al. (2024) and Lemke et al. (2012) convincingly demonstrated that local microhabitat conditions, particularly soil characteristics, nutrient availability, light exposure, and flooding regimes, exert stronger selective pressures on riparian species than broader regional factors. These findings provide substantial evidence that fine-scale environmental heterogeneity, rather than geographic distance or regional adaptations, primarily drives the observed patterns of phenotypic diversity in these ecosystems.

The mechanistic basis for these environmental adaptations in Amazonian trees is becoming clearer through molecular approaches. Landscape genomics has proven to be instrumental in identifying specific genetic loci associated with adaptive descriptors in diverse tree populations (Čalić et al., 2015). A compelling example comes from research on *Himatanthus succuba*, where populations in flooded versus non-flooded environments display significant genetic and phenotypic

differentiation, indicating that hydrological regimes function as powerful selective forces that drive local adaptation (Ferreira et al., 2010). These studies revealed that genetic variation underlying phenotypic descriptors is often structured in response to fine-scale environmental gradients rather than broader geographic patterns.

The implications of these findings extend beyond academic understanding to practical applications in conservation and germplasm collection. The absence of geographic structuring in phenotypic variation has profound implications. It suggests that microenvironmental heterogeneity within sites exceeds macrogeographic differentiation, supporting a conservation strategy focused on ecological gradients rather than geographic representation. This pattern, consistent with high gene flow in river-connected populations, indicates that future climate adaptation may depend more on standing variation within populations than on region-specific ecotypes. For germplasm collection strategies, this finding suggests that sampling multiple microhabitats within a single location may capture more functional diversity than extensive geographic sampling across uniform habitats.

### 4.3 Promising accessions for breeding programs and future applications

Having established that phenetic groups represent coherent clusters of variation driven by environmental factors, the practical application of these findings becomes paramount. The identification of elite materials becomes particularly relevant given that phenetic group 3 consistently demonstrated superior commercial traits across multiple analyses. From an applied perspective, translating our understanding of diversity patterns into actionable breeding targets represents a critical step in the domestication continuum of *M. dubia*.

The superiority of phenetic group 3 accessions extends previous findings while providing new insights for breeding strategies. The identification of elite germplasm extends beyond previous single-location evaluations by providing multi-trait selection indices. For example, Alves et al. (2013) demonstrated that phenotypic selection based on fruit quality descriptors in tropical fruits leads to significant genetic gains in subsequent breeding generations, suggesting similar potential for *M. dubia*. Similarly, Bardales-Lozano et al. (2016) reported the successful development of improved *M. dubia* varieties through selection based on phenotypic characterization of germplasm collections.

Beyond identifying superior accessions, understanding their complementary trait profiles enables strategic breeding approaches. The distinct descriptor combinations observed in the identified promising accessions suggest specific breeding strategies tailored to different market demands and processing requirements. Crossing these complementary accessions could pyramid favorable alleles while maintaining the genetic diversity essential for long-term improvement. This strategy aligns with modern approaches to crop improvement that balance genetic gain with genetic

conservation, as advocated by Tanksley and McCouch (1997) and Liu et al. (2023) for underutilized species.

### 4.4 Implications for sustainable development

The identification of promising germplasm accessions with superior trait combinations extends beyond immediate breeding applications to encompass broader socioeconomic considerations. The foundation provided by these elite materials enables development of improved varieties that could substantially increase the productivity and profitability of *M. dubia* cultivation. Penn (2006) demonstrated that improved *M. dubia* varieties could generate 2–3 times the income of traditional crops in the region while creating sustainable livelihoods that reduce pressure on forest resources. Furthermore, the commercial development of *M. dubia* can create value chains that benefit rural communities through processing and value addition (Pinedo Panduro et al., 2004; Pinedo, 2009).

The economic foundation for this transformation is already evident in current production systems. The economic potential of *M. dubia* in the Peruvian Amazon is substantial, with a current production of 14,226 tons, generating approximately \$28 million annually (MIDAGRI, 2025), demonstrating significant promise for improving rural livelihoods. Blare and Donovan (2018) demonstrated that the cultivation and value addition of high-value native fruits increased rural household income while simultaneously requiring less land and contributing to the conservation of forest cover. By adopting improved varieties derived from our identified accessions, smallholder farmers can significantly enhance yield and quality, thereby positioning themselves to meet the rising international demand for this superfruit (Fracassetti et al., 2013; García-Chacón et al., 2023).

Beyond primary production, the development of value chains creates multiplier effects throughout rural economies. Establishing sustainable *M. dubia* production systems generates employment throughout the value chain from cultivation to processing and marketing. This addresses poverty reduction and rural development goals, particularly for indigenous communities that maintain traditional knowledge of *M. dubia* management (Blare and Donovan, 2018). Value-added processing of *M. dubia* fruit into products such as freeze-dried powder, juices, and functional foods could multiply economic returns in rural communities five-fold compared to fresh fruit sales (Conceição et al., 2019; Do et al., 2021). Additionally, byproducts, such as seed oil for cosmetics, reduce waste and provide additional income streams (Azevedo et al., 2019; Conceição et al., 2019).

Ultimately, the conservation and utilization of characterized germplasm represents a model for sustainable development in the Amazon. The *ex situ* conservation approach exemplified by the INIA germplasm bank complements *in situ* conservation efforts by preserving genetic diversity, while enabling characterization and utilization. This dual approach operationalizes sustainable

development principles by balancing immediate economic needs with long-term conservation goals. As climate change threatens Amazonian ecosystems, maintaining both *ex situ* collections and developing climate-resilient production systems based on characterized germplasm are critical (Clement et al., 2010).

## 4.5 Study limitations

While our study provides valuable insights into the phenotypic diversity of *M. dubia*, several limitations should be acknowledged. First, reliance on morphological data alone provides an incomplete picture of the genetic diversity. Phenotypic descriptors are influenced by both genetic and environmental factors, and the observed variability may not fully reflect the underlying genetic diversity. Future studies should integrate molecular marker data to comprehensively assess genetic diversity patterns and population structures (Govindaraj et al., 2015).

Second, the *ex situ* germplasm bank, which is representative of the Loreto region, encompasses a relatively limited geographic area compared with the total distribution range of the species. *M. dubia* populations from other Amazonian countries, including Brazil, Colombia, and Venezuela, may harbor additional diversity not captured in the current study (Šmíd et al., 2017; Castro et al., 2018). Investigations have reported significant genetic differentiation between populations of Amazonian species from different watersheds, suggesting that broader geographic sampling may reveal additional patterns of diversity (Clement et al., 2010).

Third, our study focused primarily on morphological descriptors, with a limited assessment of biochemical characteristics beyond basic measurements of sugar content and pH. Recent studies have demonstrated significant variations in the vitamin C content and antioxidant capacity among *M. dubia* genotypes (Castro Gómez et al., 2013; Freitas et al., 2016; Grigio et al., 2016), highlighting the importance of such evaluations. Given the commercial importance of *M. dubia* fruit as a source of vitamin C and other bioactive compounds, future phytochemical profiling studies are essential.

Finally, phenotypic characterization was conducted in a single environment, limiting the assessment of genotype-by-environment interactions and phenotypic plasticity. Such interactions are crucial for understanding how different genotypes respond to varying environmental conditions, which can significantly affect descriptor expression and stability (Montesinos-López et al., 2016; Nguyen et al., 2023). Conducting multi-environment trials would provide more comprehensive insights into the stability of observed descriptors across diverse growing conditions (i.e., flooding gradients, soil types, and climate zones), facilitating the identification of resilient genotypes that can adapt to varying agricultural environments (Sakazaki et al., 2022).

## 4.6 Future research directions

The advancement of *M. dubia* improvement requires immediate deployment of genomic technologies to accelerate

breeding and conservation efforts. Whole-genome sequencing, successfully implemented in other tropical fruit species (Alves et al., 2024; Chakrabarty and Külheim, 2025), will establish the foundation for developing SNP markers and enabling marker-assisted selection for commercially important traits. Genotyping-by-sequencing approaches, proven effective for determining genetic diversity and population structure in Myrtaceae species (Klápště et al., 2021; Chen et al., 2025; Zolkafli et al., 2025; Leal et al., 2025). Most transformatively, CRISPR/Cas-based genome editing technologies targeting the L-galactose pathway for vitamin C biosynthesis (Castro et al., 2023; Mall et al., 2024; Redondo-López et al., 2025), could dramatically accelerate genetic gains. These molecular tools will support development of mapping populations from crosses between phenetic groups, particularly utilizing superior accessions from group 3, enabling quantitative trait loci identification and implementation of genomic selection protocols that could reduce breeding cycles from 8–10 years to 4–5 years (Kumar et al., 2020). Integration of speed breeding techniques using controlled environments (Watson et al., 2018) and multi-omics approaches combining genomics, transcriptomics, and metabolomics (Kumar et al., 2017) will further accelerate understanding of the genetic basis underlying the exceptional nutritional properties of *M. dubia*.

To fully exploit the genetic potential revealed by genomic approaches, integrative multi-omics strategies must elucidate the complex biological pathways underlying the exceptional nutritional properties of *M. dubia*. Integration of multi-omic approaches would significantly deepen our understanding of the genetic and biochemical bases underlying the exceptional properties of *M. dubia*. Integration of genomics with transcriptomics, proteomics, and metabolomics will reveal how genetic variation translates into the biochemical diversity of bioactive compounds. High-throughput metabolomic profiling across the germplasm collection will identify relationships between genetic markers and phytochemical composition, following successful models in other species (Pott et al., 2020; Saldanha et al., 2020; Colantonio et al., 2022; Manickavasagam et al., 2024). RNA-Seq analyses across developmental stages can identify key regulatory networks controlling fruit development and quality descriptors, building on approaches successfully employed with *Benincasa hispida* (Du et al., 2024), *Citrus sinensis* (Feng et al., 2019), and *Fragaria chiloensis* (Gaete-Eastman et al., 2022). These molecular insights will guide targeted breeding for enhanced nutritional value and optimized agronomic performance (Ceccarelli, 2015).

Additionally, conservation strategies must evolve beyond traditional *ex situ* maintenance to encompass dynamic approaches that preserve adaptive potential while supporting phenotypic evaluation. Establishment of multilocational collections across environmental gradients will capture genotype-by-environment interactions, complemented by cryopreservation protocols for long-term security (Panis et al., 2020) and on-farm conservation networks that maintain traditional varieties while preserving indigenous knowledge (Galluzzi et al., 2010). Supporting these efforts, next-generation phenotyping technologies (Tao et al., 2022; Awada et al., 2024) including

UAV-based remote sensing (Ampatzidis et al., 2020) and near-infrared spectroscopy for non-destructive fruit quality assessment (Walsh et al., 2020) will enable rapid evaluation of large germplasm collections. Multi-environment phenotyping networks across the Amazon basin, focusing on climate resilience traits such as drought and flooding tolerance (Oliveira et al., 2019), will identify germplasm capable of maintaining productivity under future climate scenarios while incorporating farmer preferences through participatory selection approaches (Ceccarelli and Grando, 2020).

The ultimate success of *M. dubia* as a driver of sustainable Amazonian development requires integration of biological research with production system optimization and socioeconomic considerations. Development of agroforestry systems that integrate *M. dubia* with complementary species (Miller and Nair, 2006), combined with research on pollinator ecology (Campbell et al., 2018) and post-harvest technologies (Neves et al., 2015), will address current production constraints. Innovation platforms uniting researchers, farmers, processors, and policymakers will ensure research relevance and facilitate technology adoption (Schut et al., 2016), while decision support tools incorporating genetic, environmental, and market information will guide variety deployment. Long-term studies quantifying ecosystem services from *M. dubia* cultivation, including carbon sequestration and biodiversity conservation (Reed et al., 2019), will demonstrate broader societal benefits beyond direct economic returns. These integrated research efforts will establish *M. dubia* as a model for sustainable utilization of underutilized Amazonian species, creating a replicable framework that simultaneously advances conservation and development objectives throughout tropical regions globally.

## 5 Conclusions

The comprehensive phenotypic characterization of the *M. dubia ex situ* germplasm bank confirmed our three primary hypotheses. First, we observed moderate overall phenotypic variability, with greater variation in fruit and seed descriptors than in vegetative descriptors (Table 1; Figure 2), reflecting the limited history of domestication of this Amazonian species. This variability pattern has critical implications: the high variation in commercially important traits provides excellent opportunities for genetic improvement through selection, whereas the stability of vegetative traits ensures that fruit quality improvements do not compromise plant adaptation or agronomic performance. Second, as hypothesized, we found no significant correlation between phenotypic characteristics and geographic origin, confirming that the observed variability is influenced more by local environmental factors than by regional adaptations. This finding fundamentally redirects conservation and collection strategies—future germplasm acquisition should prioritize diverse microhabitats and environmental gradients within regions rather than maximizing geographic coverage, potentially reducing collection costs while capturing greater functional diversity. Third, multivariate analysis successfully identified four distinct phenetic groups, with three promising accessions (PER1000416, PER1000423, and

PER1000411) from phenetic group 3 that exhibited superior combinations of commercially important descriptors.

The moderate phenotypic variability observed reflects the limited historical domestication of *M. dubia*, with natural dispersal occurring primarily via rivers and minimal anthropogenic selection prior to the 20th century. Beyond its evolutionary significance, this limited domestication status presents a unique opportunity: unlike many crops that have lost genetic diversity through domestication bottlenecks, *M. dubia* retains the adaptive variation necessary for developing climate-resilient cultivars while improvements in yield and quality remain achievable through conventional breeding.

Most significantly, our study identified three high-value accessions (PER1000416, PER1000423, and PER1000411) that demonstrated superior combinations of commercially important descriptors, increased fruit weight, higher pulp content, and reduced seed count. These promising accessions constitute valuable genetic resources for future breeding programs aimed at enhancing the productivity and quality of *M. dubia*. This characterization provides actionable guidance for stakeholders: farmers can select accessions based on specific market demands (PER1000416 for maximum fruit size, PER1000411 for processing efficiency), while conservation programs can prioritize maintaining representatives from all four phenetic groups to preserve functional diversity. Regional development agencies can use these data to establish demonstration plots showcasing superior accessions, facilitating adoption through farmer-to-farmer knowledge transfer networks. The immediate next steps should include: (1) clonal propagation of these elite accessions for multi-location trials across different flooding regimes and soil types, (2) controlled crosses among superior accessions to combine complementary traits, (3) on-farm validation with smallholder farmers to assess performance under diverse management conditions, and (4) establishment of multiplication gardens to ensure rapid variety dissemination upon release.

The promising accessions identified through this characterization represent not only valuable genetic resources for breeding programs, but also tangible assets for sustainable economic development in the Peruvian Amazon. By developing improved varieties based on these accessions, substantial economic opportunities can be created for local communities while maintaining biodiversity conservation through the continued maintenance of *ex situ* germplasm collection. The integration of this characterized germplasm into regional development programs could provide a model for the sustainable utilization of Amazonian plant genetic resources that balance conservation with economic prosperity. However, realizing this potential requires supportive policies, including: (1) public investment in multiplication infrastructure, (2) technical assistance programs for cultivation practices, (3) market linkage initiatives connecting farmers to processors, and (4) benefit-sharing mechanisms ensuring that communities retain value from their traditional knowledge.

The characterization of *M. dubia* germplasm presented in this study carries profound implications for conservation policy and sustainable development strategies in the Amazon region. Our

findings demonstrate that native fruit species like *M. dubia* represent critical components of integrated agroforestry systems that can reconcile economic development with biodiversity conservation. The absence of geographic structure in phenotypic variation supports policy frameworks that prioritize habitat diversity over simple geographic representation in conservation planning, potentially improving the cost-effectiveness of conservation investments. Furthermore, the identification of promising germplasm suitable for cultivation provides tangible pathways for implementing agroforestry systems that maintain 60–70% of native biodiversity while generating income comparable to conventional agriculture. These results underscore the need for policies that support *ex situ* germplasm conservation networks, facilitate benefit-sharing agreements with indigenous communities who maintain traditional knowledge, and incentivize the adoption of diversified agroforestry systems over monoculture expansion. As the Amazon faces unprecedented pressures from climate change and deforestation, the development of economically viable native species like *M. dubia* becomes increasingly critical for creating sustainable alternatives that preserve forest cover while supporting rural livelihoods. The success of such initiatives requires coordinated policies that integrate agricultural development, biodiversity conservation, and climate adaptation strategies, positioning characterized native germplasm as a cornerstone of sustainable Amazonian development.

Additionally, this study demonstrates that systematic germplasm characterization directly supports conservation and utilization strategies for underutilized native fruit species. By documenting and analyzing the phenotypic diversity of *M. dubia*, we provide essential baseline data for evidence-based decisions in breeding, conservation, and development programs. The lack of geographic structure in phenotypic variation fundamentally changes how future conservation efforts should prioritize phenotypic diversity over geographic representation, potentially doubling the efficiency of the limited conservation resources. Critical steps include: (1) molecular characterization using SNP markers to validate phenetic groups and identify quantitative trait loci, (2) multi-environment trials across at least three flooding gradients and two soil types to assess descriptor stability, and (3) participatory selection with local communities to incorporate traditional selection criteria and ensure variety adoption.

Finally, the integration of advanced genomic technologies, including genotyping-by-sequencing and genome editing approaches such as CRISPR/Cas, represents a transformative opportunity for *M. dubia* improvement. These tools, combined with systems biology approaches that integrate multi-omics data, could accelerate, from the current 15–20 year breeding cycle to 7–10 years, the development of improved varieties with enhanced productivity, nutritional quality, and resilience to environmental stress. However, technology alone is insufficient; successful transformation requires participatory approaches that engage local communities, incorporate traditional knowledge, and have the potential to transform throughout the innovation process. As climate change intensifies pressure on Amazonian ecosystems, continued research into the adaptive traits of *M. dubia* becomes

increasingly critical. Future investigations integrating genomic tools with traditional knowledge will accelerate the development of climate-resilient varieties while preserving cultural connections to this remarkable fruit. The phenotypic foundation established here represents the first step toward transforming *M. dubia* from an underutilized species to a model for sustainable intensification that harmonizes conservation with development in tropical regions worldwide.

## 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 author/s.

## Author contributions

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

## Funding

The author(s) declare financial support was received for the research, and/or publication of this article. This research was supported by the National Council for Science, Technology, and Technological Innovation (CONCYTEC) through the PROCENCIA program under the Basic Research Projects funding initiative “E041-2024-03” (Contract Number PE501088786-2024-PROCENCIA). We also acknowledge the Universidad Nacional de la Amazonía Peruana (UNAP) for institutional support through the approval of research projects under Resolución Rectoral N° 0449-2024-UNAP. Additional support was provided by the “Conservation and Sustainable Use of Genetic Resources of the National Germplasm Bank of INIA Maintained Under Ex Situ Conditions” project (Budget Program 0121).

## Acknowledgments

We express our sincere gratitude to the field technical staff—Lolo Lumba, Angel Vizcarra, Américo Tuesta, and Celsio Bereca—for their dedicated assistance with data collection and meticulous maintenance of the germplasm collection. Their expertise and

commitment were instrumental in the successful phenotypic characterization of the *M. dubia* accessions presented in this study. The authors also wish to acknowledge the institutional support provided by the Instituto Nacional de Innovación Agraria (INIA) for access to the *ex situ* germplasm collection that enabled this research.

## 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/fcosc.2025.1623515/full#supplementary-material>

## References

- Alves, R. M., de Abreu, V. A. C., Oliveira, R. P., Almeida, J. V. D. A., de Oliveira, M. M., Silva, S. R., et al. (2024). Genomic decoding of *Theobroma grandiflorum* (cupuassu) at chromosomal scale: evolutionary insights for horticultural innovation. *GigaScience* 13, giae027. doi: 10.1093/gigascience/giae027
- Alves, R. M., de Sousa, C. R., da Conceição, M. S., de Souza, D. C., and Sebbenn, A. M. (2013). Diversidade genética em coleções amazônicas de germoplasma de cupuaçuzeiro [*Theobroma grandiflorum* (Willd. ex Spreng.) Schum. *Rev. Bras. Frutic* 35, 818–828. doi: 10.1590/S0100-29452013000300019
- Ampatzidis, Y., Partel, V., and Costa, L. (2020). Agrovie: Cloud-based application to process, analyze and visualize UAV-collected data for precision agriculture applications utilizing artificial intelligence. *Comput. Electron Agric.* 174, 105457. doi: 10.1016/j.compag.2020.105457
- Araujo, E. C. G., Silva, T. C., da Cunha Neto, E. M., Favarin, J. A. S., da Silva Gomes, J. K., das Chagas, K. P. T., et al. (2024). Bioeconomy in the Amazon: Lessons and gaps from thirty years of non-timber forest products research. *J. Environ. Manage* 370, 122420. doi: 10.1016/j.jenvman.2024.122420
- Awada, L., Phillips, P. W. B., and Bodan, A. M. (2024). The evolution of plant phenomics: global insights, trends, and collaborations, (2000-2021). *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1410738
- Azevedo, L., de Araujo Ribeiro, P. F., de Carvalho Oliveira, J. A., Correia, M. G., Ramos, F. M., de Oliveira, E. B., et al. (2019). Camu-camu (*Myrciaria dubia*) from commercial cultivation has higher levels of bioactive compounds than native cultivation (Amazon Forest) and presents antimutagenic effects *in vivo*. *J. Sci. Food Agric.* 99, 624–631. doi: 10.1002/jsfa.9224
- Bardales-Lozano, R. M., Chagas, E. A., Smiderle, O., Abanto-Rodriguez, C., Chagas, P. C., Mota Filho, A. B., et al. (2016). Genetic divergence among camu-camu plant populations based on the initial characteristics of the plants. *J. Agric. Sci.* 8, 51–58. doi: 10.5539/jas.v8n11p51
- Blare, T., and Donovan, J. (2018). Building value chains for indigenous fruits: lessons from camu-camu in Peru. *Renew Agr Food Syst.* 33, 6–18. doi: 10.1017/S1742170516000181
- Čalić, I., Bussotti, F., Martínez-García, P. J., and Neale, D. B. (2015). Recent landscape genomics studies in forest trees—what can we believe? *Tree Genet. Genomes* 12, 1–7. doi: 10.1007/s11295-015-0960-0
- Campbell, A. J., Carvalheiro, L. G., Maués, M. M., Jaffé, R., Giannini, T. C., and Freitas, M. A. B. (2018). Anthropogenic disturbance of tropical forests threatens pollination services to açai palm in the Amazon river delta. *J. Appl. Ecol.* 55, 1725–1736. doi: 10.1111/1365-2664.13086
- Castro, J. C., Castro, C., and Cobos, M. (2023). Genetic and biochemical strategies for regulation of L-ascorbic acid biosynthesis in plants through the L-galactose pathway. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1099829
- Castro, J. C., Maddox, J. D., and Imán, S. A. (2018). "Camu-camu—*Myrciaria dubia* (Kunth) McVaugh," in *Exotic Fruits Reference Guide, First*. Eds. S. Rodrigues, E. Sousa de Brito and E. de Oliveira Silva (Cambridge, Elsevier Inc.), 97–105.
- Castro, J. C., Vásquez-Guizado, S. J., Vigil, B. E., Ascue, F., Rojas-Villa, N., Paredes, J. D., et al. (2024). Development and application of microsatellite markers for genetic diversity assessment and construction of a core collection of *Myrciaria dubia* (Kunth) McVaugh germplasm from the Peruvian amazon. *Forests* 15, 22–22. doi: 10.3390/f15111873
- Castro, J. C., Gutiérrez, F., Acuña, C., Cerdeira, L. A., Tapullima, A., Cobos, M., et al. (2013). Variación del contenido de vitamina C y antocianinas en *Myrciaria dubia* "camu camu. *Rev. Soc. Quím Peru* 79, 319–330.
- Ceccarelli, S. (2015). Efficiency of plant breeding. *Crop Sci.* 55, 87–97. doi: 10.2135/cropsci2014.02.0158
- Ceccarelli, S., and Grandi, S. (2020). Participatory plant breeding: Who did it, who does it and where? *Exp. Agric.* 56, 1–11. doi: 10.1017/S0014479719000127
- Chagas, E. A., Lozano, R. M. B., Chagas, P. C., Bacelar-Lima, C. G., Garcia, M. I. R., Oliveira, J. V., et al. (2015). Variabilidade intraespecífica de frutos de camu-camu em populações nativas na amazônia setentrional. *Crop Breed Appl. Biotechnol.* 15, 265–271. doi: 10.1590/1984-70332015v15n4a44
- Chakraborty, S., and Külheim, C. (2025). Trends in tree improvement methods: from classical breeding to genomic technologies. *Tree Genet. Genomes* 21, 1–21. doi: 10.1007/s11295-025-01698-6
- Chen, C., Chang, H., Pang, X., Liu, Q., Xue, L., and Yin, C. (2025). Genetic diversity analysis and conservation strategy recommendations for *ex situ* conservation of *Cupressus chengiana*. *BMC Plant Biol.* 25, 1–12. doi: 10.1186/s12870-025-06581-z
- Chirinos, R., Galarza, J., Betalleluz-Pallardel, I., Pedreschi, R., and Campos, D. (2010). Antioxidant compounds and antioxidant capacity of Peruvian camu camu (*Myrciaria dubia* (H.B.K.) McVaugh) fruit at different maturity stages. *Food Chem.* 120, 1019–1024. doi: 10.1016/j.foodchem.2009.11.041
- Clement, C. R., De Cristo-Araújo, M., Coppens D'Eeckenbrugge, G., Alves Pereira, A., and Picanço-Rodrigues, D. (2010). Origin and domestication of native amazonian crops. *Diversity* 2, 72–106. doi: 10.3390/d2010072
- Colantonio, V., Ferrão, L. F. V., Tieman, D. M., Bliznyuk, N., Sims, C., Klee, H. J., et al. (2022). Metabolomic selection for enhanced fruit flavor. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2115865119. doi: 10.1073/pnas.2115865119
- Conceição, N., Albuquerque, B. R., Pereira, C., Corrêa, R. C. G., Lopes, C. B., Calheta, R. C., et al. (2019). By-products of camu-camu [*Myrciaria dubia* (Kunth) McVaugh] as promising sources of bioactive high added-value food ingredients: functionalization of yogurts. *Molecules* 25, 70. doi: 10.3390/molecules25010070
- Cramér, H. (1946). *Mathematical methods of statistics*. (Princeton University Press, Princeton, NJ).
- de Oliveira, E. A., Marimon, B. S., Feldpausch, T. R., Colli, G. R., Marimon-Junior, B. H., Lloyd, J., et al. (2014). Diversity, abundance and distribution of lianas of the Cerrado-Amazonian forest transition, Brazil. *Plant Ecol. Divers.* 7, 231–240. doi: 10.1080/17550874.2013.816799

- Do, N. Q., Zheng, S., Oh, S., Nguyen, Q. T. N., Fang, M., Kim, M., et al. (2021). Anti-allergic effects of *Myrciaria dubia* (Camu-Camu) fruit extract by inhibiting histamine H1 and H4 receptors and histidine decarboxylase in RBL-2H3 cells. *Antioxidants* 11, 104. doi: 10.3390/antiox11010104
- Du, X., Liu, N., Lu, P., Wang, Y., Lu, B., Tian, S., et al. (2024). RNA-seq-based transcriptome profiling of early fruit development in Chieh-qua and analysis of related transcription factors. *Sci. Rep.* 14, 13489. doi: 10.1038/s41598-024-63871-6
- Dulloo, M. E., Thormann, I., Fiorino, E., De Felice, S., Rao, V. R., and Snook, L. (2013). Trends in research using plant genetic resources from germplasm collections: from 1996 to 2006. *Crop Sci.* 53, 1217–1227. doi: 10.2135/cropsci2012.04.0219
- Feng, G., Wu, J., and Yi, H. (2019). Global tissue-specific transcriptome analysis of *Citrus sinensis* fruit across six developmental stages. *Sci. Data* 6, 153. doi: 10.1038/s41597-019-0162-y
- Ferreira, G. A. C. (2020). *Camu-camu* (*Myrciaria dubia* (Kunth) McVaugh) e seus polinizadores: Produtividade, diversidade e interações na Amazônia Central, Brasil (Thesis, Instituto Nacional de Pesquisas da Amazônia – Inpa, Manaus, Brazil).
- Ferreira, C. S., Figueira, A. V. O., Gribel, R., Wittmann, F., and Piedade, M. T. F. (2010). "Genetic variability, divergence and speciation in trees of periodically flooded forests of the Amazon: A case study of *Himatanthus sucuuba* (Spruce) woodson," in *Amazonian Floodplain Forests, 1st edn.* Eds. W. J. Junk, M. T. Piedade and F. Wittmann (Springer, Dordrecht), 618.
- Fisher, R. A. (1925). *Statistical Methods Research Workers. 11th edn* (London: Oliver and Boyd).
- Fraccasetti, D., Costa, C., Moulay, L., and Tomás-Barberán, F. A. (2013). Ellagic acid derivatives, ellagitannins, proanthocyanidins and other phenolics, vitamin C and antioxidant capacity of two powder products from camu-camu fruit (*Myrciaria dubia*). *Food Chem.* 139, 578–588. doi: 10.1016/j.foodchem.2013.01.121
- Freitas, C. A. B., Silva, A. S., Alves, C. N., Nascimento, W. M. O., Lopes, A. S., Lima, M. O., et al. (2016). Characterization of the fruit pulp of Camu-Camu. *J. Braz. Chem. Soc.* 27, 1838–1846. doi: 10.5935/0103-5053.20160067
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453–467. doi: 10.1093/biomet/58.3.453
- Gaete-Eastman, C., Stappung, Y., Molinett, S., Urbina, D., Moya-León, M. A., and Herrera, R. (2022). RNAseq, transcriptome analysis and identification of DEGs involved in development and ripening of *Fragaria chiloensis* fruit. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.976901
- Galluzzi, G., Eyzaguirre, P., and Negri, V. (2010). Home gardens: neglected hotspots of agro-biodiversity and cultural diversity. *Biodivers Conserv.* 19, 3635–3654. doi: 10.1007/s10531-010-9919-5
- García-Chacón, J. M., Marin-Loaiza, J. C., and Osorio, C. (2023). Camu camu (*Myrciaria dubia* (Kunth) McVaugh): an amazonian fruit with biofunctional properties—A review. *ACS Omega* 8, 5169–5183. doi: 10.1021/acsomega.2c07245
- Govindaraj, M., Vetriventhan, M., and Srinivasan, M. (2015). Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet. Res. Int.* 2015, 431487. doi: 10.1155/2015/431487
- Grigo, M. L., Chagas, E. A., Berlingieri, M. F., de Andrade, A., Mota, A. B., and Cardoso, P. (2016). Determination of harvest time and quality of native camu-camu fruits (*Myrciaria dubia* (Kunth) Mc Vaugh) during storage. *Fruits* 71, 373–378. doi: 10.1051/fruits/2016029
- Hagphanah, M., Hashemipetroudi, S., Arzani, A., and Araniti, F. (2024). Drought tolerance in plants: physiological and molecular responses. *Plants* 13, 2962. doi: 10.3390/plants13212962
- Hernández-Delgado, S., Padilla-Ramírez, J. S., and Mayek-Pérez, N. (2018). Morphologic characterization of guava germplasm from México: Implications about its conservation and breeding. *Rev. Bras. Frutic* 40, 11. doi: 10.1590/0100-29452018887
- Hotelling, H. (1931). The generalization of student's ratio. *Ann. Math Stat.* 2, 360–378. doi: 10.1214/aoms/1177732979
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24, 417–441. doi: 10.1037/h0071325
- Imán, S., and Aldana, M. (2007). *Tecnología para la producción del camu camu Myrciaria dubia (H.B.K.) Mc Vaugh* (Lima: Instituto Nacional de Innovación Agraria).
- Imán, S., Chuquizuta, B., Samanamud, A., and Ochoa, M. (2022). *Descriptores para camu camu Myrciaria dubia (Kunth) Mc Vaugh. 1st edn* (Iquitos: Instituto Nacional de Innovación Agraria - INIA).
- Imán, S., Pinedo, S., and Melchor, M. (2011). Caracterización y evaluación morfoagronómica de germoplasma de camu camu *Myrciaria dubia* McVaugh. *Sci. Agropecu* 2, 189–201.
- Iqbal, M. M., Erskine, W., Berger, J. D., and Nelson, M. N. (2020). Phenotypic characterisation and linkage mapping of domestication syndrome traits in yellow lupin (*Lupinus luteus* L.). *Theor. Appl. Genet.* 133, 2975–2987. doi: 10.1007/s00122-020-03650-9
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* 20, 141–151. doi: 10.1177/001316446002000116
- Klápště, J., Ashby, R. L., Telfer, E. J., Graham, N. J., Dungey, H. S., Brauning, R., et al. (2021). The use of "Genotyping-by-sequencing" to recover shared genealogy in genetically diverse eucalyptus populations. *Forests* 12, 904. doi: 10.3390/f12070904
- Kumar, R., Bohra, A., Pandey, A. K., and Kumar, A. (2017). Metabolomics for plant improvement: Status and prospects. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01302
- Kumar, S., Hilario, E., Deng, C. H., and Molloy, C. (2020). Turbocharging introgression breeding of perennial fruit crops: a case study on apple. *Hortic. Res.* 7, 47. doi: 10.1038/s41438-020-0270-z
- Lawley, D. N. (1938). A generalization of Fisher's Z test. *Biometrika* 30, 180–187. doi: 10.1093/biomet/30.1-2.180
- Leal, B. S. S., Tavares, V. C., Watanabe, M. T. C., Cardoso, A. L. R., Tyski, L., Alves-Pereira, A., et al. (2025). Genetic-based conservation status indicators applied to endemics with restricted distributions: a case for eastern Amazonian cangas plants. *Ann. Bot.*, mcaf030. doi: 10.1093/aob/mcaf030
- Lemke, I. H., Kolb, A., and Diekmann, M. R. (2012). Region and site conditions affect phenotypic trait variation in five forest herbs. *Acta Oecol* 39, 18–24. doi: 10.1016/j.actao.2011.11.001
- Lim, T. K. (2012). "Myrciaria dubia," in *Edible Medicinal and Non Medicinal Plants, First. Ed.* T. K. Lim (Springer, Berlin), 1–159.
- Liu, U., Gianella, M., Dávila, P., Diazgranados, M., Flores, C. M., Lira-Saade, R., et al. (2023). Conserving useful plants for a sustainable future: species coverage, spatial distribution, and conservation status within the Millennium Seed Bank collection. *Biodivers Conserv.* 32, 2791–2839. doi: 10.1007/s10531-023-02631-w
- Luize, B. G., Palma-Silva, C., Siqueira, T., and Silva, T. S. F. (2024). Tree species occurring in Amazonian wetland forests consistently show broader range sizes and niche breadths than trees in upland forests. *Ecol. Evol.* 14, e11230. doi: 10.1002/ece3.11230
- Maestri, R., and Duarte, L. (2020). Evoregions: Mapping shifts in phylogenetic turnover across biogeographic regions. *Methods Ecol. Evol.* 11, 1652–1662. doi: 10.1111/2041-210X.13492
- Mall, A. K., Manimekalai, R., Misra, V., Pandey, H., Srivastava, S., and Sharma, A. (2024). CRISPR/Cas-mediated genome editing for sugarcane improvement. *Sugar Tech* 27, 1–13. doi: 10.1007/s12355-023-01352-2
- Manickavasagam, G., San, P. W. C., Gorji, S. G., Sunghong, B., Keong, Y. Y., Fitzgerald, M., et al. (2024). Unbiased metabolomics of volatile secondary metabolites in essential oils originated from myrtaceae species. *Chem. Afr* 7, 3067–3075. doi: 10.1007/s42250-024-01000-6
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.
- Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605
- MIDAGRI (2025). *Sistema Integrado de Estadísticas Agrarias Vol. 2* (Perfil productivo regional). Ministerio de Desarrollo Agrario y Riego, Lima, Peru.
- Miller, R. P., and Nair, P. K. R. (2006). Indigenous agroforestry systems in Amazonia: From prehistory to today. *Agrofor Syst.* 66, 151–164. doi: 10.1007/s10457-005-6074-1
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Pérez-Hernández, O., Eskridge, K. M., et al. (2016). A genomic Bayesian multi-trait and multi-environment model. *G3* 6, 2725–2744. doi: 10.1534/g3.116.032359
- Morales, A., Seguel, I., and Diaz, L. (2019). "Caracterización de germoplasma de quinoa del sur de Chile," in *Quinoa del sur de Chile. Alternativa productiva y agroindustrial de alto valor, 1st edn.* Ed. J. Díaz-Sánchez (Instituto de Investigaciones Agropecuarias, Temuco), 7–23.
- Murtagh, F., and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *J. Classif* 31, 274–295. doi: 10.1007/s00357-014-9161-z
- Neves, L. C., de Campos, A. J., Cisneros-Zevallos, L., Colombo, R. C., and Roberto, S. R. (2015). Post-harvest behavior of camu-camu fruits based on harvesting time and nutraceutical properties. *Sci. Hortic.* 186, 223–229. doi: 10.1016/j.scienta.2015.02.030
- Newton, P., Agrawal, A., and Wollenberg, L. (2013). Enhancing the sustainability of commodity supply chains in tropical forest and agricultural landscapes. *Glob Environ. Change* 23, 1761–1772. doi: 10.1016/j.gloenvcha.2013.08.004
- Nguyen, V. H., Morante, R. I. Z., Lopena, V., Verdeprado, H., Murori, R., Ndayiragije, A., et al. (2023). Multi-environment genomic selection in rice elite breeding lines. *Rice* 16, 1–17. doi: 10.1186/s12284-023-00623-6
- Nunes, C. F., Setotaw, T. A., Pasqual, M., Chagas, E. A., Santos, E. G., Santos, D. N., et al. (2017). *Myrciaria dubia*, an Amazonian fruit: Population structure and its implications for germplasm conservation and genetic improvement. *GMR* 16, 1–12. doi: 10.4238/gmr16019409
- Oliva, C., Vargas, V., and Linares, C. (2005). Selección de plantas madre promisorias de *Myrciaria dubia* (HBK) Mc Vaugh, camu camu arbustivo, en Ucayali-Perú. *Folia* 14, 85–89. doi: 10.24841/fa.v14i2.407
- Oliveira, R. S., Eller, C. B., Barros, F. D. V., Hirota, M., Brum, M., and Bittencourt, P. (2019). Linking plant hydraulics and the fast-slow continuum to understand resilience to drought in tropical ecosystems. *New Phytol.* 230, 904–923. doi: 10.1111/nph.17266
- Paiva, J., and Das Chagas, F. (2016). Camu-Camu super fruit (*Myrciaria dubia* (H.B.K) Mc Vaugh) at different maturity stages. *Afr. J. Agric. Res.* 11, 2519–2523. doi: 10.5897/ajar2016.11167
- Palmer, M. A., Menninger, H. L., and Bernhardt, E. (2010). River restoration, habitat heterogeneity and biodiversity: a failure of theory or practice? *Freshw. Biol.* 55, 205–222. doi: 10.1111/j.1365-2427.2009.02372.x

- Panis, B., Nagel, M., and Van den houwe, I. (2020). Challenges and prospects for the conservation of crop genetic resources in field genebanks, in *in vitro* collections and/or in liquid nitrogen. *Plants* 9, 1634. doi: 10.3390/plants9121634
- Pearson, K. (1896). VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia. *Philos. Trans. R Soc. London* 187, 253–318. doi: 10.1098/rsta.1896.0007
- Penn, J. W. (2006). The cultivation of camu camu (*Myrciaria dubia*): A tree planting programme in the Peruvian Amazon. *For Trees Livelihood* 16, 85–101. doi: 10.1080/14728028.2006.9752547
- Peters, C. M., and Vasquez, A. (1987). Estudios ecológicos de Camu-Camu (*Myrciaria dubia*). I. Producción de frutos en poblaciones naturales. *Acta Amaz* 17, 161–188. doi: 10.1590/1809-43921987171174
- Pickersgill, B. (2016). “Domestication of plants in mesoamerica: an archaeological review with some ethnobotanical interpretations,” in *Ethnobotany of Mexico, 1st edn*. Eds. R. Lira and A. Casas (Springer, New York), 207–231.
- Pickersgill, B., Heiser, C. B., and Harris, D. R. (1976). Cytogenetics and evolutionary change under domestication. *Philos. Trans. R Soc. London Ser. B* 275, 55–69. doi: 10.1098/rstb.1976.0070
- Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *Ann. Math Stat.* 26, 117–121. doi: 10.1214/aoms/1177728599
- Pinedo, M. (2009). Camu-camu: innovación del agro en la amazonía Peruana; perspectivas. *Encuentro Económico, Región Loreto* 22–3. Banco Central de Reserva del Perú, Lima. Available at: <http://www.bcrp.gob.pe/docs/Proyeccion-Institucional/Encuentros-Regionales/2009/Loreto/EER-Loreto-Mario-Pinedo.pdf>
- Pinedo, M. (2017). *Seleção de Genótipos superiores em coleções ex situ de Camu-Camu [Myrciaria dubia (Kunth) McVaugh] da Amazônia Peruana* Vol. 141 (Universidade Federal de Roraima, Programa de Pós-Graduação em Biodiversidade e Biotecnologia da Amazônia Legal, Boa Vista, Brazil).
- Pinedo Panduro, M., Linares Bensimón, C., Mendoza Zuñiga, H., and Anguiz, R. (2004). *Plan de mejoramiento genético de camu camu* (Instituto de Investigaciones de la Amazonía Peruana, Iquitos, Peru).
- Pott, D. M., Vallarino, J. G., and Osorio, S. (2020). Metabolite changes during postharvest storage: effects on fruit quality traits. *Metabolites* 10, 187. doi: 10.3390/metabo10050187
- Puente Ganz, L. (2008). *Validación Clonal de Plantas Madres Promisorias de Myrciaria dubia (H.B.K.) Mc Vaugh “camu camu arbustivo”, en Cámaras de Sub Irrigación en Ucayali – Perú* (Thesis, Universidad Nacional Agraria de la Selva, Tingo María, Peru).
- Redondo-López, A., González-Schain, N., Perales, M., and Conde, D. (2025). “CRISPR-Cas in woody perennial plants: methods, efficiency, applications, and challenges to creating commercial varieties with high ecological and economic value,” in *CRISPR-Cas Methods*. Eds. M. T. Islam, K. Molla, P. Bhowmik and K. Xie (Humana, New York, NY).
- Reed, J., van Vianen, J., Foli, S., Clendenning, J., Yang, K., MacDonald, M., et al. (2019). Trees for life: The ecosystem service contribution of trees to food production and livelihoods in the tropics. *For Policy Econ* 84, 62–71. doi: 10.1016/j.forpol.2017.01.012
- Rogers, D. J., and Tanimoto, T. T. (1960). A computer program for classifying plants. *Science* 132, 1115–1118. doi: 10.1126/science.132.3434.1115
- Roy, S. N. (1958). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Stat.* 24, 220–238. doi: 10.1214/aoms/1177729029
- Sakazaki, R. T., Chagas, E. A., Abanto-Rodríguez, C., Chagas, P., de Araujo, M. C., Neto, J. L., et al. (2022). Selection of *Myrciaria dubia* clones under conditions of the savanna/forest transition of Roraima through multivariate analysis. *Agron. Colomb* 40, 3–11. doi: 10.15446/agron.colomb.v40n1.100319
- Saldanha, L. L., Allard, P. - M., Afzan, A., de Melo, F. P. d. S. R., Marcourt, L., Queiroz, E. F., et al. (2020). Metabolomics of *Myrcia bella* populations in Brazilian savanna reveals strong influence of environmental factors on its specialized metabolism. *Molecules* 25, 2954. doi: 10.3390/molecules25122954
- Sampaio de Paiva, R. M. (2019). *Caracterização agrônômica e molecular de genótipos de Myrciaria dubia (Kunth) McVaugh provenientes da Amazonia setentrional* (Thesis, Universidade Federal de Roraima, Boa Vista, Brazil).
- Schroth, G., and Ruf, F. (2013). Farmer strategies for tree crop diversification in the humid tropics. A review. *Agron. Sustain Dev.* 34, 139–154. doi: 10.1007/s13593-013-0175-4
- Schut, M., Klerkx, L., Sartas, M., Lamers, D., Campbell, M. M., Ogbonna, I., et al. (2016). Innovation platforms: experiences with their institutional embedding in agricultural research for development. *Exp. Agric.* 52, 537–561. doi: 10.1017/S001447971500023X
- Scott, A. J., and Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. *Biometrics* 30, 507–512. doi: 10.2307/2529204
- Seleiman, M. F., Al-Suhaibani, N., Ali, N., Akmal, M., Alotaibi, M., Refay, Y., et al. (2021). Drought stress impacts on plants and different approaches to alleviate its adverse effects. *Plants* 10, 259. doi: 10.3390/plants10020259
- Shannon, C. E. (1948). A mathematical theory of communication. *BSTJ* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x
- Šmid, J., Kalousová, M., Mandák, B., Houška, J., Chládková, A., Pinedo, M., et al. (2017). Morphological and genetic diversity of camu-camu [*Myrciaria dubia* (Kunth) McVaugh] in the Peruvian Amazon. *PLoS One* 12, e0179886. doi: 10.1371/journal.pone.0179886
- Spearman, C. (1904). The proof and measurement of association between two things. *AJP* 15, 72–101. doi: 10.2307/1412159
- Stetter, M. G. (2020). Limits and constraints to crop domestication. *Am. J. Bot.* 107, 1617–1621. doi: 10.1002/ajb2.1585
- Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063
- Tao, H., Xu, S., Tian, Y., Li, Z., Ge, Y., Zhang, J., et al. (2022). Proximal and remote sensing in plant phenomics: 20 years of progress, challenges, and perspectives. *Plant Commun.* 3, 100344. doi: 10.1016/j.xplc.2022.100344
- Thom, G., Xue, A. T., Sawakuchi, A. O., Ribas, C. C., Hickerson, M. J., Aleixo, A., et al. (2020). Quaternary climate changes as speciation drivers in the Amazon floodplains. *Sci. Adv.* 6, eaax4718. doi: 10.1126/sciadv.aax4718
- Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika* 18, 267–276. doi: 10.1007/bf02289263
- Verde, W. G., and Ríos, J. C. (2018). Suelos potenciales para el cultivo de camu camu (*Myrciaria dubia* H. B. K. McVaugh) en la Provincia de Coronel Portillo, región Ucayali. *Anales Científicos* 79, 151–158. doi: 10.21704/ac.v79i1.1157
- Villachica, H. (1996). *El cultivo del camu camu (Myrciaria dubia H.B.K. McVaugh) en la Amazonia Peruana. 1st edn* (Lima: Tratado de Cooperación Amazónica).
- Walsh, K. B., McGlone, V. A., and Han, D. H. (2020). The uses of near infra-red spectroscopy in postharvest decision support: A review. *Postharvest Biol. Technol.* 163, 111139. doi: 10.1016/j.postharvbio.2020.111139
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244. doi: 10.1080/01621459.1963.10500845
- Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M. - D., et al. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nat. Plants* 4, 23–29. doi: 10.1038/s41477-017-0083-8
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* 9, 60–62. doi: 10.1214/AOMS/1177732360
- Willan, R. L. (1958). A guide to forest seed handling: with special reference to the tropics, volumen 2. FAO forestry paper No. 20/2. FAO, Rome, 379.
- Zolkafli, S. H., Marjuni, M., Abdullah, N., Singh, R., and Ithnin, M. (2025). Exploring diversity in African oil palm (*Elaeis guineensis* Jacq.) germplasm populations via genotyping-by-sequencing. *Genet. Res. Crop Evol.* 72, 6111–6127. doi: 10.1007/s10722-025-02330-1