frontiers in Soil Science

TYPE Original Research
PUBLISHED 15 April 2026
DOI 10.3389/fsoil.2026.1786422

Check for updates

Integrated multivariate analysis of morphological and yield traits in native *Capsicum chinense* ecotypes grown in acidic soils of the Peruvian Amazon

Henry Díaz-Chuquizuta [1], Luis Fernando Manrique Gonzales [2], Martín Sánchez Ojanasta [2], Juan Pablo Cuevas-Giménez [3], Boris Martínez [4], Geiner Flores Sinti [5], Juan Jose Kerry Tanchiva [5] and Geomar Vallejos-Torres [2,6*]

[1]Dirección de Servicios Estratégicos Agrarios, Estación Experimental Agraria San Ramón, Instituto Nacional de Innovación Agraria (INIA), Yurimaguas, Alto Amazonas, Peru, [2]Dirección de Servicios Estratégicos Agrarios, Estación Experimental Agraria El Porvenir, Instituto Nacional de Innovación Agraria (INIA), San Martín, Peru, [3]Estación Experimental Agraria de Baños del Inca, Dirección de Servicios Estratégicos Agrarios, Instituto Nacional de Innovación Agraria (INIA), Cajamarca, Peru, [4]Estación Experimental Agraria San Ramón. Dirección de Investigación y Desarrollo Tecnológico, Instituto Nacional de Innovación Agraria (INIA), Yurimaguas, Alto Amazonas, Peru, [5]Universidad Nacional Autónoma de Alto Amazonas, Yurimaguas, Alto Amazonas, Peru, [6]Facultad de Ciencias Agrarias, Universidad Nacional de San Martín, Tarapoto, San Martín, Peru

OPEN ACCESS

EDITED BY
Mohamed T. El-Saadony,
Zagazig University, Egypt

REVIEWED BY
Yassine Mouniane,
Ibn Tofail University, Morocco
Uzma Arif,
HITEC University, Pakistan

*CORRESPONDENCE
Geomar Vallejos-Torres
gvallejos@unsm.edu.pe

RECEIVED 12 January 2026
REVISED 17 March 2026
ACCEPTED 25 March 2026
PUBLISHED 15 April 2026

CITATION
Díaz-Chuquizuta H, Manrique Gonzales LF, Sánchez Ojanasta M, Cuevas-Giménez JP, Martínez B, Flores Sinti G, Kerry Tanchiva JJ and Vallejos-Torres G (2026) Integrated multivariate analysis of morphological and yield traits in native *Capsicum chinense* ecotypes grown in acidic soils of the Peruvian Amazon. *Front. Soil Sci.* 6:1786422. doi: 10.3389/fsoil.2026.1786422

Introduction: The comprehensive characterization of native *Capsicum chinense* ecotypes represents a strategic priority for genetic improvement, germplasm conservation, and the sustainable use of Amazonian crops. The objective of this study was to evaluate morphological, phenological, and productive variability among 12 ecotypes from the Peruvian Amazon by integrating multivariate análisis and machine learning with soil physicochemical characterization.

Methods: The research was conducted on acidic tropical soils with low organic matter content and limited availability of exchangeable bases, conditions representative of degraded Amazonian agroecosystems, which enabled the assessment of soil–plant interactions and their influence on phenotypic expression and crop yield.

Results: The results revealed a broad, well-structured range of phenotypic variability, with fruit diameter, fruit length, fruit weight, and seed weight identified as the primary morphological determinants of yield and adaptive capacity under low-fertility soil conditions. Principal component analysis indicated that four components explained more than 70% of the total variance, primarily associated with productivity, fruit morphometry, and phenological traits. Cluster análisis identified groups with high internal consistency, while linear discriminant analysis validated the phenotypic structure, achieving a classification accuracy of 91.8%. The ecotypes JEB-028 and LAG-022 exhibited superior productive performance, whereas BAL-012 and YUR-001 demonstrated greater phenotypic stability under restrictive soil conditions.

Discussion: Overall, these findings confirm the strategic value of native Amazonian germplasm and underscore the importance of integrating edaphic diagnostics into genetic selection programs and into strategies for the sustainable management and restoration of degraded agricultural soils in the Amazon.

KEYWORDS

Acidic soils, *Capsicum chinense*, multivariate analyses, Peruvian Amazon, phenotypic variability

1 Introduction

The *Capsicum* genus harbors extensive genetic and phenotypic diversity, making it a strategic resource for food security, the food industry, and biotechnology. Over the past decade, the application of multivariate and machine learning approaches to characterize phenotypic diversity and predict yield in horticultural crops has accelerated, facilitating the identification of key traits of value for breeding programs and germplasm industrialization (1, 2). These integrated methodologies, including principal component analysis (PCA), clustering techniques, and predictive models such as Random Forest (RF), have proven particularly effective in reducing the dimensionality of complex phenotypic datasets and prioritizing traits with the most significant contribution to yield (1, 2).

At the international level, research on *Capsicum* has achieved substantial advances in genetics, phenotyping, and predictive tools that facilitate the selection of materials with high productivity and enhanced phytochemical quality. Recent reviews emphasize that integrating multi-omic and advanced analytical approaches is critical for accelerating breeding programs and guiding their commercial deployment, while also creating opportunities in sectors such as functional foods and phytochemicals (3, 4). Collectively, these developments position multivariate phenotyping and machine learning at the core of applied research on species with high commercial value.

In the national context, Peru has consolidated its position as a major player in the global *Capsicum* market, with sustained increases in both export volume and value in recent years. The strengthening of value chains, growing international demand for differentiated products (such as paprika, *piquillo* peppers, and dried *Capsicum*), and the wide diversity of native varieties create a strategic opportunity to enhance the value of Peruvian germplasm through applied research aimed at identifying ecotypes with high yield potential and desirable commercial attributes (5). Multivariate characterization of native ecotypes directly supports value-added policies, genetic improvement programmes, and *in-situ* and *ex-situ* conservation strategies. In addition, the use of liquid biofertilizers derived from cocoa shells has shown positive effects on the growth of *Capsicum chinense*, promoting sustainable agricultural practices and highlighting its potential applicability in the Amazon region (6).

At the local level, in the Loreto region and the Peruvian Amazon, native varieties such as *charapita*, along with other ecotypes of *Capsicum chinense* and *C. frutescens*, constitute an agrobiodiversity heritage of high cultural value and significant productive potential. Regional studies and technical reports have documented the phenotypic variability and socio-economic relevance of these chillies in local markets and short supply chains, as well as their traditional uses and potential for the development of differentiated, value-added products (7). However, integrated studies that combine morphological and productive characterization with predictive modelling in native ecotypes of the Peruvian Amazon remain scarce, limiting the systematic identification of key traits that are truly decisive for selection and production scaling.

The multivariate analyses employed in this study are supported by the approach described by Mishra et al. (8), who applied multivariate statistical methods and machine learning techniques, including Random Forest, to assess genetic diversity, trait associations, and genotypic differentiation in guava (*Psidium guajava* L.). Long et al. (9) conducted a multivariate analysis of yield in ten agronomic traits across 59 maize hybrids to investigate yield performance, providing a theoretical foundation for the development of high-yielding varieties adapted to mountainous regions. Similarly, Tlahig et al. (10) applied multivariate statistical approaches to classify alfalfa genotypes into distinct yield groups aligned with specific breeding objectives, such as seed production optimisation and biomass yield enhancement. Sameway, Al Galib et al. (11) employed multivariate analysis to identify high-yielding rice cultivars based on seed yield and morphological traits, revealing substantial genetic diversity among the evaluated cultivars that could be harnessed in future breeding programs for varietal rice improvement. Collectively, these studies highlight the value of multivariate statistical tools in plant breeding. By integrating information from multiple correlated traits, breeders can make more informed selection decisions, facilitating the development of stable, high-yielding genotypes adapted to diverse environmental conditions. This approach not only enhances the efficiency of breeding programs but also supports the long-term sustainability and productivity of agricultural systems (12).

This study directly addresses this knowledge gap by integrating robust descriptive statistics, principal component analysis (PCA), clustering techniques, mixed-effects models, and machine learning approaches, such as Random Forest (RF) and linear discriminant analysis (LDA) to: (i) characterize phenotypic variability, (ii) identify traits that best predict yield performance, and (iii) propose selection criteria applicable to breeding programmes and regional commercial valorization strategies. The expected contribution is twofold: first, to generate reproducible, methodologically robust scientific evidence to support selection and genetic improvement efforts; and second, to provide technical inputs to strengthen local value chains in Loreto, including varietal improvement, differentiated product development, and agrobiodiversity conservation.

2 Materials and methods

2.1 Study area

The study was conducted at the San Ramón Agrarian Experimental Station of the National Institute of Agrarian Innovation (INIA), located in the district of Yurimaguas, province of Alto Amazonas, Loreto region, Peru. The site is situated within a humid tropical ecosystem, characterized by mean annual temperatures of 27–28 °C, relative humidity exceeding 80%, and annual precipitation above 2,500 mm (13). Previous studies have shown that humid tropical environments strongly influence phenotypic expression in horticultural crops, underscoring the importance of integrating environmental variables to improve agronomic interpretation (14).

2.2 Plant material

Twelve native ecotypes of *Capsicum chinense* from different Amazonian localities (Yurimaguas, Jeberos, Balsapuerto, Lagunas, Santa Cruz, and Teniente César López Rojas), in the province of Alto Amazonas, Loreto region, Peru, were evaluated (Figure 1, Table 1). Ecotype selection was based on observed morphological variability, agricultural relevance, and geographical representativeness. The characterization of native germplasm is essential for genetic improvement programmes and for elucidating the phenotypic structure and diversity of *Capsicum* species (3, 15).

2.3 Soil physico-chemical analysis

The physico-chemical analyses of the soil samples were conducted within the framework of the "Peru 2M: Know the Fertility of Your Soil" campaign at the Soil Laboratory of the San Ramón Agrarian Experimental Station and the Soil Laboratory of the El Porvenir Agrarian Experimental Station (Table 2). Soil samples were collected at 0–30 cm depth, air-dried, and sieved through a 2 mm mesh prior to analysis, as indicated by Quispe Huincho (16). Soil pH, electrical conductivity (EC), organic matter (OM) content, available phosphorus (P), exchangeable bases, texture, and effective cation exchange capacity (ECEC) were determined following the methodologies described by the Secretariat of Environment and Natural Resources (17). The interpretation of soil physico-chemical properties is essential for explaining productive and morphological differences in horticultural crops (18–20).

2.4 Experimental design

A completely randomized block design (CRBD) was employed, with 12 treatments and three replicates. The experimental population comprised native *Capsicum chinense* germplasm from

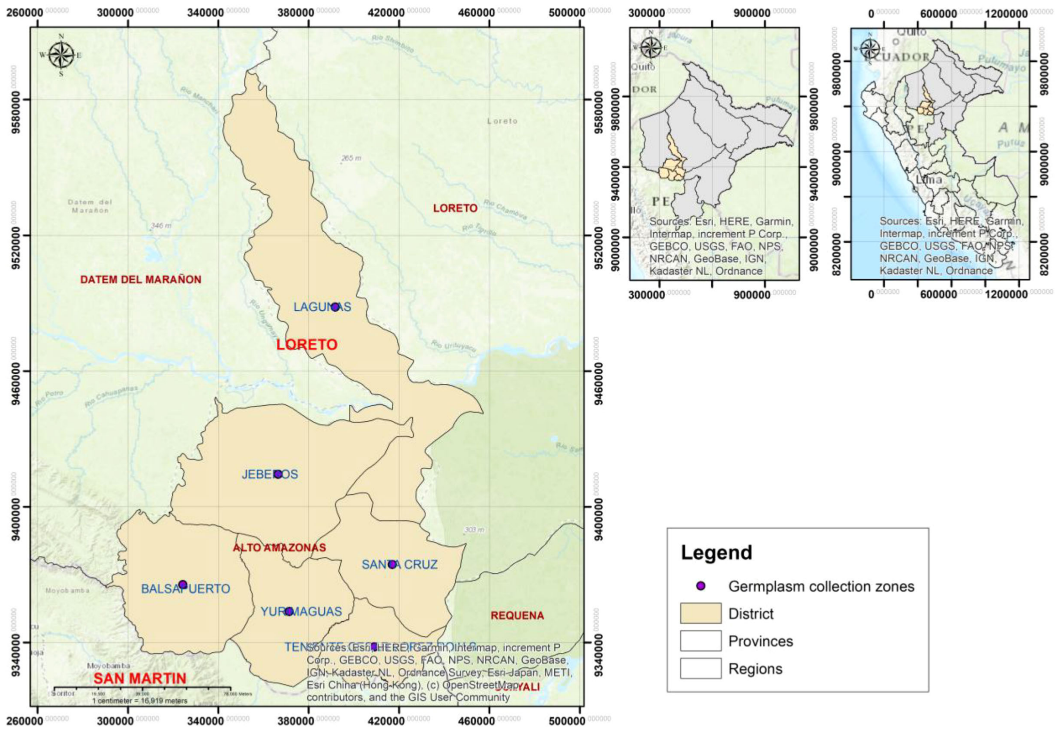

FIGURE 1
Location of the collection area of *Capsicum chinense* ecotypes.

TABLE 1 Origin and georeferencing of 12 Aji Charapita (*Capsicum chinense*) ecotypes studied in Alto Amazonas, Loreto.

N°	Ecotypes	Origin	Latitude (S)	Longitude (O)
1	YUR008	Yurimaguas	6°02'05,1"	76°15'23,2"
2	JEB027	Jeberos	5°18'07,1"	76°17'00,2"
3	TNTE CL019	Teniente César López Rojas	5°57'24,6"	75°56'44,1"
4	LAG022	Lagunas	5°13'54,0"	75°40'02,1"
5	SC024	Santa Cruz	5°30'51,5"	75°51'30,2"
6	LAG021	Lagunas	5°18'30,6"	75°43'22,1"
7	TNTE CL017	Teniente César López Rojas	6°01'23,1"	75°52'23,2"
8	YUR001	Yurimaguas	5°48'50,0"	76°07'26,9"
9	BAL012	Balsapuerto	5°50'03,2"	76°32'54,7"
10	YUR007	Yurimaguas	6°02'05,3"	76°15'21,8"
11	JEB028	Jeberos	5°16'34,4"	76°16'43,2"
12	JEB026	Jeberos	5°17'20,9"	76°16'54,9"

the Peruvian Amazon, while the sample included 12 representative ecotypes selected based on their phenotypic variability and local agricultural availability.

2.5 Identification of variables

Morphological, productive, and seed-related variables were selected following internationally recognized phenotyping standards and descriptor lists for *Capsicum* (3, 15). These variables included plant height, canopy and stem diameter, fruit length and width, pericarp thickness, number of fruits per plant, fruit weight, and seed weight (Table 3). Measurements were obtained using a measuring tape, digital calliper, and precision balance. The use of these instruments has been validated for phenotyping fruit, seed, and morphological traits in *Capsicum* species (21, 22). Observations were recorded over three consecutive harvests at 20-day intervals.

2.6 Crop management

Seeds were sown in a sterilized organic substrate composed of soil and organic matter at a 1:2 ratio. Light irrigation and preventive management practices were applied to ensure uniform seedling emergence and growth (23). Transplanting was carried out 70 days after sowing, using a spacing of 0.80 m × 0.60 m. Weed control, supplementary irrigation, and phytosanitary monitoring were conducted throughout the crop cycle. Similar management protocols have been reported in recent studies on *Capsicum* production systems (24).

TABLE 2 Results of soil physicochemical analysis.

Characteristics		Characteristics	
Chemical analysis:	Value	Exchangeable bases:	Value
pH (1:1)	4,2	Calcium (cmolc·kg^{-1})	0,71
Electrical conductivity	5,2	Magnesio (cmolc·kg^{-1})	0,2
Organic matter (%)	1,3	Sodium (cmolc·kg^{-1})	0,12
Available phosphorus (ppm)	4,5	Potassium (cmolc·kg^{-1})	<0,10
Available potassium (ppm)	17,2	Texture:	
Total nitrogen(%)	0,07	Sand (%)	63,26
ECEC	2,49	Clay (%)	13,03
		Silt (%)	23,71

2.7 Statistical analysis

Prior to statistical analysis, extreme values, data consistency, variable distributions, and normalization were assessed. Data normality was evaluated using the Shapiro–Wilk test, in accordance with recent guidelines for multivariate analyses in agronomic research (18). Descriptive statistics, including mean, median,

TABLE 3 Variables evaluated and their identification codes.

ID	Variables evaluated	CODE
1	Yield (kg·ha^{-1})	YIELD.kg.ha
2	Days to emergence	DY.E
3	Days to 50% flowering	DY.FLO.50
4	Days to 50% fruiting	DY.FRU.50
5	Number of fruits per plant	N.FRU.P
6	Weight of 10 fruits (g)	W.10.FRU (g)
7	Fruit weight per plant (g)	P.FRU.W/P (g)
8	Fruit diameter (cm)	FRU.D (mm)
9	Fruit length (cm)	FRU.L (mm)
10	Fruit wall thickness (mm)	FRU.W.T (mm)
11	100-seed weight (g)	X100.S.W (g)
12	Seed diameter (mm)	S.D (mm)
13	Seed length (mm)	S.L (mm)
14	Plant height (cm)	P.H (cm)
15	Canopy diameter (cm)	C.D (cm)
16	Stem diameter (cm)	ST.D (mm)

standard deviation, standard error, coefficient of variation, minimum, and maximum values, were calculated, as descriptive analyses are essential for characterizing the distribution and variability of plant germplasm (22). Linear mixed-effects models were fitted using the lme4 and lmerTest packages, considering ecotype as a fixed effect and block as a random effect (18). Pearson's correlation analysis was applied to explore relationships between phenotypical and productive variables.

Principal component analysis (PCA) was used to reduce dimensionality and to identify phenotypic patterns among ecotypes (3, 25). Both k-means and hierarchical clustering methods were applied, and the optimal number of clusters was determined using multiple indices implemented in the NbClust package. Multivariate clustering approaches are widely used for classifying and differentiating agricultural germplasm (19). Random Forest (RF) models were employed to identify trait relative importance and predict yield performance, as this method has demonstrated high accuracy in agricultural applications (1, 26–28). Linear discriminant analysis (LDA) was used to validate the clusters obtained, confirming the phenotypic structure among ecotypes (24). A dendrogram was generated to visualize multivariate similarity among ecotypes and to identify consistent phenotypic groups based on morphological and productive characteristics. This method provides a hierarchical representation of how materials are grouped by their statistical distance, thereby facilitating selection and comparison among ecotypes (29).

3 Results

3.1 Descriptive statistics

Descriptive statistics revealed two distinct patterns of variation under acidic soil conditions: (a) relatively stable phenology and (b) highly heterogeneous productivity. Regarding phenology, DY.E, DY.FLO.50, and DY.FRU.50 exhibited moderate to low dispersion and distributions close to symmetry (skewness near zero or moderate; negative kurtosis), suggesting that temporal variation in the crop cycle is limited within this environment. In contrast, YIELD (M = 76.51, SD = 65.65; skewness = 1.429; kurtosis = 1.861) and several production variables (e.g., W.10.FRU with skewness = 2.067 and kurtosis = 5.822; P.FRU.W/P with skewness = 1.423 and kurtosis = 1.876) displayed pronounced right-skewed distributions and extreme values. Statistically, this pattern is consistent with substantial heterogeneity among ecotypes and with high residual variability expected under edaphic stress. These distributional characteristics anticipate (i) potential overlap among means in multiple comparisons when standard errors are large and (ii) greater suitability of non-parametric or robust approaches for assessing bivariate relationships (Table 4).

3.2 Results of morphological variables in *Capsicum chinense*

Comparative analysis using box plots revealed statistically significant differences among ecotypes for productive, morphological, and phenological variables. Days to seedling emergence exhibited a distribution concentrated around 18 days (median ≈ 18), with low dispersion across most ecotypes, indicating uniformity in early developmental stages (Figure 2a). Plant height (cm) showed marked variability among ecotypes, with JEB-028 and TNTECL-017 reaching greater heights (median ≈ 40 cm), whereas YUR-001 and BAL-012 exhibited shorter stature (median ≈ 20 cm), as shown in Figure 2b. Phenological variables, including days to 50% flowering (Figure 3a) and days to 50% fruiting (Figure 3b), revealed differences in developmental timing among ecotypes. JEB-028 and TNTECL-019 displayed later phenological events (median ≈ 102 days), while BAL-012 and YUR-007 reached flowering and fruiting earlier (median ≈ 98 days).

TABLE 4 **Descriptive analysis results of the evaluated variables.**

Vars	Mean	Sd	Median	Trimmed	Mad	Min	Max	Range	Skew	Kurtosis	Se
DY.E	17.08	1.05	16.84	17.07	1.24	15.33	19.00	3.67	0.09	-0.91	0.18
DY.FLO.50	100.08	3.53	100.67	100.14	4.69	94.00	105.67	11.67	-0.08	-1.26	0.59
DY.FRU.50	109.95	2.50	110.34	110.09	2.47	103.67	113.67	10.00	-0.56	-0.58	0.42
P.H.cm.	19.56	9.82	17.40	18.50	8.81	6.52	48.44	41.92	0.99	0.60	1.64
ST.D.mm.	5.15	2.24	5.12	5.03	2.79	1.72	10.15	8.43	0.26	-0.90	0.37
C.D.cm.	0.396	0.04	0.41	0.40	0.02	0.32	0.47	0.15	-0.28	-0.74	0.01
N.FRU.P	23.78	11.14	22.86	23.25	12.32	5.83	49.56	43.73	0.36	-0.78	1.86
P.FRU.W.P.g.	22.23	18.83	16.73	19.66	12.94	1.82	83.91	82.09	1.42	1.88	3.14
FRU.L.mm.	9.26	2.18	9.41	9.17	2.39	6.19	14.59	8.40	0.24	-0.87	0.36
FRU.D.mm.	8.60	1.35	8.33	8.68	1.26	5.840	10.36	4.52	-0.28	-0.97	0.23
FRU.W.T.mm.	1.25	0.09	1.27	1.25	0.10	1.08	1.46	0.38	0.11	-0.69	0.02
W.10.FRU.g.	8.25	6.01	6.45	7.51	3.58	1.25	33.21	31.96	2.07	5.82	1.00
S.L.mm.	3.17	0.27	3.18	3.16	0.19	2.55	4.28	1.73	1.56	6.41	0.05
S.D.mm.	2.68	0.17	2.66	2.68	0.22	2.36	3.00	0.64	-0.10	-1.19	0.03
X100.S.W.g.	0.45	0.11	0.45	0.45	0.10	0.26	0.63	0.37	0.12	-1.13	0.02
YIELD.kg.ha	76.51	65.65	52.04	67.48	48.39	6.33	291.35	285.02	1.43	1.86	10.94

10.3389/fsoil.2026.1786422

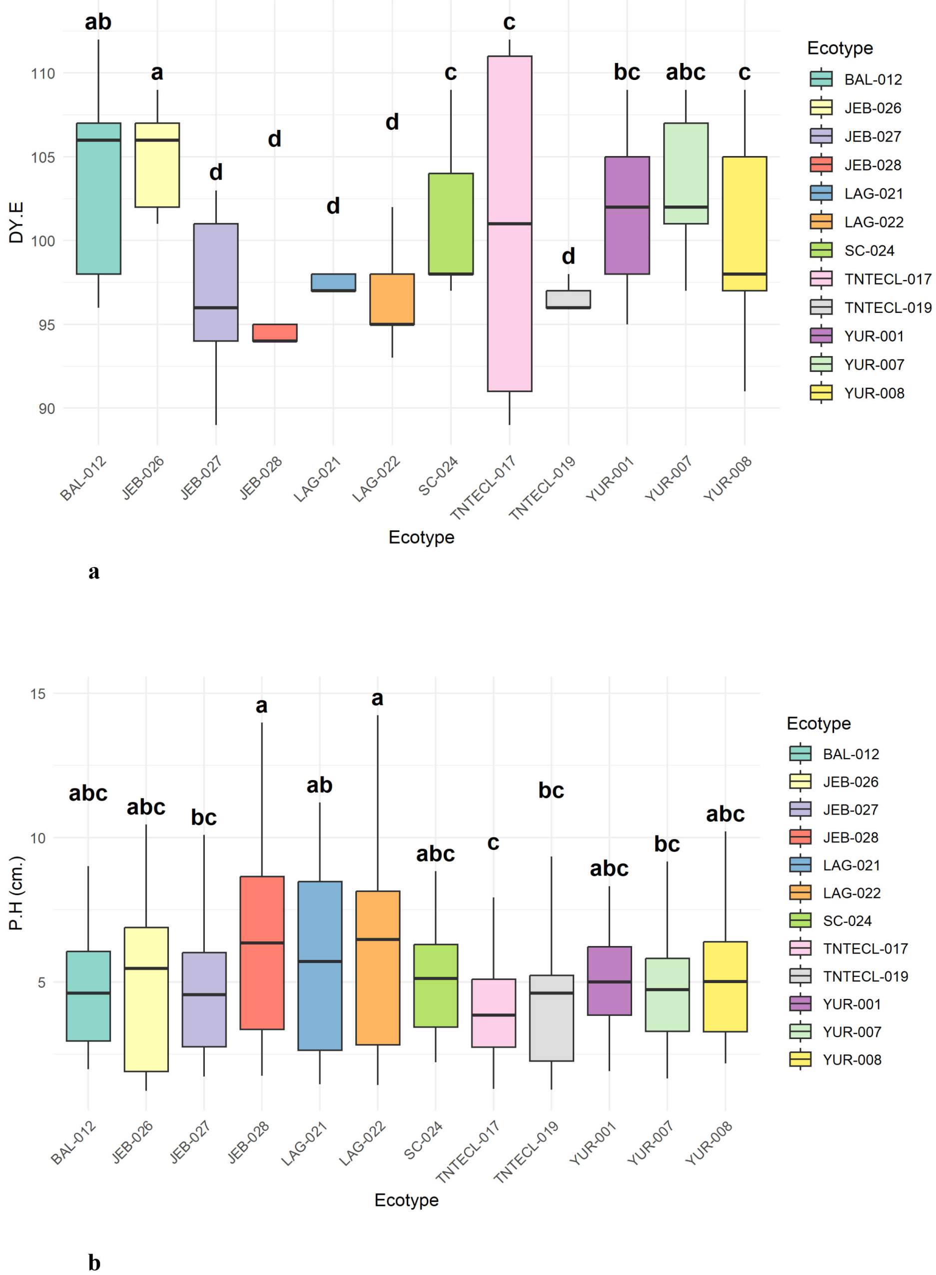

FIGURE 2
(a) Boxplot of days to seedling emergence in *Capsicum chinense* ecotypes. **(b)** Boxplot of days to plant height in *Capsicum chinense* ecotypes.

3.3 Yield results of *Capsicum chinense*

The number of fruits per plant exhibited high variability among ecotypes. TNTECL-017 and LAG-021 showed wide ranges (> 40 fruits), whereas BAL-012 and YUR-008 produced fewer fruits (median ≈ 20 fruits per plant) with greater uniformity (Figure 4a). Mean fruit weight per plant (g) was higher in JEB-028 and LAG-022 (median ≈ 11 g), with broad ranges indicating internal heterogeneity, while SC-024 and YUR-001 displayed lower values (median ≈ 7 g) and reduced dispersion (Figure 4b). Seed length (mm) showed relative stability across most ecotypes (median ≈ 2.8–3.0 mm); however, LAG-022 and SC-024 exhibited greater dispersion (IQR > 0.5 mm), possibly reflecting underlying genetic variability (Figure 5a). Regarding yield (kg·ha^{-1}), ecotypes JEB-028 and LAG-022 recorded the highest medians (≈ 320 and ≈ 300 kg·ha^{-1}, respectively), accompanied by wide interquartile ranges (IQR > 100 kg·ha^{-1}), suggesting high productive potential coupled with internal variability. In contrast, BAL-012 and YUR-001 showed lower median yields (< 200 kg·ha^{-1}) and reduced dispersion, indicating greater stability but limited productive performance (Figure 5b).

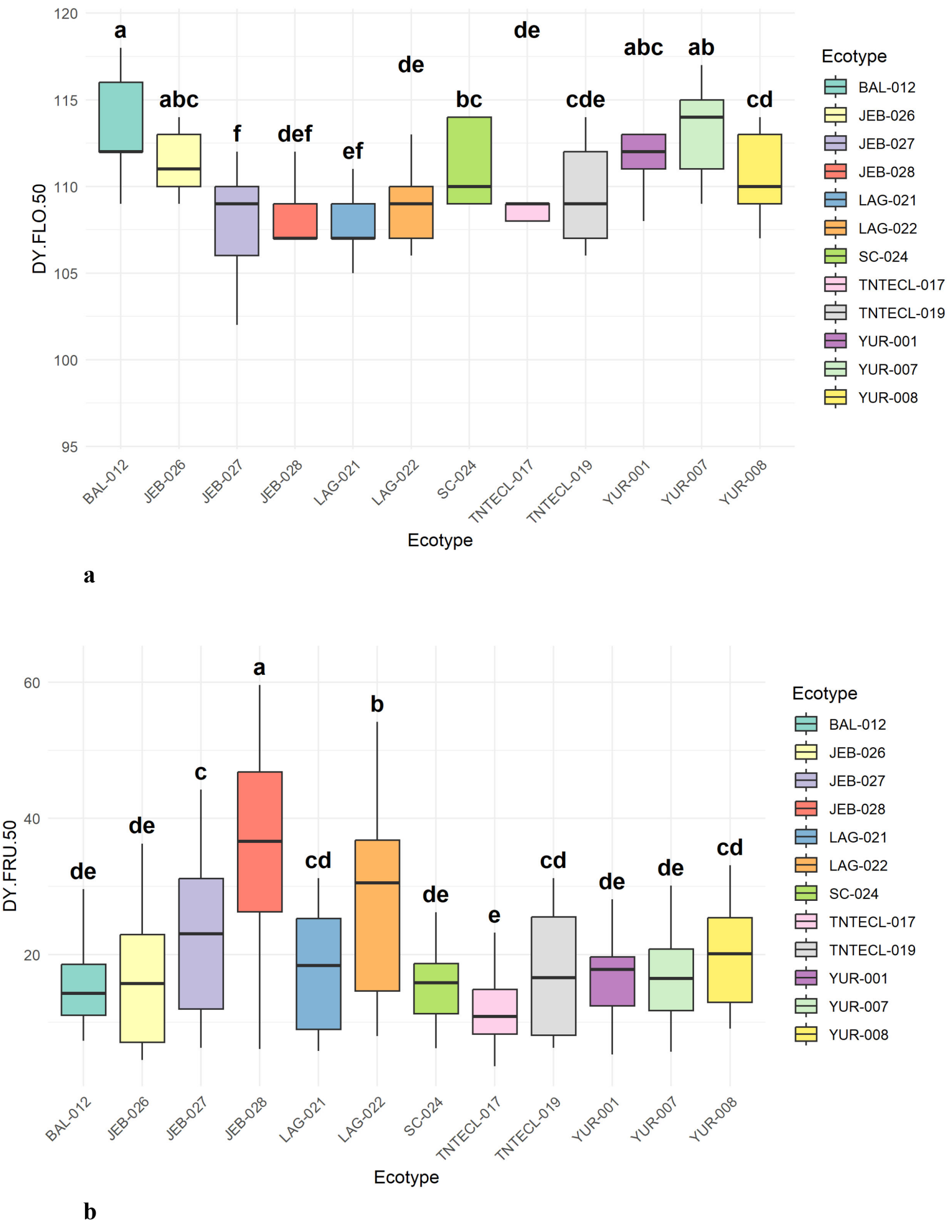

FIGURE 3
(a) Boxplot of days to 50% flowering in *Capsicum chinense* ecotypes. **(b)** Boxplot of days to days to 50% fruiting in *Capsicum chinense* ecotypes.

3.4 Correlation analysis between morphological variables and yield in *Capsicum chinense*

Correlation analysis among agronomic variables revealed significant linear associations that explain the structural, functional, and productive relationships among the evaluated ecotypes. The Pearson correlation matrix, visualized as a heat map, showed coefficients ranging from moderate negative values (r = −0.40) to strong positive values (r > 0.80), indicating both inverse and highly positive relationships. Regarding productivity, yield (kg·ha^{-1}) exhibited significant positive correlations with weight of 10 fruits (g) (r = 0.85), fruit weight per plant (g) (r = 0.82), and number of fruits per plant (r = 0.78), indicating that yield is primarily driven by cumulative fruit weight, average weight per fruit, and fruit number per plant. These relationships suggest that yield performance can be effectively explained by measurable reproductive components, making them highly relevant for direct selection strategies in breeding programmes. Fruit morphological variables also displayed strong internal correlations. Fruit diameter (mm) was positively associated with fruit length (mm) (r = 0.88) and fruit wall thickness (mm) (r = 0.81), indicating coordinated expression of fruit size across longitudinal, transverse, and thickness dimensions. Likewise, fruit weight per plant (g) showed strong correlations with fruit diameter (mm) (r = 0.79) and fruit length (mm) (r = 0.76), suggesting that individual fruit weight is primarily determined by structural fruit volume (Figure 6).

Regarding seed-related variables, 100-seed weight (g) showed moderate positive correlations with seed diameter (mm) (r = 0.65) and seed length (mm) (r = 0.62), indicating that seed mass is

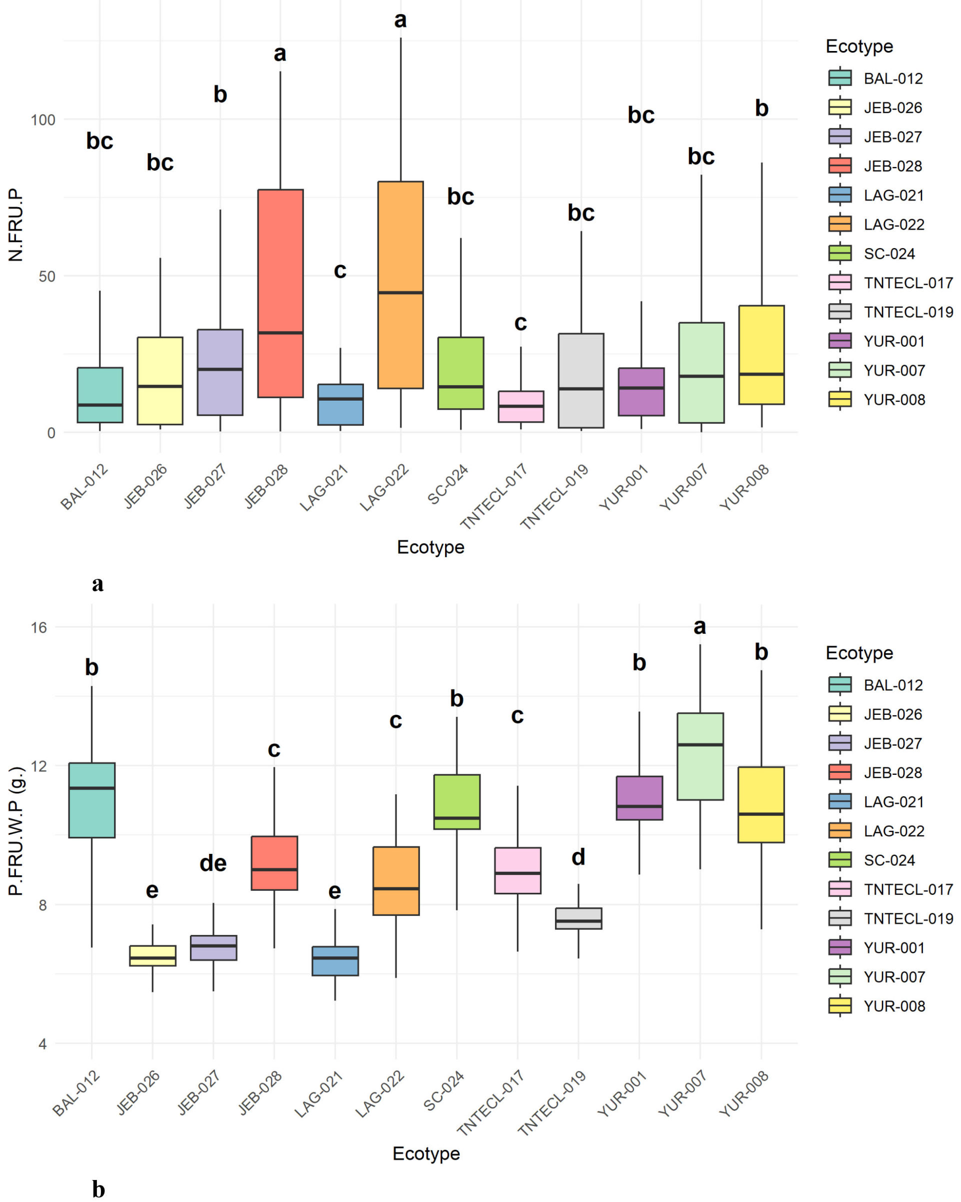

FIGURE 4
(a) Boxplot of number of fruits per plant in *Capsicum chinense* ecotypes. **(b)** Boxplot of the average fruit weight per plant in *Capsicum chinense* ecotypes.

influenced by seed size. However, these associations were weaker than those observed among fruit-related variables, suggesting a lower degree of structural dependence in seed characteristics. Phenological variables exhibited negative correlations with yield. Specifically, days to 50% fruiting and days to 50% flowering were negatively associated with yield (kg·ha^{-1}) (r = −0.42 and r = −0.38, respectively), indicating that ecotypes with later flowering and fruiting tended to have lower yield levels. This inverse relationship suggests that earliness may represent a favourable attribute under intensive production conditions. Finally, days to seedling emergence showed low correlations with the remaining variables (r < 0.30), indicating that early development is mainly independent of the morphological and reproductive traits evaluated. This finding suggests that days to emergence may be more strongly influenced by physiological seed traits or environmental conditions rather than by the structural attributes measured in this study.

3.5 Correlation analysis between morphological variables and yield in *Capsicum chinense*

The Spearman heat map indicated that yield (YIELD, kg·ha^{-1}) was primarily associated with variables reflecting direct productivity and plant vigor, showing very strong positive correlations: P.FRU.W/P (ρ = 0.99, p < 0.001), N.FRU.P (ρ = 0.91, p < 0.001), W.10.FRU (ρ = 0.90, p < 0.001), P.H (ρ = 0.90, p < 0.001), and ST.D (ρ = 0.88, p < 0.001). This pattern suggests a dominant vigor–load–commercial biomass axis, whereby more vigorous plants support a greater fruit load and/or heavier fruits, ultimately increasing YIELD

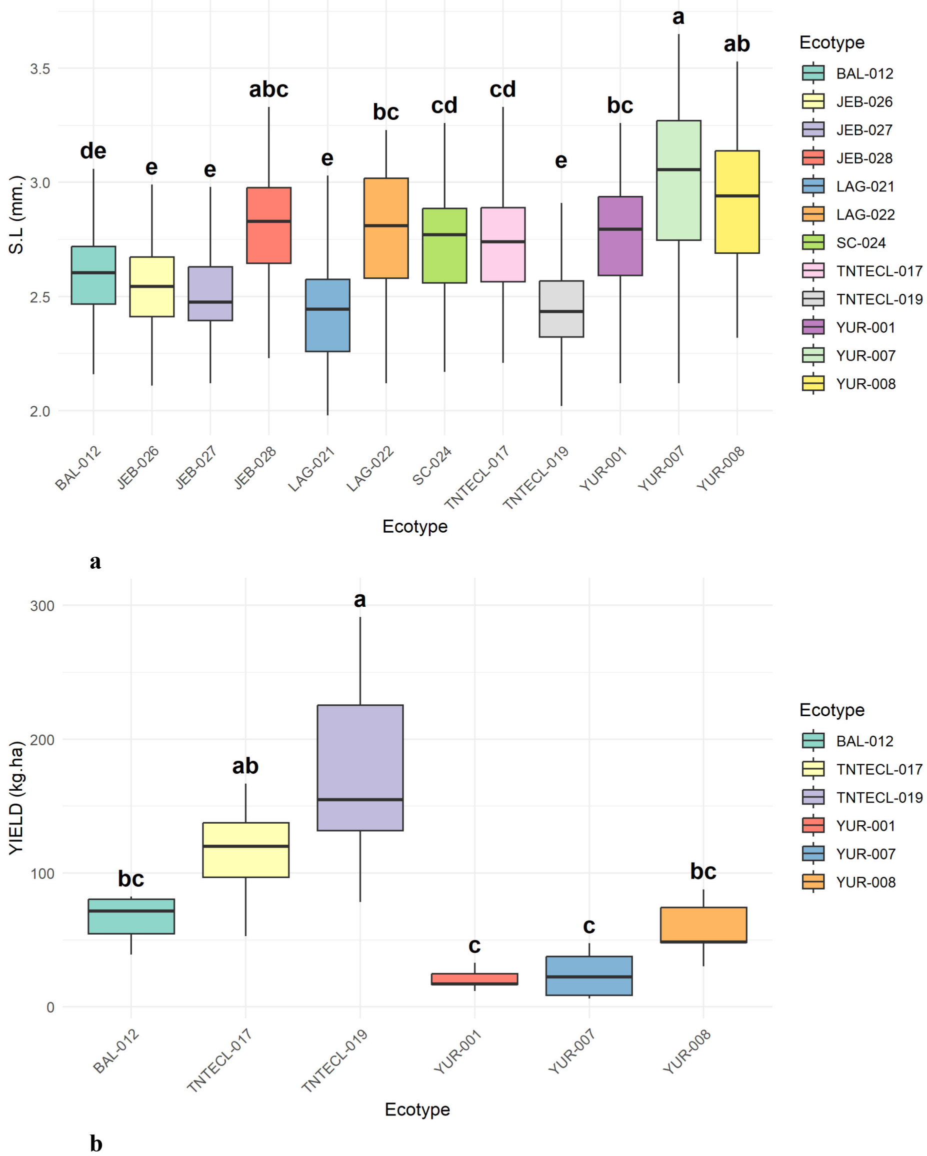

FIGURE 5
(a) Boxplot of seed length in *Capsicum chinense* ecotypes. **(b)** Boxplot of yield in *Capsicum chinense* ecotypes.

(Figure 7). In contrast, YIELD exhibited weak negative associations with phenological variables (DY.E and DY.FLO.50, $\rho \approx -0.14$; no evidence of statistical significance in the figure) and low correlations with fruit and seed morphometric traits (e.g., FRU.L, FRU.D, 100.S.W; small $|\rho|$ values without significance indicators). These results suggest that, within this dataset, yield is driven primarily by fruit number and plant vigour rather than by systematic increases in fruit or seed size.

3.6 Principal component analysis

Principal component analysis (PCA) showed that PC1 explained 33.2% of the variance and PC2 explained 23.9%, accounting for 57.1% cumulatively. These components captured two dominant axes consistent with the univariate and bivariate patterns. PC1 (productive–vigour gradient): The vectors for YIELD, P.FRU.W/P, N.FRU.P, P.H, and ST.D were closely aligned, reflecting the same covariation block identified in the Spearman analysis (high correlations). This alignment explains why these traits, although noisy at the univariate level, define a stable multivariate axis associated with plant productivity and vigour. PC2 (fruit–seed morphometry gradient): Variables such as FRU.L, FRU.D, S.D, and 100.S.W contributed strongly to this component. This pattern is consistent with the Tukey test results (which showed separation among ecotypes) and with the descriptive analysis, which indicated moderate variation and lower distortion from heavy-tailed distributions. Overall, the PCA integrates the evidence into two main dimensions: (i) a performance-driven axis, where YIELD is statistically anchored to plant vigour and fruit load, and (ii) a reproductive structure axis (fruit and seed traits) that more clearly discriminates among ecotypes (Figure 8).

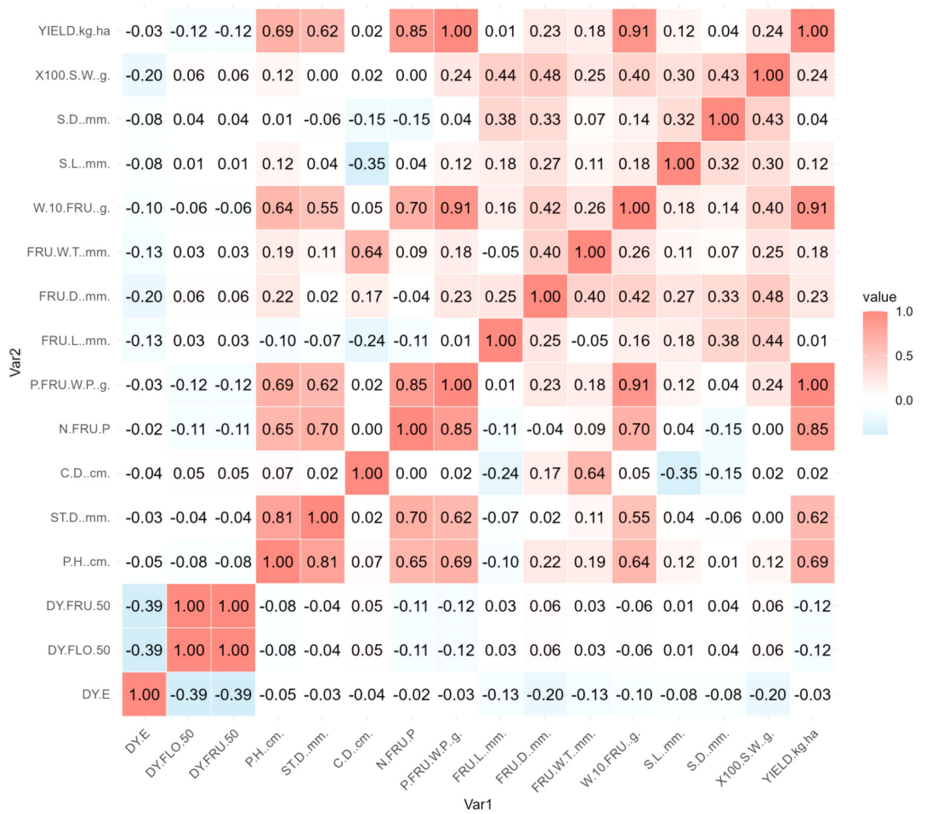

FIGURE 6
Heat map of correlations among variables evaluated in *Capsicum chinense* ecotypes.

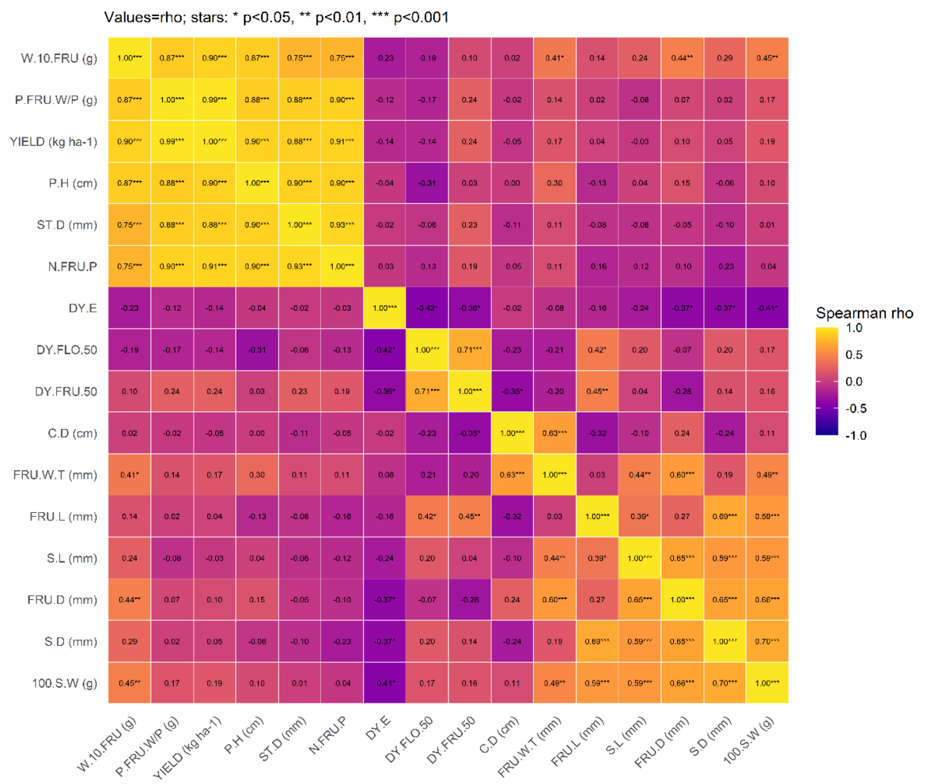

FIGURE 7
Spearman correlation heamap of variables evaluated in *Capsicum chinense* ecotypes.

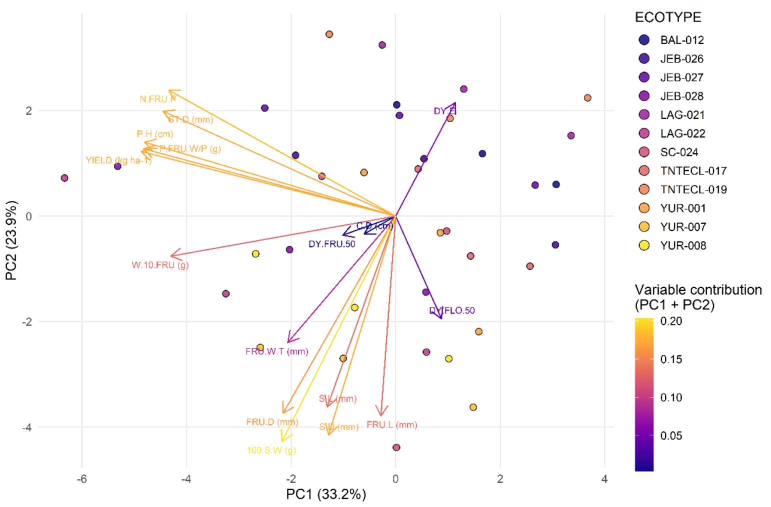

FIGURE 8
PCA biplot of variables evaluated in *Capsicum chinense* ecotypes.

3.7 Cluster analysis

K-means clustering applied to ecotype centroids on the PC1–PC2 plane (explaining 57.1% of the total variance) partitioned the ecotypes into three multivariate profiles, consistent with the PCA axes and the univariate evidence. Cluster 2: LAG-022 and JEB-028 (extreme negative PC1). This group corresponds to a profile closer to the productive axis. The pattern is consistent with the Tukey results, in which LAG-022 ranked among the highest values for yield and related metrics, and with the Spearman correlations linking yield to fruit load and plant vigour. Cluster 1: YUR-007, YUR-008, SC-024, and YUR-001 (negative PC2). This group aligns with the fruit–seed morphometry gradient and Tukey separations observed in structural traits, suggesting that these ecotypes are grouped primarily by reproductive traits rather than yield performance. Cluster 3: TNTECL-019, LAG-021, BAL-012, JEB-026, JEB-027, and TNTECL-017 (positive PC1 and moderate to high PC2), as shown in Figure 9. This cluster comprises ecotypes with relatively closer multivariate profiles in the PCA plane, consistent with the overlaps anticipated from the descriptive statistics and the partial separation observed in the Tukey analysis.

3.8 Linear discriminant analysis

The LD1–LD2 plot was interpreted as an exploratory multivariate similarity map. LD1 represented the primary axis of separation, distinguishing a group of LD1-negative individuals (e.g., YUR-007 and YUR-008) from a cluster of LD1-positive individuals (e.g., JEB-026, JEB-027, TNTECL-019, and LAG-021). This pattern suggests a high degree of internal similarity and potential overlap among the ecotypes located within the LD1-positive cluster. LD2 functioned as a complementary axis, highlighting BAL-012 as the most distinct ecotype (with high LD2 values) and positioning SC-024 and YUR-001 toward the positive LD2 region, which was separated from the right-hand cluster (Figure 10).

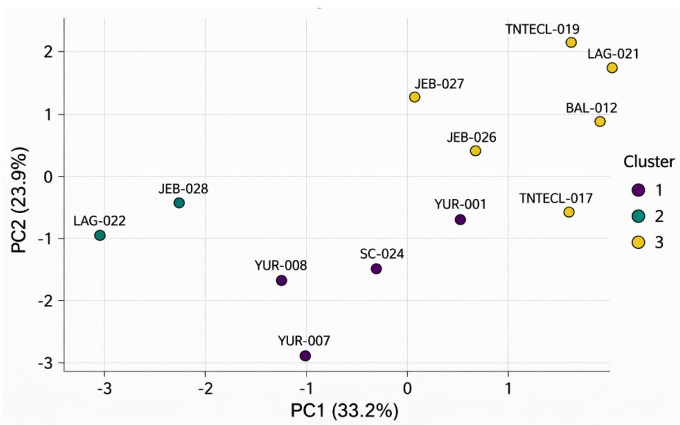

FIGURE 9
K-means Clustering of PCA centroids evaluated in *Capsicum chinense* ecotypes.

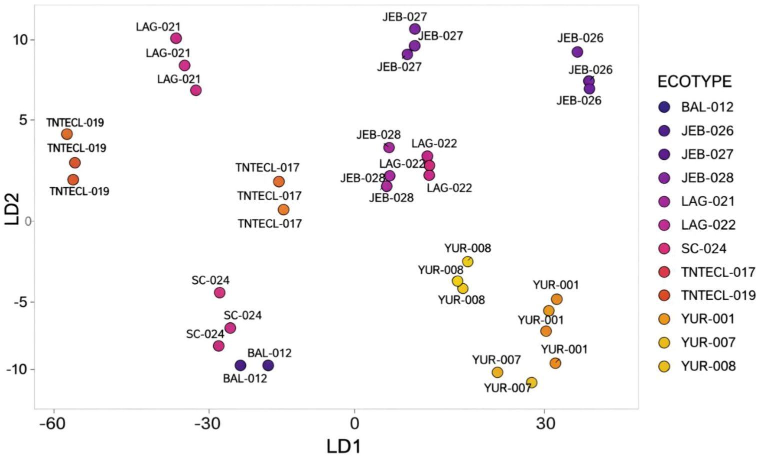

FIGURE 10
PCA-LDA: LD1 vs LD2 evaluated in *Capsicum chinense* ecotypes.

3.9 Dendrogram

The dendrogram generated using Ward's method (Ward.D2) was consistent with the PCA and k-means results, separating the ecotypes into a small group of high-performance and a broader pool of intermediate or mixed profiles under acidic soil conditions. Cluster 1 (JEB-028, LAG-022): This cluster forms a very compact group associated with high yield. The pattern is consistent with the Tukey results, in which LAG-022 ranked within the highest YIELD group, and JEB-028 showed high to intermediate performance. It also aligns with the Spearman correlation structure, where YIELD is primarily associated with the vigour–load axis (P.FRU.W/P, N.FRU.P, W.10.FRU, P.H, and ST.D). Cluster 2 (LAG-021, JEB-027, TNTECL-019): This compact group exhibits a profile that is more morphophenological than yield-extreme. For example, LAG-021 showed a lower S.D., whereas TNTECL-019 displayed a longer phenological duration and a lower 100.S.W. These patterns suggest a stable combination of traits distinct from that of the high-yield cluster. Cluster 3 (SC-024, YUR-007, YUR-008, BAL-012, YUR-001, JEB-026, TNTECL-017): This larger, more heterogeneous cluster exhibits internal substructure. It includes ecotypes with intermediate or contrasting profiles depending on whether fruit and seed traits (e.g., YUR-007, YUR-008) or earliness and plant architecture traits (e.g., BAL-012, YUR-001) are more prominent (Figure 11).

3.10 Random forest

The Random Forest (RF) analysis based on permutation importance indicated that multivariate differentiation among ecotypes under acidic soil conditions is primarily driven by structural fruit and seed traits. The highest importance scores were observed for FRU.L (0.0327), FRU.D (0.0245), and 100.S.W (0.0241), followed by C.D (0.0204) and FRU.W.T (0.0169). Intermediate contributions were associated with phenological variables (DY.FLO.50 = 0.0143; DY.FRU.50 = 0.0120) and S.D (0.0134), whereas W.10.FRU (~0.0013) and S.L (~0.0012) showed minimal contributions. In contrast, variables associated with the yield–vigor axis exhibited negative importance values (ST.D = −0.0076; P.FRU.W/P = −0.0053; P.H = −0.0050), a pattern consistent with redundancy or collinearity and with the high residual variability observed in these traits. Overall, the RF results converge with the Tukey and PCA analyses in indicating that the fruit–seed trait cluster provides more consistent discrimination among ecotypes. In contrast, traits closely linked to YIELD, although strongly correlated in the Spearman análisis, contribute

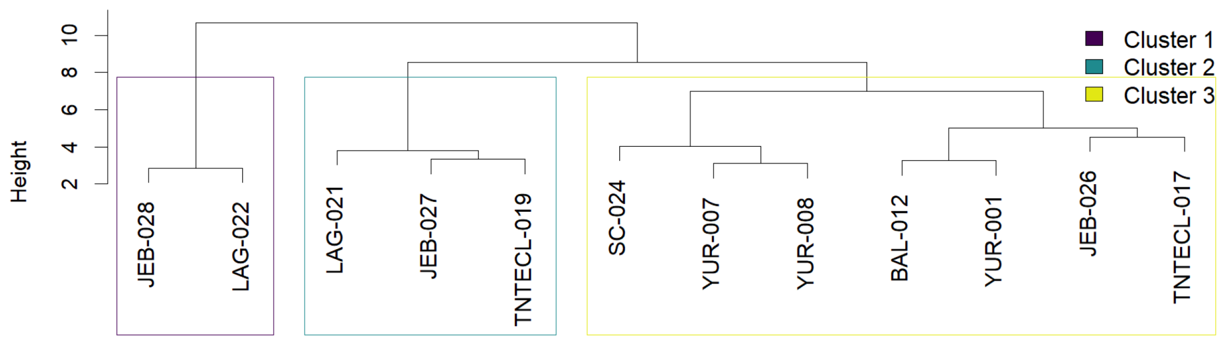

FIGURE 11
Dendrogram on standardized in *Capsicum chinense* ecotypes.

less incremental information within the multivariate framework (Figure 12).

4 Discussions

The results indicate that the morphological and productive traits of *Capsicum chinense* exhibit a well-defined multivariate structure, in which fruit-related attributes, such as individual fruit weight, cumulative fruit weight, and total yield, emerge as the main determinants of variation. This trend has been widely documented in recent studies, which demonstrate that reproductive traits are robust predictors of yield and should be prioritized in breeding programmes under modern approaches such as marker-assisted selection and multi-omic models (3). Moreover, the fact that a limited number of principal components accounted for more than 70% of the total variance confirms the robustness and coherence of the observed phenotypic structure. This finding is consistent with previous research reporting similar patterns in landraces and local populations, in which principal component analysis has proven effective at clearly differentiating genetic and functional groups based on agronomic performance (15).

The multimodal distributions observed in phenological traits, such as days to flowering and fruiting, suggest the presence of subpopulations with contrasting developmental strategies. This pattern has been reported in studies of intraspecific variability, where phenology acts as a key driver of morpho-productive divergence and adaptive differentiation (30). In addition, the negative relationship between precocity and yield reflects a physiological trade-off documented in yield prediction models, underscoring the importance of incorporating phenology as a critical variable in machine-learning-based agricultural predictions (1). The convergence of relevant variables identified through principal component analysis, correlation analysis, and Random Forest modelling further supports the robustness and internal consistency of the dataset. This methodological concordance highlights the value of integrating classical multivariate statistics with modern machine learning algorithms to enhance the reliability and precision of priority trait identification in crop improvement programmes (31).

The prominence of 100-seed weight and fruit size as predictors of yield is consistent with recent literature highlighting the role of early vigour and fruit morphometry as determinant factors in final productivity, particularly in tropical cropping systems where environmental conditions tend to amplify phenotypic differentiation (32, 33). These findings indicate that integrating seed- and fruit-related traits is an efficient strategy for accelerating the selection process in *Capsicum* breeding programmes. Moreover, integrating digital phenotyping methods and multisensor data fusion has the potential to further improve measurement accuracy and predictive performance, enabling the development of higher-resolution predictive models for yield forecasting and trait prioritization, as demonstrated by recent advances in high-throughput phenotyping platforms (34).

Regarding productivity and fruit morphology, PCA indicated that the first principal component (31.3%) accounted for the largest proportion of total variability, clearly separating high-yielding ecotypes (JEB-028 and LAG-022), which were associated with greater fruit diameter and length, from low-productivity ecotypes (BAL-012 and YUR-001). The separation of centroids and the amplitude of the confidence ellipses reflect underlying genotypic

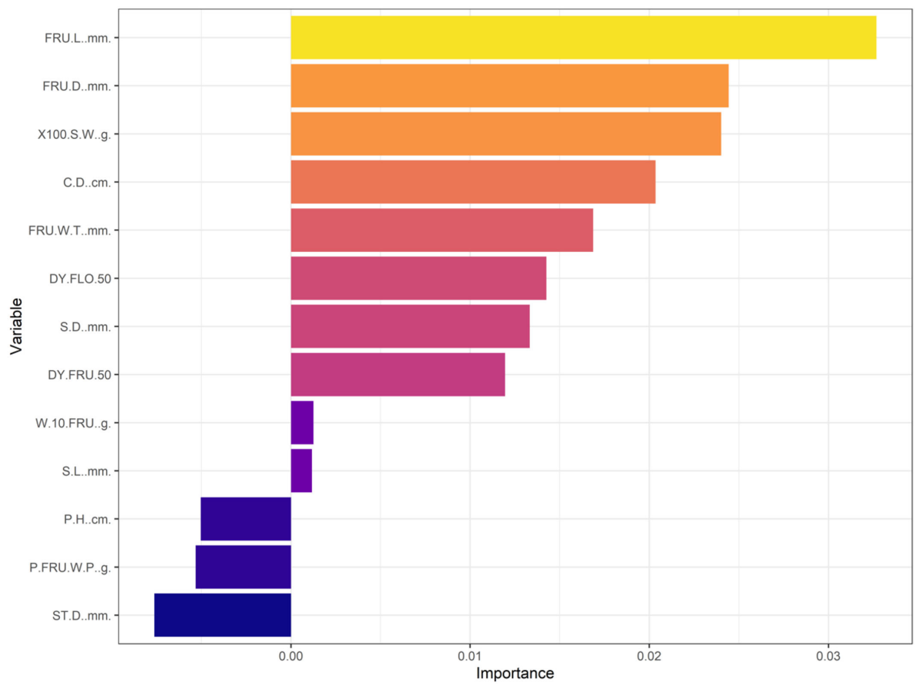

FIGURE 12
Random Forest plot showing the relative importance of agronomic variables in *Capsicum chinense* ecotypes.

diversity that can be effectively exploited in selection programmes. Similar studies have demonstrated that yield is strongly correlated with fruit morphological attributes and cumulative fruit weight, enabling efficient indirect selection strategies (35, 36). In terms of phenology and vegetative vigour, the second principal component (17.0%) captured phenological and vegetative variability, differentiating late- and vigorous-maturing ecotypes (TNTECL-019 and SC-024) from earlier-maturing ecotypes. This pattern confirms the central role of phenology in agroecological adaptation and yield stability. Recent research in peppers and other crop species has shown that PCA is an effective tool for identifying phenotypic diversity and population structure, highlighting phenology as a critical axis in varietal characterization (15).

Hierarchical cluster analysis, supported by a high agglomeration coefficient, revealed well-defined phenotypic groupings, providing a robust framework for developing varietal recommendations tailored to contrasting management conditions. This finding is consistent with previous studies that have successfully used dendrograms and hierarchical clustering to identify agronomically relevant groups in *Capsicum* germplasm collections (33). The high accuracy of linear discriminant analysis for classifying individuals further reinforces the internal consistency of these groups; however, the slight overlap observed among certain clusters highlights the influence of environmental factors and underscores the need to validate these patterns across multiple environments and crop cycles (1). The high coefficients of variation observed for several productive traits not only indicate significant potential for genetic improvement but also reveal opportunities for optimizing agronomic management practices. In this context, approaches based on explainable artificial intelligence (XAI) have emerged as valuable tools for understanding how specific factors influence crop yield performance and for guiding field practices using quantitative evidence (37). In parallel, the incorporation of single-nucleotide polymorphism (SNP) markers generated through genotyping-by-target sequencing (GT-seq) in *Capsicum* represents a promising alternative for integrating genetic variability into predictive models, thereby accelerating the selection of superior lines through combined phenotypic and genomic information (32).

The morphological and productive variability observed among native ecotypes of *Capsicum chinense* was closely associated with the soil conditions at the experimental site, characterized by acidic, low-fertility tropical soils. Soil acidity restricts the availability of essential nutrients and negatively affects root development, thereby conditioning crop yield performance and favouring only ecotypes with higher adaptive capacity. Previous studies have demonstrated that, under acidic soil conditions, soil–plant interactions and tolerance to chemical stress, particularly aluminium toxicity, are critical determinants of phenotypic expression and crop productivity (38, 39). In this context, the present results underscore the importance of incorporating edaphic diagnostics into selection strategies and sustainable management practices for *Capsicum chinense* in Amazonian agroecosystems.

One limitation of the present study is the absence of molecular marker analyses to identify significant associations among multiple traits related to yield performance under drought stress. The use of molecular markers could provide an innovative approach to enhancing drought tolerance in *Capsicum chinense*. Molecular analyses may also reveal a high level of genetic diversity within native *C. chinense* germplasm, which could be exploited in future breeding programs for varietal improvement. Furthermore, molecular tools highlight the potential to accelerate genetic improvement efforts and facilitate the development of stress-tolerant *C. chinense* cultivars, thereby contributing to sustainable fruit production under diverse climatic conditions throughout the year. Another important consideration for future research is the maintenance of soil quality, particularly regarding soil acidity management. Effective control of soil acidity requires an integrated strategy that combines the application of organic amendments, liming, balanced inorganic fertilisation, and beneficial microorganisms such as arbuscular mycorrhizal fungi and *Trichoderma* spp., among others (40, 41).

5 Conclusions

This study demonstrates that native *Capsicum chinense* ecotypes from the Peruvian Amazon exhibit a clearly differentiated phenotypic structure, characterized by high variability in morphological, productive, and phenological traits. The integration of multivariate analyses, mixed-effects models, and machine-learning algorithms enabled a robust and reproducible characterization of the evaluated germplasm. Ecotypes such as JEB-028 and LAG-022 emerged as elite materials, consistently exhibiting high yield and superior fruit morphology. In contrast, BAL-012 and YUR-001 showed lower productivity profiles but high phenotypic stability, positioning them as valuable genetic resources for breeding strategies focused on resilience, adaptation, or exploratory hybridization. Overall, the results confirm that an integrative framework combining classical statistical approaches, multivariate analyses, machine learning techniques, and discriminant validation is highly effective for phenotypic characterization of *Capsicum chinense*. Moreover, this study underscores the strategic value of Amazonian germplasm as a key resource for developing high-yielding varieties, differentiated products, and agrobiodiversity conservation strategies, thereby contributing to sustainable agricultural innovation in tropical environments. The results indicate that the physicochemical properties of the evaluated Amazonian soil, characterized by high acidity, low organic matter content, and limited availability of exchangeable bases, significantly condition the morphological, phenological, and productive expression of native *Capsicum chinense* ecotypes. These findings demonstrate that the observed phenotypic variability arises not only from genetic differences but also from soil–plant interactions. In this context, ecotypes exhibiting higher yield and enhanced morphological development reflect a superior capacity to adapt to acidic, low-fertility soils. This highlights the critical importance of integrating edaphic diagnostics into selection strategies, sustainable soil management practices, and genetic improvement programmes for humid tropical agroecosystems.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

HD: Investigation, Writing – original draft, Conceptualization, Funding acquisition, Writing – review & editing. LM: Writing – original draft, Data curation, Methodology. MS: Supervision, Writing – review & editing, Writing – original draft. JC: Writing – review & editing, Investigation, Data curation. BM: Data curation, Investigation, Writing – review & editing. GV: Writing – original draft, Formal analysis, Investigation, Writing – review & editing. GF: Writing – review & editing, Validation. JK: Visualization, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. The research was funded by the Instituto Nacional de Innovación Agraria, within the framework of the project: Mejoramiento de los servicios de investigación y transferencia tecnológica en el manejo y recuperación de suelos agrícolas degradados y aguas para riego en la pequeña y mediana agricultura en los departamentos de Lima, Áncash, San Martín, Cajamarca, Lambayeque, Junín, Ayacucho, Arequipa, Puno y Ucayali" CUI 2487112.

Conflict of interest

The author(s) declared that this work 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) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

1. Morales A, Villalobos FJ. Using machine learning for crop yield prediction in the past or the future. *Front Plant Sci*. (2023) 14:1128388. doi: 10.3389/fpls.2023.1128388

2. Khatibi SMH, Ali J. Harnessing the power of machine learning for crop yield prediction and decision support. *Front Plant Sci*. (2024). 15:1417912. doi: 10.3389/fpls.2024.1417912

3. Lozada DN, Booij JA, García-Pérez L. Chile pepper (Capsicum) breeding and improvement. *Front Plant Sci*. (2022) 13:879182. doi: 10.3389/fpls.2022.879182

4. Castaño E, Vega-Muñoz MA, Grisales-Vásquez NY, Loaiza-Loaiza OA, Henao-Rojas JC, Montoya G. Capsicum germplasm targeted valorization using physicochemical and phytochemical descriptors. *Front Sustain Food Syst*. (2025) 1:1571012. doi: 10.3389/fsufs.2025.1571012

5. MIDAGRI. *Comercio exterior agrario 2024 (Anuario)*. Ministerio de Desarrollo Agrario y Riego del Peru (2025). Available online at: https://siea.midagri.gob.pe/portal/phocadownload/datos_estadisticas/anuarios/comercio_exterior/comercio_exterior_2024.pdf (Accessed August 11, 2025).

6. Solórzano R, Cruz J, Gaona-Jimenez N, Lozano A, Díaz-Chuquizuta H, Vallejos-Torres G, et al. Impact of liquid biofertilizer from cocoa shells on the growth and chlorophyll content of sweet peppers (Capsicum chinense L.) in San Martín, Peru. *arXiv*. (2024). 7:1673914. doi: 10.3389/fagro.2025.1673914

7. Cruz ACI. Variabilidad del fruto en el ecotipo ají "charapita" (Capsicum frutescens L.) de la Amazonía Peruana. *Ciencia. Amazónica*. (2020) 8(2):167–84. Available online at: https://ojs.ucp.edu.pe/index.php/cienciaamazonica/article/download/295/145/.

8. Mishra DS, Yadav V, Berwal MK, Ravat P, Singh A, Rane J, et al. Integrative multivariate and machine learning analyses reveals trait architecture and selection potential in guava (Psidium Guajava L.). *BMC Plant Biol*. (2026) 1–59. doi: 10.1186/s12870-026-08216-3

9. Long Y, Zeng Y, Liu X, Yang Y. Multivariate analysis of grain yield and main agronomic traits in different maize hybrids grown in mountainous areas. *Agriculture*. (2024) 14:1703. doi: 10.3390/agriculture14101703

10. Tlahig S, Mohamed A, Triki T, Yahia Y, Yehmed J, Yahia H, et al. Integrated agro-morphological and molecular characterization for progeny testing to enhance alfalfa breeding in arid regions of Tunisia. *J Agric Food Res*. (2025) 20:101793. doi: 10.1016/j.jafr.2025.101793

11. Al Galib MA, Farzana S, Chakrobarty T, Islam MZ, Shirazy BJ, Rahman MA, et al. Multivariate analysis for identifying high-yielding rice cultivars based on seed yield and morphological traits. *Discov Plants*. (2025) 2:256. doi: 10.1007/s44372-025-00345-7

12. Shirzad A, Asghari A, Moharramnejad S, Shiri M, Ebadi A. Integrated analysis of genotype by yield trait and genotype by environment interactions for selecting superior maize genotypes. *Sci Rep*. (2025) 15:41372. doi: 10.1038/s41598-025-25254-3

13. Servicio Nacional de Meteorología e Hidrología del Peru. *Boletín climatológico anual 2024*. SENAMHI (2024). Available online at: https://www.senamhi.gob.pe (Accessed September 18, 2025).

14. Powadi AA, Jubery TZ, Tross M, Shrestha N, Coffey L, Coffey L, et al. Enhancing yield prediction from satellite imagery via genotype–environment disentanglement. *Front Plant Sci*. (2025) 16:1617831. doi: 10.3389/fpls.2025.1617831

15. Ortega-Albero N, Pérez-Ramírez P, Torres-García A. Genetic diversity and phenotyping in Capsicum. *Front Plant Sci*. (2024) 15:1435427. doi: 10.3389/fpls.2024.1435427

16. Quispe Huincho MR. *Muestreo de suelos y servicios de análisis*. INIA (2022). Available online at: https://repositorio.inia.gob.pe/items/c79f3f8a-6ef4-4c4a-9eda-005034875b69 (Accessed June 27, 2025).

17. NOM-021-RECNAT-2000. Natural resources and environment secretary. Mexican official standard. Specifications of fertility, salinity, and soil classification. Study Sampling Analysis. (2022). Available online at: https://www.ordenjuridico.gob.mx/Documentos/Federal/wo69255.pdf

18. Madden LV. The value of linear mixed models for horticultural research. *Front Hortic*. (2024) 1:1423462. doi: 10.3389/fhort.2024.1423462

19. Sainju UM, Liptzin D. Multivariate interpretation of soil chemical properties. *Front Environ Sci*. (2022) 10:1005114. doi: 10.3389/fenvs.2022.1005114

20. Shahane AA, Shivay YS. Soil health and novel agronomic practices. *Front Agron*. (2021) 3:680456. doi: 10.3389/fagro.2021.680456

21. Brainard SH, Bustamante JA, Dawson JC, Spalding EP, Goldman IL. A digital image-based phenotyping platform for analyzing fruit and seed traits. *Plants*. (2021) 12:1–13. doi: 10.3389/fpls.2021.690031

22. Pereira-Dias L, Fita AA, Vilanova S, Sánchez-López E, Rodríguez-Burruezo A. Phenomics of peppers using high-throughput phenotyping. *Sci Hortic*. (2020) 275:109701. doi: 10.1016/j.scienta.2020.109701

23. López DL, Rodriguez LE, Azabache YF, Rodriguez RF, Chuquimbalqui D, et al. Impact of Bacillus subtilis as a growth promoter in the cultivation of Tabasco chili pepper (Capsicum frutescens). *Adv Agric*. (2025) 2025:018672. doi: 10.1155/aia/5018672

24. Maazallahi A, Thota S, Kondaboina NP, Muktineni V, Annem D, Rokkam AS, et al. Naïve Bayes and Random Forest for crop yield prediction. *arXiv*. (2025) 1–59. Available online at: https://arxiv.org/abs/2404.15392.

25. Ullah A, Shakeel A, Muhu-Din HG, Naeem M, Ali M, Shah AN, et al. Principal components and genetic basis in crops. *Front Plant Sci*. (2022) 13:981369. doi: 10.3389/fpls.2022.981369

26. Dhillon MS, Dahms T, Kuebert-Flock C, Rummler T, Arnault J, Steffan-Dewenter I, et al. Integrating random forest and process-based crop models to improve predictive skill. *Front Remote Sens*. (2023) 4:1010978. doi: 10.3389/frsen.2022.1010978

27. Jeong J-H, Kim Y-H, Kang S-W, Lee K-H. A random forest ranking approach for yield prediction using UAV-based indices. *Comput Electron. Agric*. (2020) 178:105791. doi: 10.1016/j.compag.2020.105791

28. Mohanarajan P, Usha P. Enhancing crop yield prediction using Random Forest. *J Data Min Manage*. (2025) 10:38–45. Available online at: https://matjournals.net/engineering/index.php/JoDMM/article/view/1659

29. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: Why it works. *J Classif*. (2014) 31:274–95. doi: 10.1007/s00357-014-9161-z

30. Carvalho LC, Gonçalves EF, da Silva JM, Costa JM. Potential phenotyping methodologies to assess inter- and intra-specific variability. *Front Plant Sci*. (2021) 12:718202. doi: 10.3389/fpls.2021.718202

31. Ghysels S, Baets BD, Reheul D, Maenhout S. Image-based yield prediction using RF and CNN approaches. *Front Plant Sci*. (2025) 16:1549099. doi: 10.3389/fpls.2025.1549099

32. Jo J, Kim Y, Kim GW, Kwon J-K, Kang B-C. Development of a panel of genotyping-in-thousands by sequencing (GT-seq) markers for Capsicum spp. *Front Plant Sci*. (2021) 12:769473. doi: 10.3389/fpls.2021.769473

33. Hong J-P, Ro N, Lee H-Y, Kim GW, Kwon J-K, Yamamoto E, et al. Genomic selection for prediction of fruit-related traits in pepper (Capsicum spp.). *Front Plant Sci*. (2020) 11:570871. doi: 10.3389/fpls.2020.570871

34. Zhou H, Yang J, Lou W, Sheng L, Li D, Hu H. Improving grain yield prediction through fusion of multi-source imagery and agronomic data. *Front Plant Sci*. (2023) 14:1217448. doi: 10.3389/fpls.2023.1217448

35. Ding C, Alghabari F, Rauf M, Zhao T, Javed MM, Alshamrani R, et al. Optimization of soybean physiochemical, agronomic, and genetic responses under varying regimes of day and night temperatures. *Front Plant Sci*. (2024) 14:1332414. doi: 10.3389/fpls.2023.1332414

36. Tanin MJ, Sharma A, Saini DK, Singh S, Kashyap L, Srivastava P, et al. Ascertaining yield and grain protein content stability in wheat genotypes having the Gpc-B1 gene. *Front Genet*. (2022) 13:1001904. doi: 10.3389/fgene.2022.1001904

37. Mohan RNVJ, Rayanoothala PS, Sree RP. Next-gen agriculture: integrating AI and XAI for precision crop management. *Front Plant Sci*. (2024) 15:1451607. doi: 10.3389/fpls.2024.1451607

38. Bojórquez-Quintal E, Escalante-Magaña C, Echevarría-MaChado I, Martínez-Estévez M. Aluminum, a friend or foe of higher plants in acid soils. *Front Plant Sci*. (2017) 8:1767. doi: 10.3389/fpls.2017.01767

39. Huang C, Zhao Y, Wang L, Zhang J. Soil acidity shapes plant nutrient uptake and root traits: Implications for crop productivity. *Front Plant Sci*. (2023) 14:1206820. doi: 10.3389/fpls.2023.1206820

40. Gurmessa B. Challenges of soil acidity and the importance of liming and organic amendments in tropical agricultural land with reference to Ethiopia. *Environ Dev Sustain*. (2021) 23:77–99. doi: 10.1007/s10668-020-00615-2

41. Vallejos-Torres G, Sánchez T, García MA, Trigoso M, Arévalo LA. Efecto de hongos formadores de micorrizas arbusculares en clones de Coffea arabica, variedad caturra en Peru. *Acta Agronómica*. (2019) 68:278–84. doi: 10.15446/acag.v68n4.72117