Genet Resour Crop Evol (2026) 73:168
https://doi.org/10.1007/s10722-026-02799-4

RESEARCH ARTICLE

Field-image endocarp phenotyping for provenance discrimination and traceability in *Spondias tuberosa*

Maria Suzana Oliveira da Silva · Natali Aparecida Santana · Valdinete Vieira Nunes · Matheus Emannuel Oliveira Vieira · Crislaine Costa Calazans · Daniela Aparecida de Castro Nizio · Renata Silva-Mann

Received: 8 December 2025 / Accepted: 26 March 2026 / Published online: 9 April 2026
© The Author(s) 2026

Abstract Understanding how fruit traits capture provenance structure is essential for traceability and the effective use of native fruit trees in conservation and pre-breeding. We assessed 30 mother trees of *Spondias tuberosa* (umbuzeiro) from three provenances in Sergipe, Brazil; due to asynchronous fruiting and field-access constraints, ripe fruits were obtained for 23 trees within the sampling window. Basic shape descriptors were computed under a simple, standardized setup, and basic shape descriptors were computed (length, width, aspect ratio, circularity) and CIELAB color indices. Exploratory analyses revealed clear phenotypic structure by provenance, with color channels—especially lightness and the green–red axis—carrying much of the discriminating signal, while simple shape metrics added complementary information. Models based on a compact subset of descriptors distinguished provenances with moderate performance, indicating that low-cost, image-derived features are sufficient to recover geographic structure. Dendrometric traits and fruit-quality measures (total soluble solids, pH, titratable acidity, firmness) provided additional context but did not replace the information captured by endocarp images. We translate these results into a concise descriptor panel and a field-ready workflow that can be implemented in collections and nurseries to support seed-lot traceability, germplasm management, and the selection of mother trees. The findings demonstrate a practical path to phenotyping at scale for conservation and sustainable use of umbuzeiro.

Keywords Caatinga · *Spondias tuberosa* · Phenotypic diversity · Mother-tree design · Chromatic descriptors · Endocarp shape

M. S. O. da Silva · N. A. Santana · V. V. Nunes · M. E. O. Vieira
Graduate Program in Agriculture and Biodiversity, Federal University of Sergipe, São Cristóvão, Sergipe, Brazil
e-mail: suzanaoliveira@academico.ufs.br

N. A. Santana
e-mail: nataliknd@academico.ufs.br

V. V. Nunes
e-mail: val.ufs@gmail.com

M. E. O. Vieira
e-mail: matheusemannuel.ov@gmail.com

C. C. Calazans
Department of Forestry Sciences, Federal University of Sergipe, São Cristóvão-SE, Brazil
e-mail: crislainecalazans@academico.ufs.br

D. A. de Castro Nizio
Federal Institute of Education, Science and Technology of Sergipe (IFS), São Cristóvão-SE, Brazil
e-mail: danielanizio82@gmail.com

R. Silva-Mann (✉)
Department of Agronomic Engineering, Federal University of Sergipe, São Cristóvão, Sergipe, Brazil
e-mail: renatamann@academico.ufs.br

Introduction

Brazil harbors exceptional biological diversity, and the Caatinga—the only Brazilian biome restricted entirely to the national territory—holds major ecological, social, and biotechnological importance. Most of its vegetation is deciduous as a water-saving strategy, whereas the umbu tree (*Spondias tuberosa* Arruda), a native species widely distributed in the semiarid, remains green year-round owing to underground xylopodia that store water and nutrients and confer high drought tolerance (Aguiar et al. 2023; Fernandes et al. 2020; Matos et al. 2020). Despite this relevance, *S. tuberosa* remains a neglected and underutilized genetic resource, even with well-documented food, pharmacological, and nutraceutical potential (Xavier et al. 2022). Beyond its economic value, the species sustains sociocultural practices in extractive communities of the Caatinga (Santos et al. 2019). Ongoing erosion of genetic diversity jeopardizes adaptive capacity under environmental change and entails the loss of valuable information for breeding (Stange et al. 2021).

In Bahia, advances in evaluation and selection culminated in the registration of four cultivars—BRS 48, BRS 52, BRS 55, and BRS 68 (Santos et al. 2021). In Sergipe, however, phenotypic characterization of native populations remains incipient (Santana et al. 2024), constraining conservation and pre-breeding initiatives in a production system heavily dependent on extractivism (IBGE 2023). Characterizing phenotypic variability within and among populations is essential to understand geographic differentiation, management effects, and breeding potential. For *S. tuberosa*, dendrometric descriptors (e.g. equivalent diameter at breast height, total height, crown area), together with fruit and endocarp morphological and chromatic traits, are particularly informative for detecting divergence among provenances and for guiding mother-tree selection and conservation strategies (Bianchi et al. 2016; Motta et al. 2015; Ribeiro et al. 2019; Tamura 2024).

Plant phenomics and image-based phenotyping have emerged as key approaches to overcome the phenotyping bottleneck, enabling standardized, high-throughput, and increasingly low-cost measurements of complex traits in the field (Furbank and Tester 2011; Araus and Cairns 2014; Cobb et al. 2013; Li et al. 2014). Simple imaging setups can generate large numbers of reproducible geometric and color descriptors, which are particularly useful for perennial crops and underutilized species where molecular resources are scarce and evaluation budgets are limited. Developing and testing robust, field-deployable image-based workflows for native fruit trees such as S. tuberosa is therefore an important step toward integrating phenomics into genetic resource management and pre-breeding.

A key challenge in in situ research is aligning trait definitions and statistical modeling with the biological sampling unit so that genetic and environmental components of variation can be disentangled (Bolker et al. 2009; Harrison et al. 2018). Because the fruit pericarp—including the endocarp—derives from the ovary wall and is therefore maternal tissue, multiple fruits from a given tree provide replicate observations of the same maternal genotype (Seymour et al. 2013; Dardick and Callahan 2014; Consonni et al. 2005). Accordingly, analyses are well served by a hierarchical mixed-effects framework with provenance as a fixed effect and mother tree (nested) as a random effect, partitioning variance into among–mother-tree (genetic + permanent microsite) and within-tree (microenvironment + measurement) components (Pinheiro and Bates 2000; Bolker et al. 2009; Harrison et al. 2018; Resende 2016). Under field-image conditions, trait definitions should also favor descriptors less sensitive to boundary discretization: sphericity as width/length (W/L, the ratio of minor to major axis) is typically more robust than circularity ($4\pi A/P^2$), because perimeter estimation is resolution- and edge-noise-dependent and can destabilize circularity across segmentations (Gonzalez and Woods 2018; Russ 2018; Sladoje and Nyström 2005; Phillip et al. 2021; Berrezueta et al. 2019). Where tree-level repeatability is reported, it is interpreted as an operational upper bound to broad-sense heritability, with classic caveats regarding environmental covariance and sampling scope (Lessells and Boag 1987; Dohm 2002).

Contribution and novelty. We introduce a practical, low-cost plant phenomics framework to characterize genetic resources of a native fruit tree by combining field-image-based endocarp morphometry and chromatic descriptors with a mother-tree analytical design. A standardized imaging workflow (calibrated lighting, in-frame scale, white balance) yields reproducible features; among these, sphericity (W/L) and

CIELAB channels consistently capture provenance structure better than raw size. Methodologically, we respect the sampling hierarchy—using mixed-effects variance partitioning and mother-tree–stratified cross-validation—which provides unbiased estimates and remains robust when per-population sample sizes are modest. The outcome is a compact descriptor panel and a field-ready image-based phenotyping protocol that support seed-lot traceability, conservation ranking, and pre-breeding in under-sampled, extractive systems.

Against this backdrop, the present study develops and tests an image-based phenotyping workflow for *S. tuberosa* and uses it to characterize and compare the phenotypic variability of Sergipe populations by integrating dendrometric, morphometric, chromatic, and physicochemical descriptors within a mother-tree analytical design. Our goals are to (i) document provenance-level differentiation under natural, extractive conditions; (ii) illustrate how trait choices and hierarchical modeling affect genetic inference in a native woody species; and (iii) provide an operational trait panel to inform in situ conservation planning, mother-tree selection, prospective origin verification/traceability, and pre-breeding of the umbu tree. Given the single-season sampling and the high within-provenance variability typical of perennial outcrossers, we frame the application to origin verification as a proof of concept that requires multi-season, multi-site validation and standardized imaging before routine use.

Material and methods

Study area

Three field expeditions in February 2023 surveyed 30 reproductively mature umbu trees (*Spondias tuberosa* Arruda) across three provenances in Sergipe State, Brazil—Canindé de São Francisco (Santana et al. 2024), Poço Verde, and Gararu. The sites span the Caatinga climatic gradient from semiarid steppe (Köppen–Geiger BSh) to tropical with a dry summer (As), with long-term means of ~ 541–805 mm annual precipitation and ~ 24–35 °C air temperature. All mother trees were georeferenced (WGS84). The geographic context and sampling locations are shown in Fig. 1.

This research has protocol from the National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SisGen) in compliance with the provisions of Law nº 13.123/2015 (Brasil 2015) under registration number A47C741.

Climate data (February 2023)

Municipal precipitation totals (mm) were retrieved from EMDAGRO's official rainfall table. Monthly air temperature (T2M, T2M_MAX, T2M_MIN; °C) and all-sky surface shortwave irradiance (ALLSKY_SFC_SW_DWN; MJ m^{-2} day^{-1}; also reported as kWh m^{-2} day^{-1} = MJ/3.6) were extracted from the NASA POWER monthly API at the geographic centroids of Canindé de São Francisco, Gararu, and Poço Verde. As an independent check, hourly data from the nearest INMET automatic station (Nossa Senhora da Glória, A453) were aggregated to monthly values: RAD_GLO (kJ m^{-2} h^{-1}) was summed per day and converted to MJ m^{-2} day^{-1}; days with < 10 valid hourly records were excluded; daily precipitation was summed; and air-temperature statistics (mean, daily maxima/minima) were averaged over the month. Differences between POWER and INMET are expected given gridded satellite–reanalysis versus point-measurement and local microclimate. Summary plots are shown in Fig. 2.

Characterization of mother trees

Thirty reproductively mature mother trees were sampled with a minimum ~ 200 m inter-tree spacing to limit short-range spatial autocorrelation (spacing is not interpreted as proof of unrelatedness). For each tree we recorded total height (H, m), equivalent diameter at breast height (D_{eq}, cm; measured at 1.30 m on the upslope side, avoiding buttresses), and crown area (CA, m^2). For multi-stemmed individuals, $Deq = \sqrt{\sum_{i=1}^{k} d_i^{\,2}}$ where d_i are diameters at breast height (DBH, 1.3 m) measured on the upslope side (avoiding buttresses), and k is the number of stems; for single-stem trees, D_{eq} = DBH. CA was computed from two horizontal crown diameters (D_1 = maximum; D_2 = perpendicular) assuming an ellipse: $CA = (p_i * D_1 * D_2)/4$. All measurements were taken with a calibrated diameter tape and a hypsometer/clinometer; a subset was remeasured to verify

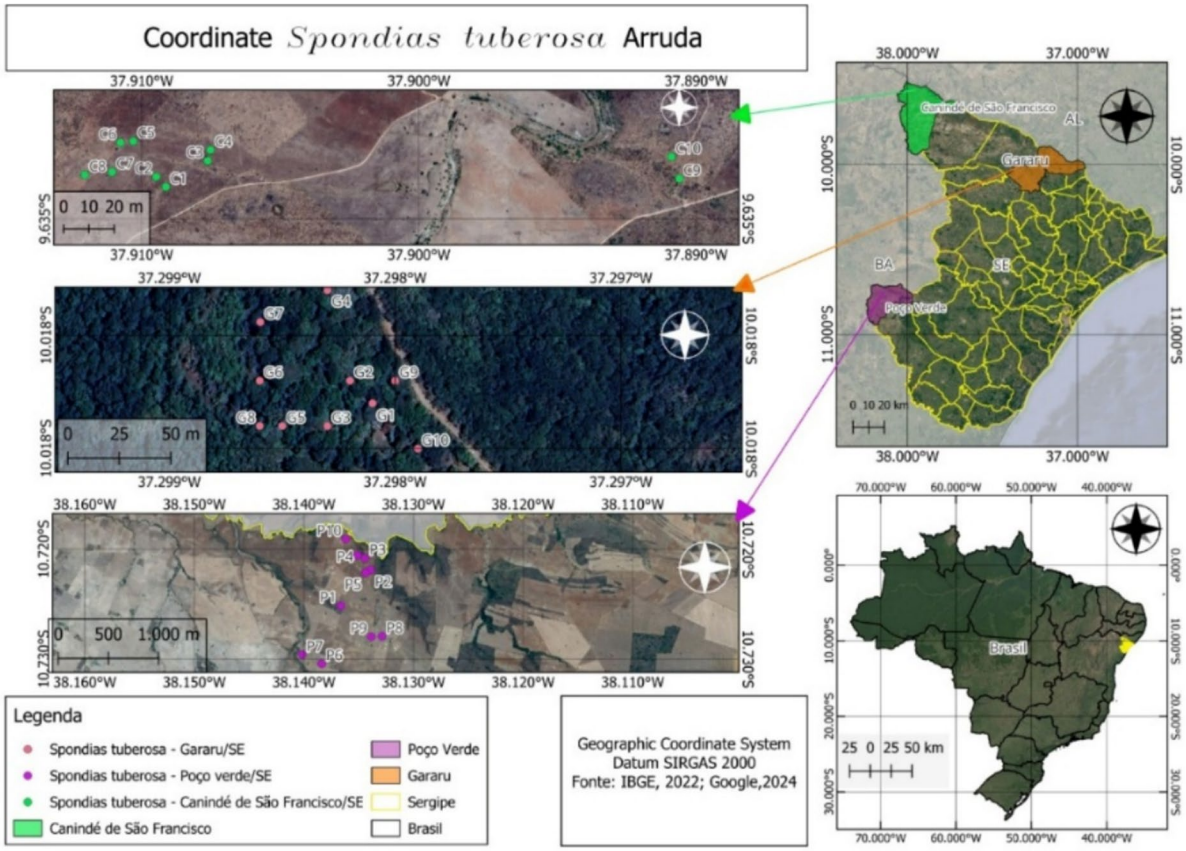


Fig. 1 Geographic location of sampled *Spondias tuberosa* (umbu) mother trees from three provenances in Sergipe State, Brazil—Canindé de São Francisco, Poço Verde, and Gararu (February 2023). Circles denote georeferenced tree positions recorded during field expeditions

repeatability; and GNSS positions were stored for spatial validation.

Fruit and endocarp sampling

Ripe fruits were harvested from 23 mother trees ($n = 508$ fruits). After cleaning and routine preparation (de-pulping for endocarps), 499 intact endocarps remained (damaged pieces were excluded a priori by QC rules). Where indicated, fruit-level values were averaged at the mother-tree level for downstream analyses.

Image acquisition and processing

Image-based phenotyping, image acquisition, and trait extraction. Fruits and endocarps were phenotyped using a GroundEye (Tbit) imaging system under a standardized, calibrated setup, following a hierarchical design in which multiple fruits from each mother tree represent replicate observations of the same maternal genotype. The system consisted of a fixed overhead RGB camera at a constant camera–sample distance, diffuse and homogeneous illumination, and an in-frame length scale to convert pixels to metric units. At the beginning of each imaging session, white balance and color calibration were performed against a reference target.

Images were processed through a reproducible pipeline to derive geometric and color descriptors. First, objects were segmented into binary masks using rule-based thresholding on the RGB channels, followed by limited manual editing under predefined quality-control criteria to correct obvious segmentation errors. From each binary mask, we extracted basic morphometric traits (area,

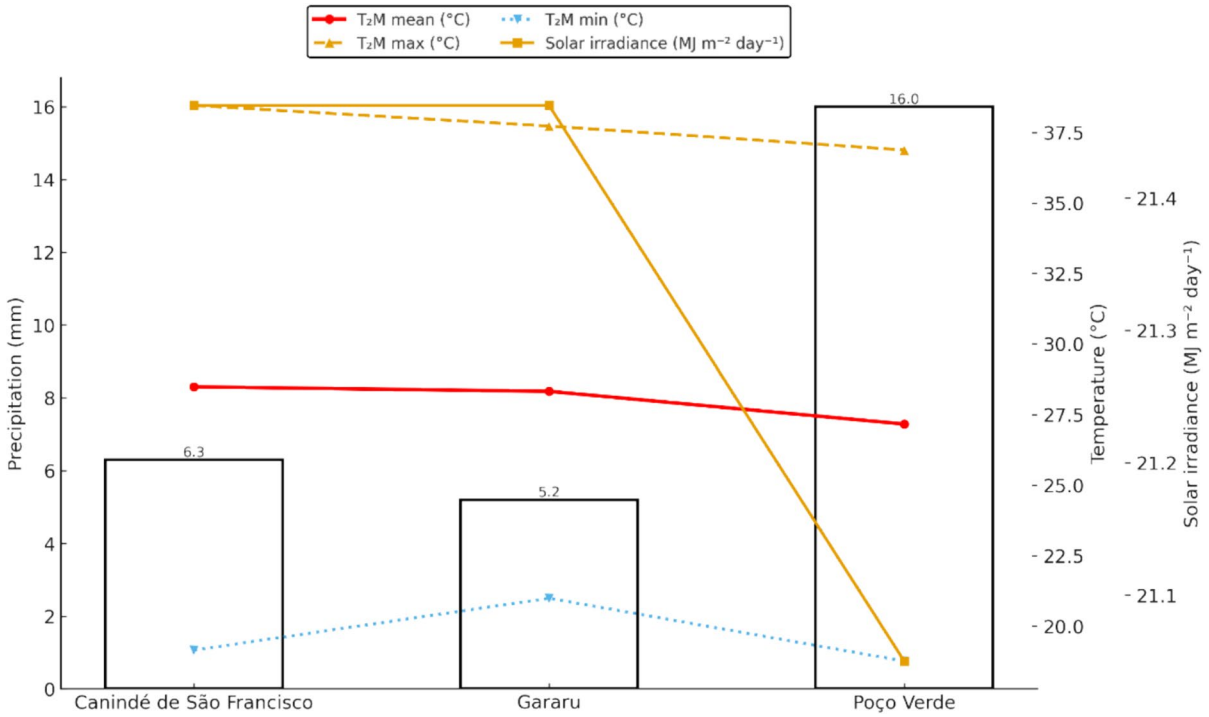


Fig. 2 February 2023 climate—precipitation, air temperature (T_2M) and solar irradiance in Canindé de São Francisco, Gararu and Poço Verde (Sergipe, Brazil)

perimeter, major and minor axes) and computed shape descriptors, including the width/length ratio (W/L, sphericity) and circularity ($4\pi A/P^2$). Color information was obtained from the original RGB images, which were converted to the CIELAB space after calibration; from these, we extracted L*, a*, b* (and, where relevant, derived h_ab and C*_ab). All descriptors were then aggregated at the mother-tree level, consistent with the hierarchical phenotyping design, for downstream mixed-effects modeling and multivariate analyses.From the masks we extracted morphometric descriptors—length, width, projected area (A), perimeter (P), and the aspect-ratio "sphericity" W/L (minor- to major-axis ratio). Circularity ($4\pi A/P^2$) was computed as a secondary metric due to its higher sensitivity to contour/segmentation noise.

Chromatic descriptors were computed over a consistent exocarp region of interest, avoiding specular highlights. Raw sRGB (0–255) values were stored; derived indices included HSV and CIELAB (D65, 2°) components L^*, a^*, b^*, plus hue (h_{ab}, degrees) and chroma ($C_{ab} = \sqrt{(a^{*2} + b^{*2})}$). Any negative values reported for individual channels reflect mean-centering/standardization prior to modeling; raw (non-centered) values are available in the dataset.

Quality control

All image-derived variables were stored at the per-fruit level with traceable IDs linking to mother tree and provenance. Obvious segmentation artifacts and missing values were flagged and removed under pre-specified QC rules. Measurement logs (instrument model, calibration notes, operator, session time) were recorded. A data dictionary accompanies all matrices to ensure variable-name consistency across steps.

Physicochemical analyses

Physicochemical traits were assessed on three ripe fruits per mother tree ($n = 69$): total soluble solids (TSS, °Brix), pulp pH, titratable acidity (TA, g citric acid 100 g^{-1} fresh weight), and pulp firmness (N = Newtons). Instruments were calibrated at the start of each session; for each fruit, three technical readings

were taken and averaged. TSS was read on expressed juice with a temperature-compensated refractometer (Manufacturer: Azzota Corporation, Model: DRB0-45nD); pH was measured with a bench pH meter; TA was obtained by acid–base titration and expressed as citric-acid equivalents; firmness was determined by axial compression using a penetrometer setup.

Statistical analysis

Univariate contrasts. We assessed group contrasts with Kruskal–Wallis tests followed by Dunn's post-hoc with Benjamini–Hochberg false discovery rate (FDR) correction (or analysis of variance (ANOVA) + Tukey when assumptions held).

Analysis level and preprocessing. Analyses were performed at the mother-tree level (one record per tree). When multiple fruits/endocarps were measured per tree, we averaged technical replicates to avoid pseudo-replication. Continuous predictors (shape and color indices) were checked for outliers; missing values (< 1% of cells) were imputed with the training-fold median only.

Leakage control and principal component analysis (PCA). To avoid leakage, all preprocessing and any feature selection were fit inside cross-validation on the training split and applied to the held-out split. PCA (on centered, scaled predictors) was used only for visualization; we inspected loadings and retained components until ~ 80% of explained variance.

Supervised classification. We evaluated linear discriminant analysis (LDA) and Random Forest (RF) for provenance prediction using stratified outer four-fold cross-validation at the mother-tree level (the 4 outer folds were chosen to respect the smallest class, $n = 4$). We compared pipelines without feature selection and pipelines with a pre-specified univariate selector (SelectKBest, *f_classif*, K = 6). Models used fixed settings: LDA (*solver* = svd); RF (*n_estimators* = 1,000; other parameters at defaults). No additional hyperparameter tuning was performed. Performance was computed from outer-fold predictions using accuracy, macro-F1, balanced accuracy, and Cohen's kappa (κ); confusion matrices aggregate all out-of-fold predictions. 95% confidence intervals for accuracy were obtained with the Wilson method. We also compared feature groups ("shape-only", "color-only", "shape + color").

Software and reproducibility. Analyses used Python 3.11, scikit-learn 1.4, numpy, scipy, pandas, and plots in matplotlib. The classification pipeline was: StandardScaler → (optional) SelectKBest(K = 6) → [LDA or RF].

Variance components and ICC. To partition within–versus between-tree variation and report intraclass correlation coefficients (ICC), we fit linear mixed-effects models at the fruit/endocarp level (before averaging). For each continuous trait we used a random-intercept model with provenance as a fixed effect and mother tree as a random effect:

$$y_{ij} = \beta_0 + \gamma_{prov[j]} + \mu_i + \varepsilon_{ij}; \mu_i \sim N\left(0, \sigma^2_{tree}\right); \varepsilon_{ij} \sim N\left(0, \sigma^2 e\right)$$

Models were fit by REML (Python 3.11.4, statsmodels 0.14.2, *MixedLM*). When warranted, traits were log-transformed. The ICC for mother tree was computed as:

$$ICCtree = \sigma^2 tree / \left(\sigma^2 tree + \sigma^2 e\right)$$

We obtained 95% CIs for variance components and ICC by parametric bootstrap (2,000 resamples). These mixed-effects fits were used only for variance partitioning and ICC reporting; all classification analyses remained at the mother-tree level using features averaged per tree.

Results

Dendrometric variation across provenances

Dendrometric profiles differed among provenances: Deq and CA varied clearly among sites, whereas total height overlapped broadly. On average, trees from Canindé de São Francisco had thicker stems, those from Poço Verde were taller with larger crowns, and Gararu showed the smallest values for all traits (Fig. 3), reinforcing Deq and CA as more informative than H for provenance discrimination.

Canindé de São Francisco exhibited the widest spread to Equivalent Diameter (Deq) (≈25–120 cm; median ≈50–55 cm). Poço Verde was slightly smaller (median ≈45–50 cm) with a narrower IQR, whereas Gararu had the smallest stems (median ≈25–30 cm) and comparatively low dispersion (Fig. 4).

Total height overlapped broadly among sites, with medians around 6–8 m, indicating that

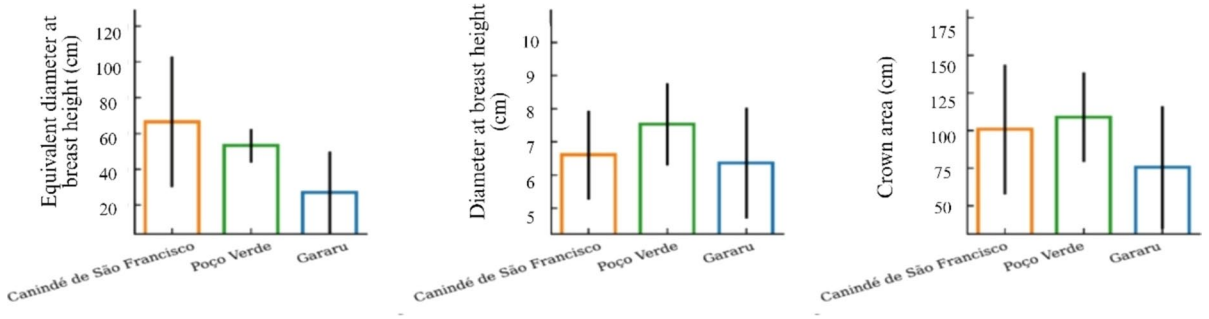


Fig. 3 Comparison of structural traits among *Spondias tuberosa* populations. Bars represent mean values and error bars indicate standard deviation (SD)

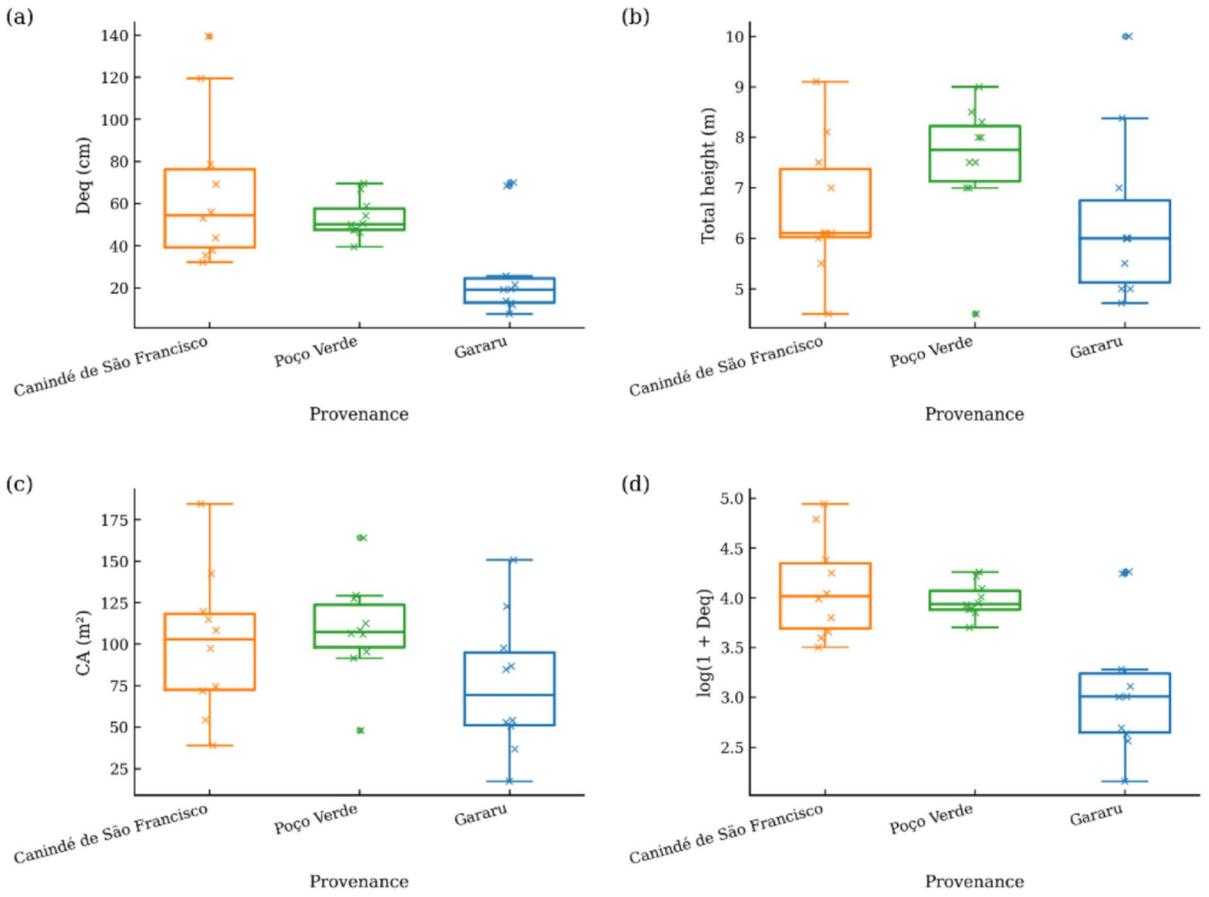


Fig. 4 Provenance-wise distributions of dendrometric descriptors for *Spondias tuberosa* in Sergipe, Brazil. **a** Equivalent diameter at breast height (Deq, cm). **b** Total height (m), **c** crown area ($m.^2$). **d** log(1 + Deq)

vertical growth is comparatively insensitive to provenance. In contrast, crown area (CA) and equivalent diameter (Deq) showed clear provenance differences—Poço Verde > Canindé de São Francisco > Gararu—with larger among-provenance and among–mother-tree variance components than

height. Thus, Deq and CA provide stronger provenance discrimination than total height, although σ^2m reflects combined genetic and micro-site effects rather than purely additive genetic variance.

Most of the phenotypic variance in the three dendrometric traits came from the within-tree/within-provenance component (σ^2e), with only modest variance among mother trees (σ^2m) and a small among-provenance share (Fig. 5). Deq and CA showed slightly more among-provenance variation than total height, which was more conservative. Importantly, σ^2m reflects combined genetic and permanent micro-site effects and cannot be taken as purely genetic; additive genetic variance (σ^2A) cannot be separated under the present design.

Fruit morphometry

The partitioning of fruit-trait variance among provenances revealed contrasting patterns across populations (Table 1). In Canindé de São Francisco (CSF), most colour and texture traits showed negligible among–mother-tree variance (σ^2m ≈ 0; ICCm = 0) and were largely driven by residual variation. In contrast, fruit width and area displayed strong maternal components (ICCm ≥ 0.85), indicating marked differentiation among families for these size traits, whereas fruit length was mainly influenced by residual variance.

In Poço Verde (POV), among–mother-tree variance was consistently important for fruit size traits, with ICCm values between 0.50 and 0.75 for length, width and area, and a moderate maternal component for yellow skin colour (ICCm = 0.33). This pattern points to a pronounced genetic structure for size and yellow chromatic attributes and greater potential for selecting superior genotypes based on these traits. Other descriptors, such as circularity, green channel, lightness and texture, showed little or no maternal variance and were comparatively uniform.

The Gararu (GAR) provenance was generally characterized by low σ^2m and ICCm values close to zero for most traits. Fruit length, width, yellow channel and texture exhibited virtually no among–mother-tree variance, whereas area, green channel and lightness showed only small maternal components. Both maternal and residual coefficients of variation were low, indicating relatively uniform fruits. Negative coefficient of variation (CV) values for chromatic traits simply reflect negative means and have no direct biological meaning.

Overall, fruit size descriptors (length, width and area) showed the highest maternal variance and intraclass correlations and were therefore the most informative for discriminating against differences among families, while most color and texture traits behaved as comparatively stable attributes less influenced by genetic origin (Table 1).

Size traits overlapped broadly among provenances (Fig. 6; all pairwise $p > 0.05$), with variability driven mainly by the within-provenance component (σ^2e). In

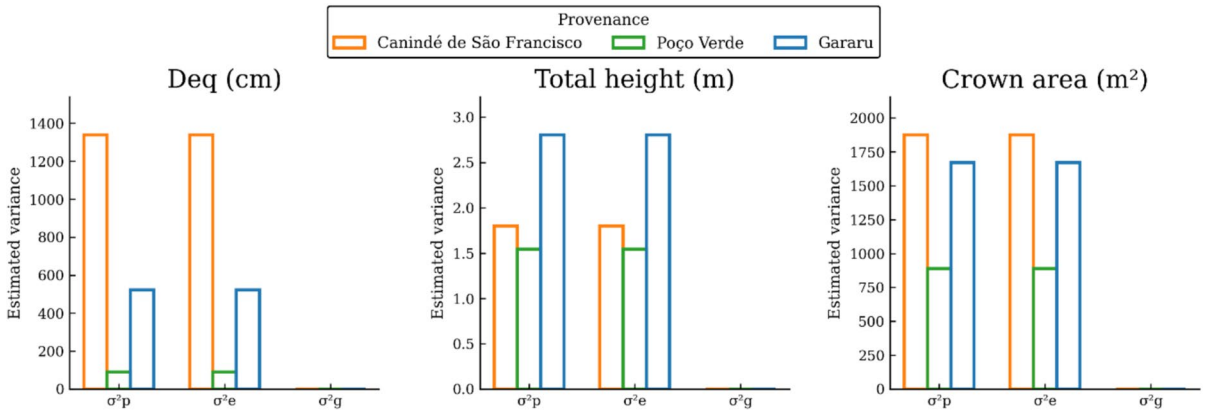


Fig. 5 Variance-component display for dendrometric descriptors of *Spondias tuberosa* in Sergipe, Brazil. Bars show the decomposition of the total phenotypic variance (σ^2p) into its main components for diameter at breast height (Deq), total height, and crown area (CA). Across traits, within-tree/within-provenance variance (σ^2e) dominates the variance structure, whereas the contribution of higher-level components (mother trees and provenances) is comparatively smaller, with Deq and CA showing a slightly stronger provenance signal than total height

Table 1 Variance partitioning of fruit traits by provenance

Provenance	Trait	N	Mean	σ^2e	σ^2m	σ^2p	ICCm	CVm (%)	CVe (%)	CVr
CSF	Length	216	3.49	0.06	0.00	0.06	0.00	0.00	7.37	–
CSF	Width	216	3.13	0.01	0.07	0.08	0.88	8.32	3.70	2.25
CSF	Area	216	8.58	0.34	1.91	2.25	0.85	16.13	6.80	2.37
CSF	Green	216	22.08	6.22	0.00	6.22	0.00	0.00	11.30	–
CSF	Yellow	216	47.24	18.31	0.00	18.31	0.00	0.00	9.06	–
CSF	Luminosity	216	50.49	31.36	0.00	31.36	0.00	0.00	11.09	–
CSF	Texture	216	115.43	202.35	0.00	202.35	0.00	0.00	12.32	–
POV	Length	240	3.51	0.02	0.06	0.08	0.75	6.81	4.39	1.55
POV	Width	240	3.04	0.02	0.02	0.04	0.50	5.06	4.66	1.09
POV	Area	240	8.43	0.49	0.87	1.36	0.64	11.08	8.27	1.33
POV	Circularity	240	0.87	0.00	0.00	0.00	–	3.90	3.87	1.00
POV	Green	240	−21.11	4.89	0.00	4.89	0.00	−0.00	−10.48	–
POV	Yellow	240	46.29	10.29	5.04	15.33	0.33	4.85	6.93	0.69
POV	Luminosity	240	50.13	32.21	0.00	32.21	0.00	0.00	11.32	–
POV	Texture	240	117.52	202.30	0.00	202.30	0.00	0.00	12.10	–
GAR	Length	94	3.50	0.02	0.00	0.02	0.00	0.00	4.56	–
GAR	Width	94	3.00	0.02	0.00	0.02	0.00	0.00	5.15	–
GAR	Area	94	8.21	0.45	0.12	0.57	0.21	4.21	8.20	0.51
GAR	Green	94	−2.31	2.47	0.30	2.76	0.11	−0.44	−7.04	0.35
GAR	Yellow	94	49.12	4.41	0.00	4.41	0.00	0.00	4.27	–
GAR	Luminosity	94	55.83	21.82	1.18	23.01	0.05	1.95	8.37	0.23
GAR	Texture	94	131.86	134.59	0.00	134.59	0.00	0.00	8.80	–

σ^2m = between-mother-tree variance; σ^2e = residual variance; σ^2p = σ^2m + σ^2e; ICCm = σ^2m/(σ^2m + σ^2e). CVm and CVe were computed as 100 × (standard deviation/Mean); negative signs can arise when Mean < 0 (chromatic channels). '–' indicates not defined. Circularity and Texture are dimensionless. CSF = Canindé de São Francisco; POV = Poço Verde; GAR = Gararu.

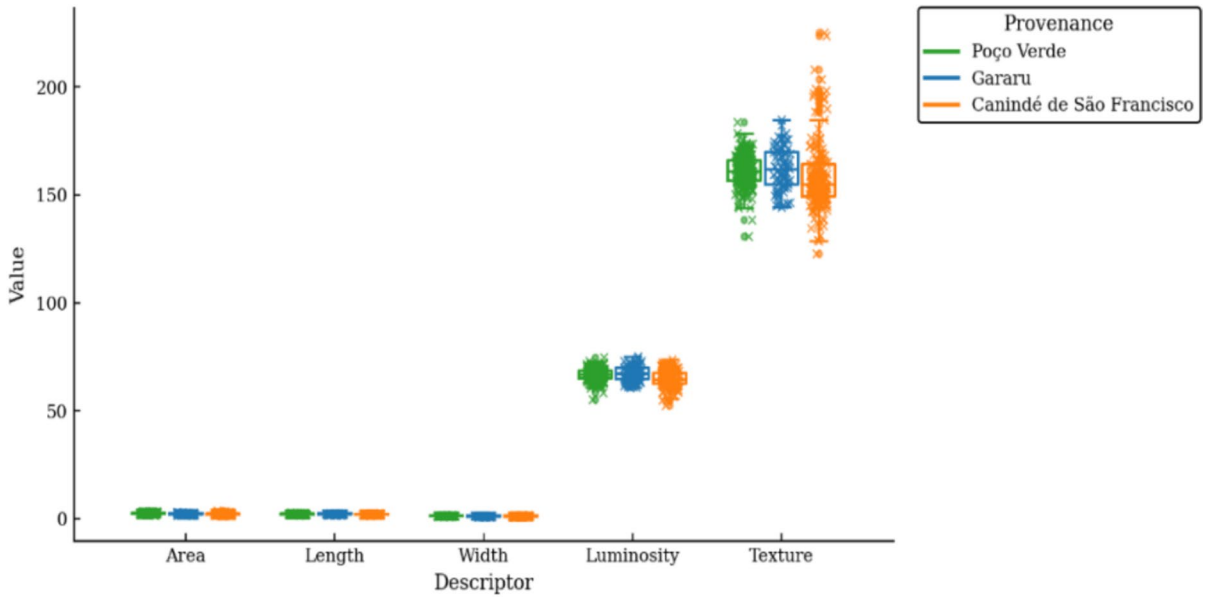


Fig. 6 Provenance-wise distributions of umbu endocarp descriptors (Sergipe, Brazil)

Canindé de São Francisco, mean length/width were 3.49/3.13 cm; ICCm was 0.00 for length and 0.88 for width, and area showed higher dispersion ($\sigma^2p = 2.25$) with ICCm = 0.85, indicating a strong mother-tree effect. In Poço Verde, means were similar (3.51/3.04 cm), with modest between–mother-tree variance for length ($\sigma^2m = 0.06$) and yellow intensity ($\sigma^2m = 5.04$; $ICC_m \approx 0.33$); area had $\sigma^2p = 1.36$ and ICCm = 0.64. In Gararu, width (3.00 cm) and area (8.21 cm^2) were smaller on average; $\sigma^2m \approx 0$ for most traits except area ($ICC_m = 0.21$). Chromatic/texture traits showed the largest residual dispersion (e.g., texture $\sigma^2p \approx 264.62$, luminosity 31.36) with ICCm ≈ 0, reinforcing that shape and color outperform raw size for provenance-level discrimination.

Size traits. Length, width, and area overlapped extensively across provenances (all pairwise $p > 0.05$), indicating that basic fruit size is driven mainly by within-provenance variability. Accordingly, size traits have limited discriminating power at the provenance scale (moderate CVs). For finer separation, either increase within-provenance sampling or prioritize more informative descriptors—especially shape (sphericity W/L, circularity $4\pi A/P^2$) and chromatic/texture traits.

The subset ($n = 69$; three fruits per mother tree) showed provenance-level differences in total soluble solids (TSS), titratable acidity (TA), pH, and pulp firmness, with effect sizes varying by trait. The largest contrasts occurred for TSS and firmness, whereas pH and TA exhibited narrower ranges with partial overlap among provenances (Fig. 7). Results are reported as estimated marginal means (±95% CIs) with FDR-adjusted pairwise contrasts.

Physicochemical attributes

The Principal Component Analysis (PCA) ordination of TSS (°Brix), pH, titratable acidity (TA), and firmness revealed two interpretable gradients. PC1 contrasted pH (positive loadings) with TA and, to a lesser extent, firmness (negative), so samples to the right tend to have higher pH/lower acidity, whereas those to the left show higher acidity (with a slight increase in firmness). PC2 was dominated by TSS, placing sweeter fruits higher on the plot. Along these axes, Canindé de São Francisco clusters at higher PC2 (higher TSS), Gararu shifts toward positive PC1 (higher pH, lower acidity), and Poço Verde tends toward negative PC1 (higher acidity, slightly greater firmness), with TSS generally lower than in Canindé. Despite these tendencies, provenances still overlap, and separation is driven chiefly by the TSS gradient and the pH–acidity balance (Fig. 8).

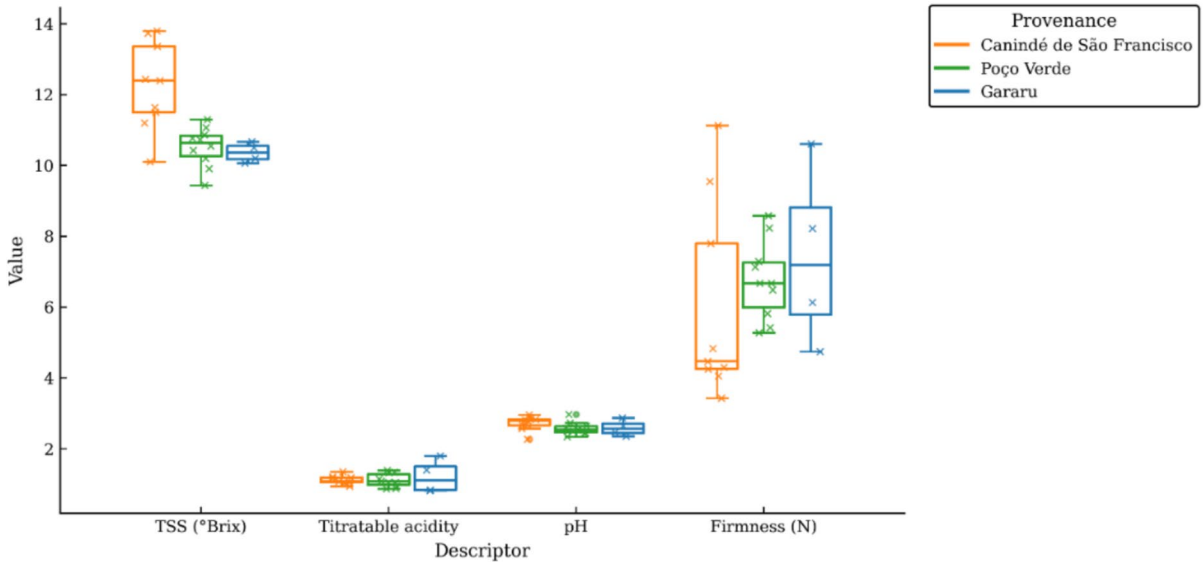


Fig. 7 Provenance-wise distributions of physicochemical attributes. Total soluble solids (TSS, °Brix), titratable acidity, pH, and firmness (N) for *Spondias tuberosa* fruits

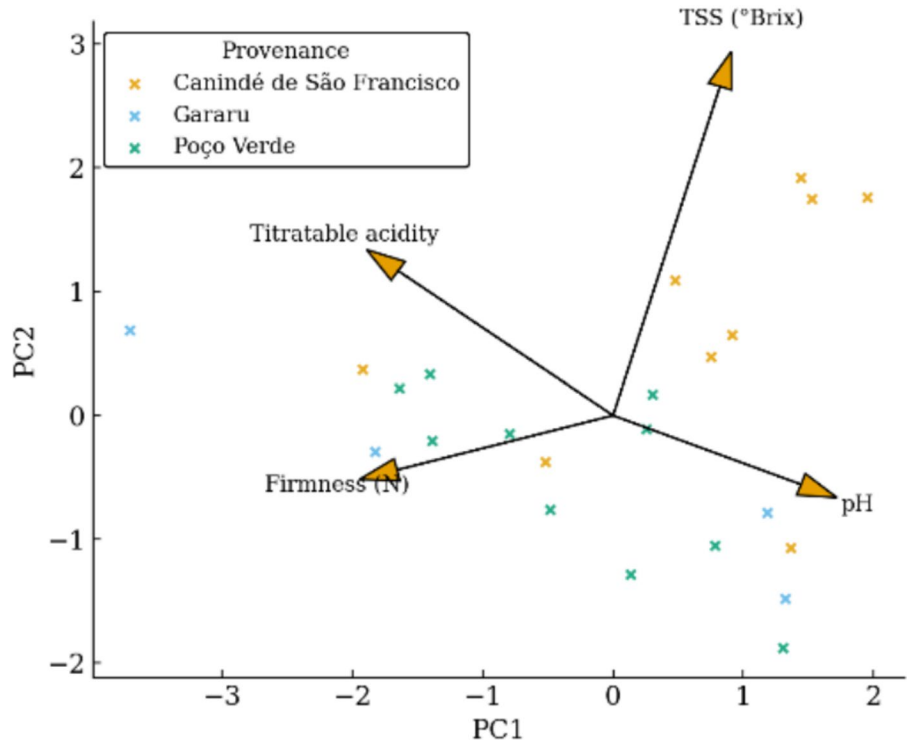


Fig. 8 Principal component analysis (PCA) biplot for provenance-wise distributions of physicochemical attributes (February 2023). Total soluble solids (TSS, °Brix), titratable acidity, pH, and firmness (N) for *Spondias tuberosa* fruits

Despite overlap among provenances, separation is driven chiefly by the TSS gradient (PC2) and the pH–acidity balance (PC1), which are therefore the most informative descriptors for provenance-level discrimination.

Endocarp morphometry

Fruit-derived analyses were restricted to the 23 of 30 trees that bore mature fruit within the sampling window, owing to asynchronous phenology and field-access constraints. Across 499 endocarps from 23 mother trees, endocarp traits provided the clearest provenance-level signatures. Shape descriptors—notably sphericity (W/L) and circularity ($4\pi A/P^2$)—discriminated provenances more strongly than size metrics (length, width, projected area), which overlapped broadly after accounting for mother-tree replication.

The endocarp morphometric descriptors exhibited high coefficients of variation, which is expected in species occurring across heterogeneous environments and under distinct management conditions. In *S. tuberosa*, populations from different sites are subject to variation in soil type, moisture availability, phenological timing, and degrees of domestication, all of which influence fruit and seed development. Such environmental and management heterogeneity is therefore reflected in the dispersion of morphometric measurements. Similar levels of morphological dispersion have been reported in other research with fruits and seeds of *Spondias* species, reinforcing the strong environmental component of endocarp development (e.g., Silva et al. 2025; de Sales et al. 2024). The summary statistics for each descriptor are presented in Table 2.

Across 499 endocarps (23 mother trees), shape traits—notably sphericity (W/L) and circularity ($4\pi A/P^2$)—provided the clearest provenance separation, whereas size metrics (length, width, area) overlapped broadly and showed small $\sigma^2 p$ (area 0.11–0.16, length 0.03–0.04, width 0.01–0.02). Texture was far more variable—especially in Canindé de São Francisco ($\sigma^2 p = 264.62$ vs 52.73–96.45 elsewhere). Chromatic traits also differed among provenances, though several channels exhibited high ennviromental coefficient of variation (CVe) (e.g., hue/brown > 50% in Canindé and Poço Verde), while others—and multiple texture indices—had manageable CVe (≈4.5–18.5%). Practically, Poço Verde tended to produce larger endocarps,

Table 2 Phenotypic variance (σ^2p) and environmental coefficient of variation (CVe, %) for endocarp traits of the umbu tree (*Spondias tuberosa* Arruda) across three natural provenances in Sergipe State, Brazil. Evaluated variables include chromatic attributes (hue, brown, green, red, yellow), morphometric attributes (area, length, width), and image-based texture

Provenance	Traits	N	Mean	σ^2p	CVe (%)
Canindé de São Francisco	Hue	191	4.81	6.32	52.26
Canindé de São Francisco	Brown	191	2.88	2.28	52.39
Canindé de São Francisco	Yellow	191	28.85	12.56	12.28
Canindé de São Francisco	Area	191	2.18	0.16	18.50
Canindé de São Francisco	Length	191	2.09	0.04	9.04
Canindé de São Francisco	Width	191	1.28	0.02	11.01
Canindé de São Francisco	Texture	191	158.85	264.62	10.24
Gararu	Hue	88	2.50	1.26	45.00
Gararu	Green	88	1.79	1.12	59.02
Gararu	Red	88	28.54	6.00	8.58
Gararu	Area	88	2.31	0.12	14.88
Gararu	Length	88	2.23	0.04	8.57
Gararu	Width	88	1.26	0.01	8.04
Gararu	Texture	88	162.08	96.45	6.06
Poço Verde	Hue	220	5.50	8.29	52.34
Poço Verde	Red	220	1.23	0.32	46.33
Poço Verde	Yellow	220	28.30	6.78	9.20
Poço Verde	Area	220	2.49	0.11	13.36
Poço Verde	Length	220	2.22	0.03	8.05
Poço Verde	Width	220	1.36	0.01	7.25
Poço Verde	Texture	220	160.96	52.73	4.51

and the chromatic/texture variables with moderate CVe are promising targets for selection and traceability. Overall, shape and color—rather than size—carry the strongest provenance signals (Fig. 9).

Endocarp size distributions (length, width, area) were consistently right shifted in Poço Verde, intermediate in Gararu, and lowest in Canindé de São Francisco, indicating larger endocarps in Poço Verde.

Multivariate structure

Fruit and endocarp morphometrics revealed partial provenance separation along the first two components (Fig. 11). Size- and shape-related loadings jointly structured the ordination, consistent with the univariate patterns (Fig. 10).

Principal component analysis (PC1 = 33.4%, PC2 = 21.6%; ≈55% cumulative) yielded partially overlapping point clouds at the mother-tree level, indicating only moderate provenance separation. The 95% concentration ellipses largely overlap; their size and orientation suggest substantial within-provenance dispersion—most notably in Gararu (broader ellipse)—while Canindé de São Francisco and Poço Verde occupy similar regions with only modest shifts in centroids (Fig. 11).

Overall, the pattern reflects gradual tendencies rather than discrete clusters, consistent with the univariate results, where within-provenance variability limits the discriminating power of size traits and elevates the importance of shape and chromatic/texture descriptors.

To assess provenance discrimination in a supervised setting, we trained Linear Discriminant Analysis (LDA) and Random Forest (RF) classifiers and evaluated them with stratified four-fold cross-validation at the mother-tree level, ensuring that fruits from the same mother tree never appeared in both training and test folds. Metrics (accuracy, precision, recall, F1) were macro-averaged by class and then averaged across folds; per-class confusion matrices are reported in Table 3.

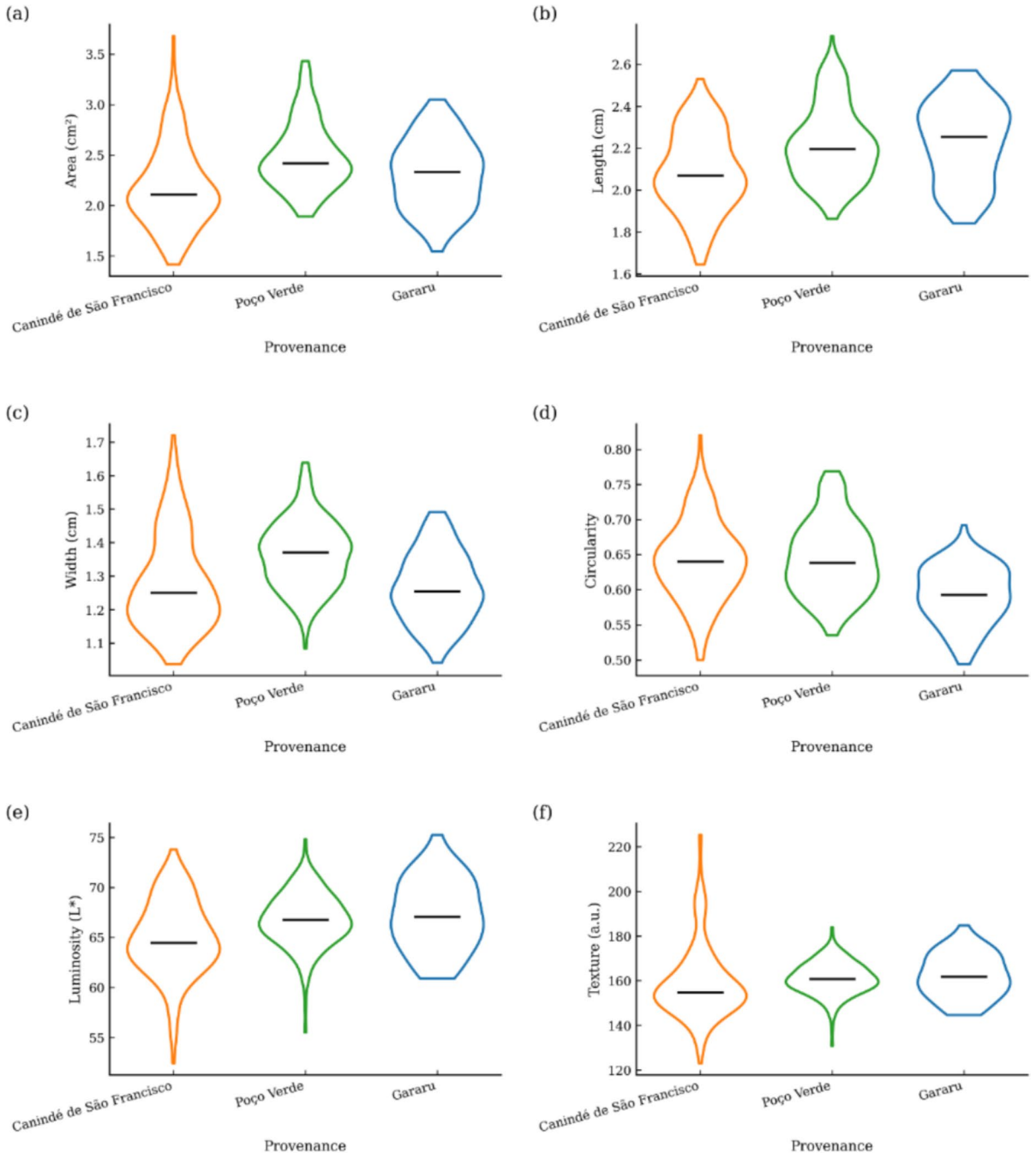


Fig. 9 Provenance-wise distributions of endocarp descriptors for *Spondias tuberosa* (Sergipe, Brazil; February 2023). Violin plots (unfilled, colored borders) show as **a** area (cm^2), **b** length (cm), **c** width (cm), **d** circularity, **e** luminosity (L*), and **f** texture (a.u.). Horizontal lines denote medians; shaded envelopes depict the distribution shape

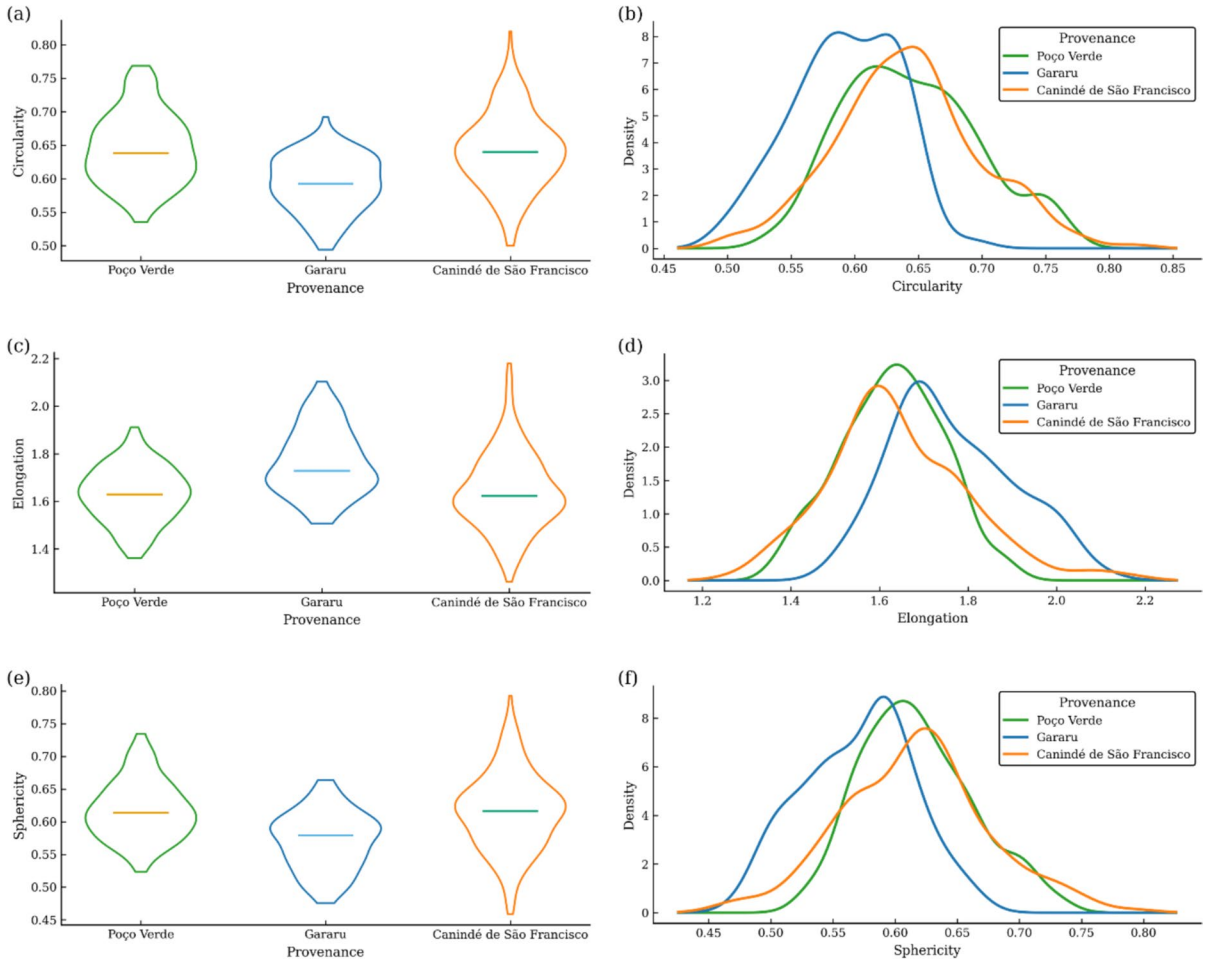


Fig. 10 Provenance-wise distributions of endocarp shape descriptors for *Spondias tuberosa* in Sergipe State, Brazil (February 2023; $n = 499$ endocarps from 23 mother trees). Panels **a**, **c**, and **e** show outline-only violin plots (no fill) for circularity (dimensionless), elongation (=length/width), and sphericity (=width/length); horizontal bars mark medians and violin width reflect probability density. Panels **b**, **d**, and **f** display the corresponding kernel density estimates (KDEs)

Allometry among dendrometric traits

We quantified allometric scaling between stem size and tree architecture with ordinary least squares on log–log pairs. Specifically:

$$ln(Height[m]) = \beta 0 + \beta 1 ln(Deq[cm]) + \varepsilon$$

$$ln\left(Crownarea\left[m^2\right]\right) = \beta 0 + \beta 1 ln(Deq[cm]) + \varepsilon$$

We report slopes, intercepts, 95% confidence intervals, and goodness-of-fit, and compare model specifications by AICc (Table 4).

Global log–log fits ($n = 30$) showed a shallow height–diameter relationship ($\beta 1 \approx 0.155$; 95% CI 0.05–0.26) and a steeper crown area–diameter scaling ($\beta 1 \approx 0.547$; 95% CI 0.35–0.75), indicating that canopy expansion outpaces vertical growth. Provenance-specific fits had wide, overlapping CIs, and model selection favored a common-slope (global) specification over separate-by-provenance slopes (ΔAICc ≈ 5 for height; ≈ 4 for crown area), implying shared allometric trends across sites, while intercepts capture provenance-level differences (Fig. 12).

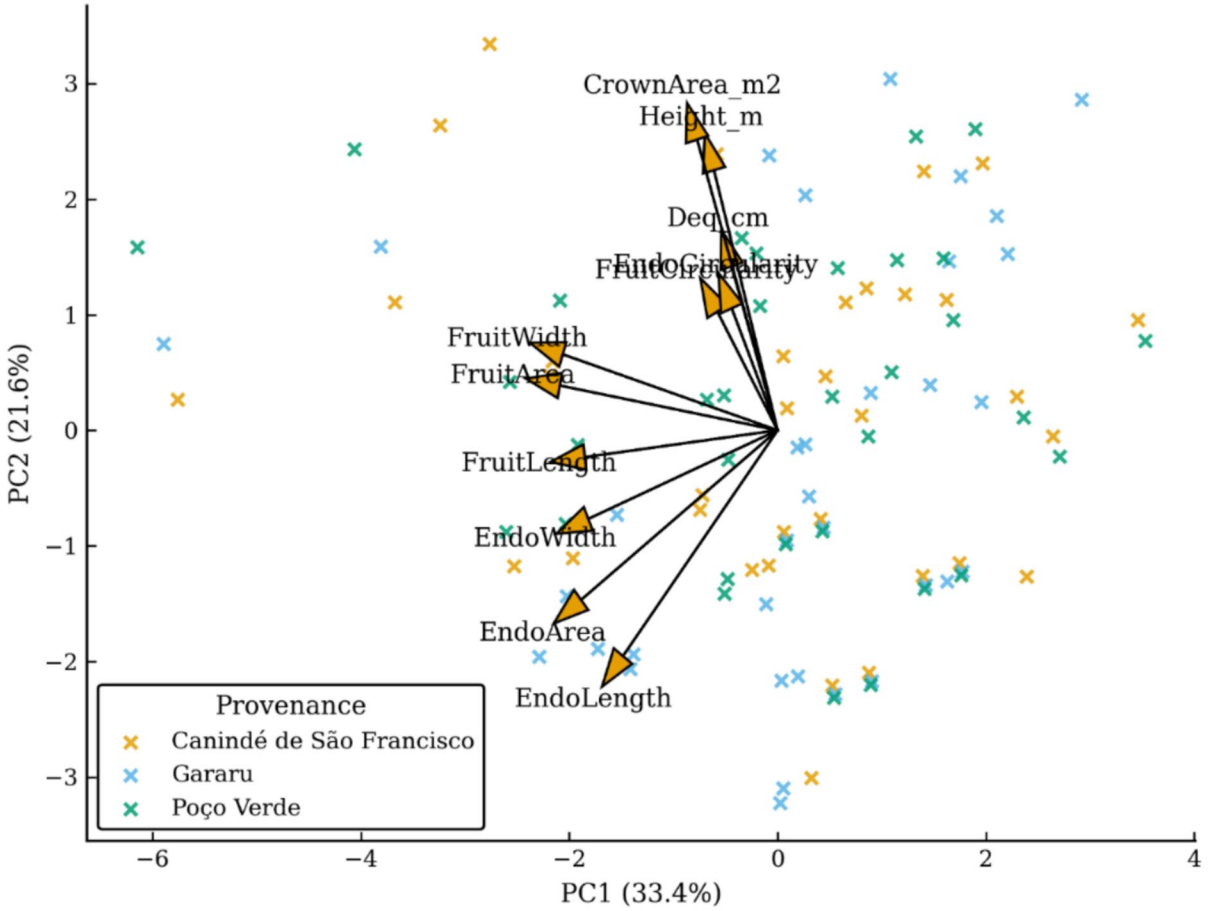


Fig. 11 Principal component analysis (PCA) biplot global of dendrometric, fruit, and endocarp morphometry by provenance (PC1 = 33.4%, PC2 = 21.6%)

Table 3 Classification performance (outer, stratified fourfold coefficient of variation (CV); mother-tree level; 95% confidence intervals (CIs)) (Wilson, proportions)

Model	Accuracy (%)	Macro-F1	Balanced accuracy	Cohen's κ	95% CI (accuracy)	N
Most-frequent baseline	43.5				(0.256–0.632)	23
LDA (no feature selection)	69.6	0.531	0.559	0.498	(0.491–0.844)	23
RF (no feature selection)	60.9	0.456	0.493	0.345	(0.408–0.778)	23
LDA (SelectKBest, K = 6)	56.5	0.439	0.452	0.294	(0.368–0.744)	23
RF (SelectKBest, K = 6)	65.2	0.488	0.526	0.416	(0.449–0.812)	23

Confusion matrix for the best model (LDA (no feature selection), aggregated over outer folds

Actual/Predicted	Canindé de São Francisco	Gararu	Poço Verde
Canindé de São Francisco	7	1	1
Gararu	3	0	1
Poço Verde	0	1	9

Table 4 Global log–log OLS fits for allometry (Height vs Deq; Crown area vs Deq)

Response	n	Intercept (β0)	Slope (β1)	β1 95% CI	R^2	AICc
Height (m)	30	1.327	0.155	[0.046, 0.263]	0.233	−10.302
Crown area (m^2)	30	2.425	0.547	[0.346, 0.748]	0.525	26.716

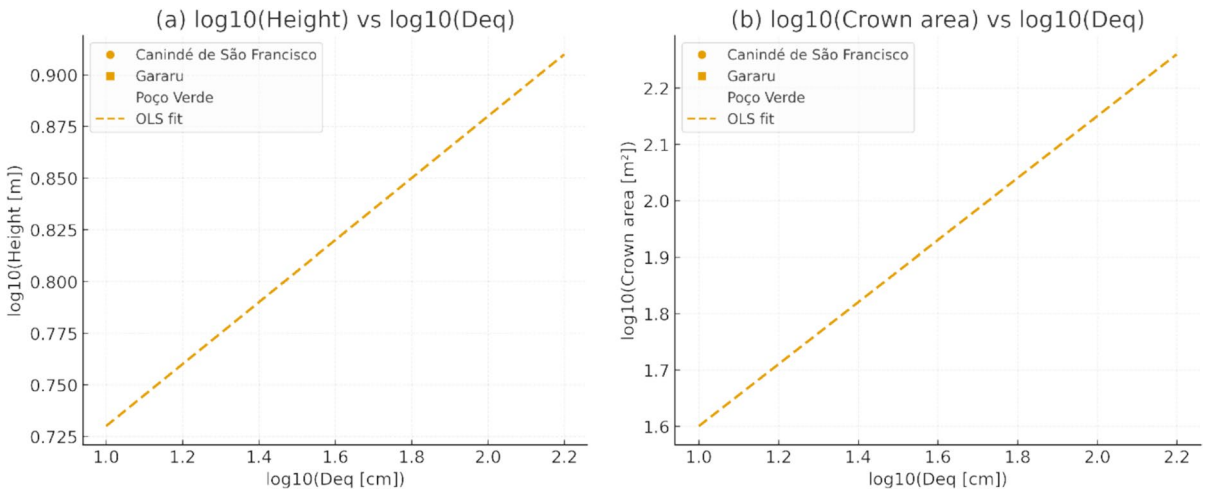


Fig. 12 Global log–log OLS fits (natural log). **a** ln(Height [m]) versus ln(Deq [cm]): slope b1 ≈ 0.155 (95% CI 0.046–0.263), intercept b0 ≈ 1.327. **b** ln(Crown area [m^2]) versus ln(Deq [cm]): slope b1 ≈ 0.547 (95% CI 0.346–0.748), intercept b0 ≈ 2.425

Discussion

Within the limited sample of 23 fruiting trees from three provenances and a single season, this study shows that provenance-level differentiation in *Spondias tuberosa* from Sergipe is expressed most clearly by image-derived color and shape descriptors, whereas classical size metrics (length, width, area) overlap broadly among sites. Similar patterns have been reported for other *Spondias* species and semi-arid fruit trees, where fruit shape and pigmentation vary more consistently with provenance or ecotype than absolute dimensions (Cavalcante et al. 2018; Santos et al. 2020; Silva et al. 2022). In this sense, our results align with the broader plant phenomics literature, which has emphasized geometric and color traits extracted from images as sensitive, low-cost proxies for structure and divergence when molecular resources or large progeny tests are limited (Furbank and Tester 2011; Araus and Cairns 2014; Li et al. 2014).

At the tree level, dendrometric traits help to interpret the landscape signal. Equivalent diameter (Deq) and crown area (CA) varied more consistently among provenances than total height, suggesting stronger sensitivity of lateral growth and canopy deployment to site conditions and possible local adaptation within this portion of the species' range (Poorter et al. 2012; Lima et al. 2019). The observed log–log allometry, with a shallow height–diameter slope and a steeper crown area–diameter slope, is consistent with general models for tropical deciduous trees (King 2011; Mensah et al. 2018) and supports the ecological interpretation that canopy expansion is favored over vertical elongation under seasonal drought stress. Wider and more robust trees improve the transport of nutrients to photosynthetic areas, promoting greater fruit production (Clement et al. 2021). These patterns strengthen confidence in the field measurements and highlight Deq and CA as operational criteria for mother-tree ranking in dry-forest environments similar to those sampled here.

Among fruit descriptors, texture showed moderate discriminative power, but its sensitivity to illumination and segmentation noise reinforces the need to pair it with more robust geometric ratios such as W/L (Berrezueta et al. 2019). For endocarps, W/L (round vs. elongated) and selected CIELAB components

were again the most responsive to provenance differences, echoing image-based research in *S. purpurea*, *S. mombin* and other drupe- or nut-bearing species (Silva et al. 2022; Borges et al. 2020). Together, these descriptors form a compact, reproducible panel that can be implemented in routine phenotyping campaigns using simple RGB imaging setups.

Methodologically, this work illustrates a "phenotyping-on-a-budget" approach in the sense of plant phenomics: an inexpensive RGB imaging system, a standardized but field-compatible acquisition protocol, and an analysis pipeline that respects the biological hierarchy (Furbank and Tester 2011; Cobb et al. 2013; Araus and Kefauver 2018). By treating the mother tree as the analytical unit, multiple fruits and endocarps are correctly handled as replicate observations of the same maternal genotype, avoiding pseudoreplication and producing variance components that are interpretable for genetic resources and prebreeding (Resende 2016; Piepho et al. 2015; Bolker et al. 2009). Mother-tree–stratified cross-validation in supervised classification further prevents information leakage across training and test sets and yields moderate but credible accuracy estimates (~70%) that reflect the partial overlap observed in PCA and univariate summaries. This level of performance is consistent with morphological discrimination among natural populations of outcrossing perennials (Ledig 1992; Hamrick and Godt 1996; Dawson et al. 2014) and should be viewed as a baseline for provenance-of-origin applications rather than as a fully validated certification tool.

From a plant phenomics and genetic resources perspective, the workflow is deliberately simple and accessible: a fixed overhead RGB camera, diffuse illumination, an in-frame scale and color reference, rule-based segmentation, and a small set of intuitive descriptors (Deq, CA, W/L, CIELAB). This combination makes the protocol reproducible across sites and realistic for resource-limited programs that manage native or underutilized fruit-tree germplasm, in line with calls for scalable, low-cost phenotyping platforms beyond highly instrumented facilities (Araus and Cairns 2014; Araus and Kefauver 2018). It also complements ongoing efforts to integrate phenotypic, environmental and genetic data in support of conservation and breeding of tree crops and neglected species (Engels et al. 2021).

In terms of conservation and pre-breeding, our findings translate into preliminary but actionable guidance for S. tuberosa in Sergipe. For traceability and quality control, CIELAB color channels—especially lightness and the green–red axis—combined with W/L provide a promising first screen to assist provenance identification in nurseries and extractive supply chains (Silva et al. 2023). For mother-tree selection and core collections, combining these image-derived descriptors with dendrometric information (Deq and CA) and basic fruit quality traits (such as total soluble solids and firmness) can improve representativeness in ex situ collections and support on-farm conservation decisions (Ribeiro et al. 2019; Santos et al. 2021). However, these recommendations are grounded in the specific populations and conditions sampled here and should not be generalized uncritically across the entire distribution of *S. tuberosa*.

Two main limitations define the scope of inference and priorities for next steps. First, the current dataset covers a single fruiting season, three provenances and 23 fruiting mother trees, providing a temporal and spatial snapshot of phenotypic variation. Multi-season and multi-site sampling, including additional provenances and larger numbers of mother trees, will be essential to disentangle climatic plasticity from heritable differentiation, quantify year-to-year stability of descriptors, and establish thresholds suitable for operational certification or breeding decisions (Dawson et al. 2014; Pennington et al. 2018). Second, although non-normality and heteroscedasticity were addressed via mixed models and variance components, future work should extend hierarchical modeling to partition additive genetic variance once larger progeny arrays or molecular marker data become available (Resende 2016; Piepho et al. 2015). Making raw images, extracted traits and analysis code publicly available under a persistent DOI will further facilitate reuse, benchmarking and integration with broader phenomics and genetic resource initiatives (FAO 2010; Engels et al. 2021).

Taken together, our results show that image-derived W/L and CIELAB color, embedded in a mother-tree–aware statistical framework, provide a robust and scalable path for phenotyping under-sampled native fruit genetic resources. The workflow exemplifies "phenotyping on a budget": it is technically simple, reproducible and compatible with the constraints of extractive production systems, yet rich enough to

inform curation, conservation planning and pre-breeding. With multi-season and multi-site validation in place, the same framework could underpin more routine origin verification and support the integration of low-cost plant phenomics into genetic resource management for *S. tuberosa* and related species.

Although our results show that image-derived phenotypic descriptors capture provenance structure and can support origin verification, full traceability in biological systems is generally strengthened by the integration of genetic variability, phenotypic characterization, and environmental data. Molecular markers have long been used to quantify intra- and inter-population diversity and to guide conservation and germplasm management in both crops and forest trees (Porth and El-Kassaby 2014; Ferreira 2005), and for *S. tuberosa* substantial genetic structure has been reported using AFLP and SSR markers (Santos et al. 2008; Balbino et al. 2019; Nunes-dos-Santos et al. 2021; de Sales et al. 2024).

Phenotyping, in turn, can distinguish wild, managed, or cultivated stands and has been shown to reflect underlying genetic patterns in natural populations (Silva et al. 2025), while environmental covariates are essential to interpret adaptive responses and climate-change vulnerability in tree populations (Alberto et al. 2013). In this context, the present study provides the low-cost, field-based phenotypic component, which can later be integrated with genetic and environmental information to increase the robustness of provenance-verification and conservation frameworks for *S. tuberosa*.

Conclusion

We propose a low-cost, image-based framework for characterizing *Spondias tuberosa* genetic resources that combines standardized field imaging with a mother-tree analytical design. In the sampled Sergipe provenances, CIELAB color channels and the shape ratio W/L captured provenance-level structure more consistently than raw size traits, while dendrometric allometry (Deq and crown area) provided additional criteria for ranking mother trees. These findings support an actionable descriptor panel centered on W/L and CIELAB, suitable for seed-lot traceability, conservation ranking and pre-breeding triage under field conditions. However, the current study is based on one fruiting season, three provenances and 23 fruiting mother trees, so provenance-of-origin discrimination should be regarded as a proof of concept that still requires multi-season, multi-site validation and larger sample sizes. Taken together, the workflow offers a promising basis for curation and quality control of under-sampled native fruit resources and a realistic starting point for future origin-verification tools.

Author contributions MSOS—Conceptualization; Methodology; Investigation; Data curation; Formal analysis; Visualization; Writing—original draft; Writing—review & editing. NAS—Investigation; Field sampling; Data curation; Resources; Writing—review & editing. VVN—Investigation; Resources; Validation; Writing—review & editing. MEOV—Software; Formal analysis; Visualization; Validation; Writing—review & editing. CCC—Investigation; Laboratory analyses; Data curation; Validation; Writing—review & editing. DACN—Writing—review & editing, Statistical review. RSM—Conceptualization; Supervision; Project administration; Funding acquisition; Methodology; Writing—review & editing.

Funding The Article Processing Charge (APC) for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001, through a scholarship awarded to the first author. R.S.M. is a productivity fellow of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil; PQ-2). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability The datasets generated and/or analyzed during this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no competing financial or non-financial interests that could have influenced the work reported in this manuscript. Fellowship and grant support are disclosed in the Funding section and do not constitute competing interests. There are no patents, products in development, or marketed products to declare.

Ethical approval Not applicable. Field sampling and use of plant materials complied with relevant institutional, national, and international guidelines.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aguiar FS, Brito FS, Gonçalves LMF, Pinheiro JMS, Rodrigues MLM, Mizobutsi GP, Oliveira GB (2023) Effect of alternative rooting inducers on umbu-cajazeira. Contrib Las Cienc Soc 16(10):24021–24031. https://doi.org/10.55905/revconv.16n.10-312

Alberto FJ, Aitken SN, Alía R, González-Martínez SC, Hänninen H, Kremer A, Lefèvre F, Lenormand T, Yeaman S, Whetten R, Savolainen O (2013) Potential for evolutionary responses to climate change: evidence from tree populations. Glob Change Biol 19(6):1645–1661. https://doi.org/10.1111/gcb.12181

Araus JL, & Cairns JE (2014) Field high-throughput phenotyping: the new crop breeding frontier. Trends in plant science, 19(1), 52-61. https://doi.org/10.1016/j.tplants.2013.09.008

Araus JL, & Kefauver SC (2018) Breeding to adapt agriculture to climate change: affordable phenotyping solutions. Current opinion in plant biology, 45, 237-247. https://doi.org/10.1016/j.pbi.2018.05.003

Balbino E, Martins G, Morais S, Almeida C (2019) Genome survey and development of 18 microsatellite markers to assess genetic diversity in *Spondias tuberosa* Arruda Câmara (Anacardiaceae) and cross-amplification in congeneric species. Mol Biol Rep 46(3):3511–3517. https://doi.org/10.1007/s11033-019-04768-w

Berrezueta E, Alvarado Y, Benavente D et al (2019) Representativity of 2D shape parameters for mineral particles in quantitative petrography. Minerals 9:768. https://doi.org/10.3390/min9120768

Bianchi VJ, Fajardo TVM, Morales RGF (2016) Multivariate analysis of seed morphology in tropical fruit trees. Genet Resour Crop Evol 63(8):1353–1365. https://doi.org/10.1007/s10722-015-0311-7

Bolker BM, Brooks ME, Clark CJ et al (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24:127–135. https://doi.org/10.1016/j.tree.2008.10.008

Brasil (2015) Lei nº 13.123, de 20 de maio de 2015, dispõe sobre o acesso ao patrimônio genético, sobre a proteção e o acesso ao conhecimento tradicional associado e sobre a repartição de benefícios para conservação e uso sustentável da biodiversidade. Presidência da República.

Cavalcante ACP, Cavalcante AG, da Silva Henrique J, Zuza JFC, Dantas MMM, da Costa Araújo R (2018) Organic substrate and organic foliar fertilizer effect on chlorophyll index and growth in Umbuzeiro seedlings. Agron Colomb 36(2):152–157

Clement CR, de Cristo-Araújo M, Coppens d'Eeckenbrugge G, Alves Pereira A, Picanço-Rodrigues D (2021) Domestication and diversification of Amazonian fruit trees: opportunities for on-farm conservation of agrobiodiversity. Front Ecol Evol 9:651459. https://doi.org/10.3389/fevo.2021.651459

Cobb JN, DeClerck G, Greenberg A, Clark R, & McCouch S (2013) Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics, 126(4), 867–887. https://doi.org/10.1007/s00122-013-2066-0

Cobb RE, Chao R, Zhao H (2013) Directed evolution: Past, present, and future. AIChE J 59(5):1432–1440. https://doi.org/10.1002/aic.13995

Consonni G, Gavazzi G, Dolfini S (2005) Genetic analysis as a tool to investigate the molecular mechanisms underlying seed development in maize. Ann Bot 96:353–362. https://doi.org/10.1093/aob/mci187

Dardick C, Callahan AM (2014) Evolution of the fruit endocarp: molecular mechanisms underlying adaptations in seed protection and dispersal strategies. Front Plant Sci 5:284. https://doi.org/10.3389/fpls.2014.00284

Dawson IK, Vinceti B, Weber JC, Neufeldt H, Russell J, Lengkeek AG, Kalinganire A, Kindt R, Lillesø J-PB, Roshetko JM, Jamnadass R (2014) Climate change and tree genetic resource management: Maintaining and enhancing the productivity and value of smallholder tropical agroforestry landscapes. For Ecol Manag 333:52–59. https://doi.org/10.1016/j.foreco.2014.01.021

de Sales RP, da Silva LC, da Silva Neves AG, Fajardo CG, Pinheiro LG, Vieira FA (2024) Addressing conservation needs: genetic diversity and population ecology of the endemic tree *Spondias tuberosa* Arruda. Scientifica Cairo 2024:5023974. https://doi.org/10.1155/2024/5023974

Dohm MR (2002) Repeatability estimates do not always set an upper limit to heritability. Funct Ecol 16:273–280. https://doi.org/10.1046/j.1365-2435.2002.00621.x

Engels JMM, Ebert AW (2021) A Critical Review of the Current Global Ex Situ Conservation System for Plant Agrobiodiversity. I. History of the Development of the Global System in the Context of the Political/Legal Framework and Its Major Conservation Components. Plants 10(8):1557. https://doi.org/10.3390/plants10081557

FAO (2010) The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Rome: Food and Agriculture Organization of the United Nations

Fernandes GW, Santos JC, Tabarelli M (2020) Phenotypic plasticity and ecological resilience in tropical dry forests. Biotropica 52(1):1–12. https://doi.org/10.1111/btp.12734

Ferreira ME (2005) Molecular analysis of gene banks for sustainable conservation and increased use of crop genetic resources. In Use of molecular markers for characterization and conservation of genetic resources. Food and Agriculture Organization of the United Nations (FAO). https://doi.org/https://www.fao.org/4/a0399e/A0399E08.htm

Furbank RT, Tester M (2011) Phenomics–technologies to relieve the phenotyping bottleneck. Trends Plant Sci 16(12):635–644

Gonzalez RC, Woods RE (2018) Digital image processing, 4th edn. Pearson

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Philosophical Transactions of the Royal Society B: Biol Sci 351(1345):1291–1298

Harrison XA, Donaldson L, Correa-Cano ME et al (2018) A brief introduction to mixed effects modelling and multimodel inference in ecology. PeerJ 6:e4794. https://doi.org/10.7717/peerj.4794

IBGE (Instituto Brasileiro de Geografia e Estatística) (2023) Produção da extração vegetal e da silvicultura (PEVS). https://doi.org/https://sidra.ibge.gov.br/pesquisa/pevs/quadros/brasil/2023. Accessed 3 Oct 2025

King DA (2011) Size-related changes in tree proportions and their potential influence on the course of height growth. In: Meinzer, F. C., Lachenbruch, B., & Dawson, T. E. (eds.), Size- and Age-Related Changes in Tree Structure and Function. Dordrecht: Springer, p. 165–191

Ledig FT (1992) Human impacts on genetic diversity in forest ecosystems. Oikos, 87-108

Lessells CM, Boag PT (1987) Unrepeatable repeatabilities: a common mistake. Auk 104:116–121

Li L, Zhang Q, Huang D (2014) A Review of Imaging Techniques for Plant Phenotyping. Sens 14(11):20078–20111. https://doi.org/10.3390/s141120078

Lima DC, Castricini A (2019) Propriedades físico-mecânicas de frutos e implicações para o processamento. Cienc Rural 49(6):e20180823. https://doi.org/10.1590/0103-8478cr20180823

Liu Y, Hou H, Ma J, et al. (2022) Fruit color genetics and genomics: from model species to crops. Hortic Res 9:uhac010. https://doi.org/10.1093/hr/uhac010

Matos FS, Freitas IAS, Pereira VLG, Pires WKL (2020) Effect of gibberellin on growth and development of *Spondias tuberosa* seedlings. Rev Caatinga 33(4):1124–1130. https://doi.org/10.1590/1983-21252020v33n427rc

Mensah S, Pienaar OL, Kunneke A, du Toit B, Seydack A, Uhl E, Pretzsch H, Seifert T (2018) Height–diameter allometry in South Africa's indigenous high forests: Assessing generic models performance and function forms. For Ecol Manag 410:1–11. https://doi.org/10.1016/j.foreco.2017.12.030

Motta TBN, Silva MR, Resende ED (2015) Quality attributes of tropical fruits related to texture and color. Sci Hortic 197:82–89. https://doi.org/10.1016/j.scienta.2015.09.041

Nunes-dos-Santos V, Santos CAF, Oliveira VR, Costa AES, Santos-da-Silva FF (2021) Diversity and genetic structure of *Spondias tuberosa* (Anacardiaceae) accessions based on microsatellite loci. Rev Biol Trop 69(2):640–648. https://doi.org/10.15517/rbt.v69i2.44194

Pennington RT, Lehmann CER, Rowland LM (2018) Tropical savannas and dry forests. Curr Biol 28(9):R541–R545. https://doi.org/10.1016/j.cub.2018.03.014

Phillip JM, Han KS, Chen WC, Wirtz D, Wu PH (2021) A robust unsupervised machine-learning method to quantify the morphological heterogeneity of cells and nuclei. Nat Protoc 16:754–774. https://doi.org/10.1038/s41596-020-00432-x

Piepho H-P (2018) Letters in mean comparisons: What they do and don't mean. Agron J 110(2):431–434. https://doi.org/10.2134/agronj2017.10.0580

Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer

Poorter H, Niklas KJ, Reich PB, Oleksyn J, Poot P, Mommer L (2012) Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control. New Phytol 193(1):30–50. https://doi.org/10.1111/j.1469-8137.2011.03952.x

Porth I, El-Kassaby YA (2014) Assessment of the genetic diversity in forest tree populations using molecular markers. Diversity 6(2):283–295. https://doi.org/10.3390/d6020283

Resende MDV (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. Crop Breed Appl Biotechnol 16:330–339. https://doi.org/10.1590/1984-70332016v16n4a49

Ribeiro LdeO, Viana EdeS, Godoy RLO, Freitas SC, Freitas SP, Matta VM (2019) Nutrients and bioactive compounds of pulp, peel and seed from umbu fruit. Ciência Rural 49(4):e20180806. https://doi.org/10.1590/0103-8478cr20180806

Russ JC (2018) The image processing handbook, 7th edn. CRC Press/Taylor & Francis

Santana JR, Souza VC, Oliveira FN, Silva PA, Ribeiro LM (2025) Phenotypic selection and diversity in *Spondias tuberosa* for breeding in drylands. Genet Resour Crop Evol 72(2):451–467

Santana NA, Nunes VV, Silva MSO, Silva-Mann R (2024) *Spondias tuberosa* Arr. Câm.: a natural population in Sergipe—Contributions to the understanding of genetic variability and conservation of the species. Research Square [Preprint]. https://doi.org/10.21203/rs.3.rs-3835713/v1

Santos CAF, Rodrigues MA, Zucchi MI (2008) Variabilidade genética do umbuzeiro no Semi-Árido brasileiro, por meio de marcadores AFLP [Genetic variability of umbu trees in Brazilian Semi-Arid region, based on AFLP markers]. Pesqui Agropecu Bras 43(8):1037–1043. https://doi.org/10.1590/S0100-204X2008000800013

Santos HSN, Mizobutsi GP, Pinheiro JMS, Neto JAS, Castricini A, Aspiazú I, Cruz CG, Sousa IPS, Jesus MO, Aguiar FS (2019) Physical and chemical characterization of fruits of different umbu (*Spondias tuberosa*) accesses. J Agric Sci 11(16):124. https://doi.org/10.5539/jas.v11n16p124

Santos RTS, Santos CAF, Oliveira VR, Aidar ST, Rybka ACP, Castro CDPC, Biasoto AC (2021) The potential of "umbuzeiro" genotypes for the development of preserves. Cienc Rural 51(2):e20191008. https://doi.org/10.1590/0103-8478cr20191008

Saraiva MMT, Martins LP, Araújo RDC, Cruz OND, Lucena FAD, Costa AD (2022) Variação colorimétrica e correlação de parâmetros biométricos entre matrizes de umbu. Rev Bras Frutic 44:e-479. https://doi.org/10.1590/0100-29452022479

Seymour GB, Østergaard L, Chapman NH, Knapp S, Martin C (2013) Fruit development and ripening. Annu Rev Plant Biol 64:219–241. https://doi.org/10.1146/annurev-arplant-050312-120057

Silva JHCS, da Silva JN, Rodrigues CM, da Silva ELF, de Almeida LGA, Bernardo MKF, do Nascimento KL, Ramos RS, do Nascimento NFF, Alves EU (2025) Phenotypic expression of diaspores indicates genetic diversity in natural populations of *Spondias tuberosa* (Anacardiaceae). Biology 14(12):1641. https://doi.org/10.3390/biology14121641

Silva SL, de Souza EVN, Couto LA, de Souza Moreira E, Landim LB, Rusciolelli LB, Cafieiro CSP (2022) Parâmetros proximais da polpa de umbu (Spondias tuberosa Arr. Cam.) comercializada no município de Guanambi, Bahia, Brasil. Braz J Sci 1(12):76–82

Sladoje N, Nyström I, Saha PK (2005) Measurements of digitized objects with fuzzy borders in 2D images. Image vis Comput 23:123–132. https://doi.org/10.1016/j.imavis.2004.06.011

Stange V, Martins AM, Silva JP, Clement CR (2021) Genetic diversity and conservation of tropical fruit tree species. Tree Genet Genomes 17(3):45. https://doi.org/10.1007/s11295-021-01516-3

Tamura Y (2024) Allometric scaling laws in fruit trees: implications for phenotypic plasticity. Ann Bot 133(4):561–573. https://doi.org/10.1093/aob/mcad023

Xavier VL, Feitoza GS, Barbosa JML, Araújo KS, Silva MV, Correia MTS, Souza MP, Carneiro-da-Cunha MDG (2022) Nutritional and technological potential of umbu (*Spondias tuberosa* Arr Cam) processing by-product flour. An Acad Bras Ciênc 94(1):e20200940. https://doi.org/10.1590/0001-3765202220200940

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.