



Using morphometrics to distinguish the restinga and cerrado ecotypes of wild *Anacardium occidentale* (Anacardiaceae): a preliminary study in northeastern Brazil

Juliana Cardozo Farias¹ · Simon Joseph Mayo² · Irlaine Rodrigues Vieira³ · Ivanilza Moreira de Andrade^{1,4}

Received: 23 March 2023 / Accepted: 26 November 2023 / Published online: 6 January 2024
© The Author(s) 2024

Abstract

The cashew tree *Anacardium occidentale* is a globally important food species and in Brazil is a natural component of cerrado and restinga vegetation and has an important role in local economies as a food and medicinal plant. Nevertheless, diversity studies of wild populations—the most important reservoir of natural variation—of this naturally widespread species are urgently needed for conservation strategies because of human impact on tropical American ecosystems where it occurs. Here, we sought to differentiate the previously recognized restinga and cerrado ecotypes using multivariate morphometrics of wild populations in a limited area of northeast Brazil (northern Piauí and northeastern Maranhão states). Thirty individuals were sampled for 47 vegetative and reproductive descriptors from six wild populations (three of each ecotype). Data analysis included principal component analysis and three classification methods (linear discriminant analysis, k-nearest neighbour analysis and classification and regression tree analysis); cross-validation tests were applied to classification results. Univariate statistics obtained for each population and ecotype were found to be consistent with the previously proposed distinction between the two ecotypes. Although principal component analysis showed extensive variation and overlap between populations, cross-validations of classification analyses showed strong differentiation between ecotypes and between most populations; vegetative and reproductive variables were equally important for discrimination. Cerrado populations were more variable and somewhat less distinct than restinga ones. New molecular genetic investigations of wild populations are needed to provide genetic support for the morphological differentiation of the ecotypes reported here.

Keywords Anacardiaceae · Cajuí · Conservation · Multivariate morphometrics · Population diversity

Introduction

Anacardium occidentale L. (cashew, caju, cajuí) is a species adapted to nutrient-poor soils and habitats that experience high temperatures and water stress (Serrano and Pessoa 2016). It belongs to the plant family Anacardiaceae and was last revised taxonomically by Mitchell and Mori (1987). Because of the high nutritive value of its seeds it was probably distributed widely in neotropical regions by human communities prior to 1500 (Mitchell and Mori 1987). Today this species has global economic importance and has spread by human agency to all regions of the tropics and subtropics, both cultivated in plantations and as naturalized self-reproducing populations, for example in India (Archak et al. 2009). It is characterized by the fruit ("double fruit" of Mitchell and Mori 1987) in which the peduncle develops into a yellow to red, fleshy, sweet and adstringent hypocarp

Handling Editor: Mike Thiv.

✉ Simon Joseph Mayo
simonjosephmayo@gmail.com

- ¹ Programa de Pós-Graduação em Desenvolvimento e Meio Ambiente, Universidade Federal do Piauí, Teresina, Piauí, Brazil
- ² Herbarium, Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AE, UK
- ³ Coordenação de Laboratórios, Universidade Federal do Delta do Parnaíba, Parnaíba, Piauí, Brazil
- ⁴ Herbário Delta do Parnaíba (HDELTA), Universidade Federal do Delta do Parnaíba, Parnaíba, Piauí, Brazil

(cashew apple, *maçã*) supporting a hard-walled drupe (nut, *castanha*).

In Brazil, the agronomic production of cashew is concentrated in the Northeast region (Brainer 2021), where it began in the 1950s (Serrano and Pessoa 2016). In this region, in addition to the importance of the drupe as food, the hypocarp is used in juices, sweet conserves, cordials (*cajuína*), wine (Serrano and Pessoa 2016) and beer (Pereira et al. 2020). Another cashew product is cashew nut shell liquid (CNSL), used in industry and in research as a repellent and insecticide and for many other applications (Brainer and Vidal 2018; Silveira et al. 2019).

For centuries *A. occidentale* has been widely cultivated, selected and manipulated to obtain desirable characteristics, and random crossing has taken place between cultivated and wild populations, with naturalization of their descendants. These processes have resulted in great difficulty in determining the natural area of distribution of the species and for determining the natural patterns of infraspecific variation (Mitchell and Mori 1987; Andrade et al. 2019). According to Mitchell and Mori (1987), natural populations of *A. occidentale* occur in the cerrado, in littoral regions in restinga vegetation and in transitional areas with caatinga. These wild plants are commonly differentiated from domestic forms by their smaller fruits weighing no more than 3 g (Barros et al. 2002), and they are called *cajuí* instead of *caju* (Andrade et al. 2019). The wild fruits are collected and used by local populations both in the littoral region and in the interior (Farias et al. 2020).

The need to study the morphological and genetic variation of wild populations of *A. occidentale* (*cajuzeiro*) has become more urgent with growing human pressures within the ecotonal region of Piauí and Maranhão, not only because of the extractive harvesting of the fruits, but also because of the destructive collection of wood and even the clearing of individuals to enable other kinds of land use (Rufino et al. 2007). Uncontrolled extractive activities can alter the morphological characteristics and impact on reproductive and vegetative structures of plants (Vieira et al. 2016). The resulting smaller population sizes also reduce genetic variability and capacity for adaptation to future environmental changes, impacting the natural bank of variability of a species of great economic importance. There are to date relatively few molecular genetic studies of wild *A. occidentale* populations (Borges 2015; Borges et al. 2018; Santos et al. 2019; Gomes et al. 2021), and the possible erosion of its genetic variability by extractivism remains little studied. In the related cerrado species *A. humile* A.St.-Hil. there are a similar number of genetic studies of wild populations (Carvalho et al. 2012; Cota et al. 2017; Santos et al. 2020).

Brazil is considered to be the most important centre of natural diversity of the species (Mitchell and Mori 1987; Barros and Crisóstomo 1995; Silva-Luz et al. 2023), and

the morphological variation observed there has led to the description of several distinct species (e.g. *A. microcarpum* Ducke and *A. othonianum* Rizzini) in cerrado, caatinga and restinga vegetation (see Andrade et al. 2019 for a more detailed discussion). However, in the opinion of taxonomic specialists of the genus (Mitchell and Mori 1987; Silva-Luz et al. 2023) and many ecologists who study cerrado vegetation (Castro et al. 2009; Castro 2013; Ratter et al. 2017), these names are all synonymous with *A. occidentale* and these authors recognize only a single variable species.

In Piauí and in eastern Maranhão there is a diversity of vegetation types including cerrado, caatinga, restinga and various transitional types (Castro 2013), in which extensive natural populations of *A. occidentale* trees occur, locally known as *cajuzeiros*. These two Brazilian states thus offer a suitable context for a study of infraspecific variation. Various research studies have been carried out in this region primarily aimed at differentiating natural restinga populations (restinga ecotype) from domesticated ones using morphometrics (Andrade et al. 2019), genetic variation (Santos et al. 2019) and bioactivity (Ribeiro et al. 2021). Souza et al. (2021) studied the ecology of wild populations in the restingas of Piauí in relation to dune migration.

Morphometric methods have previously been used in studies of *A. occidentale* for recognition and classification of accessions in germplasm collections representing different genotypes (see references in Andrade et al. 2019). Morphometric techniques have been shown to be very useful generally for delimitation of species (Marhold 2011), providing support for conservation inventories, recuperation of plant communities, species cataloguing and field identification (Barreto and Ferreira 2011).

Commenting on the lack of systematic studies of the infraspecific variation of wild populations of *A. occidentale*, Mitchell and Mori (1987) proposed an informal subdivision of the species into two kinds of wild cashews: the restinga ecotype occurring in coastal areas in eastern and northeastern Brazil, and the cerrado ecotype occurring in the savannas of central and Amazonian Brazil, Colombia, Venezuela and the Guianas as well as in caatingas transitional to cerrado in northeastern Brazil (Andrade et al. 2019). For Mitchell and Mori (1987), the main differences between the two ecotypes are morphological characters of the leaves, the size and acidity of the fruits, and the architecture of the canopy (Mitchell and Mori 1987 p. 40). Hitherto, however, this hypothesis has not been tested by more detailed analytical methods. Our objective in the present study was thus to verify whether the two ecotypes could be recognized by morphometric study of wild populations and determine their most diagnostic characters.

Material and methods

Study design

This study compared 180 individuals from six wild populations, three from restinga areas and three from cerrado areas. Two partition models of the individuals were evaluated in the analyses: (1) the six populations, justifiable as natural groups based on population genetic theoretical grounds (Picó et al. 2021: 4); (2) the two ecotypes, modelling the hypothesis of Mitchell and Mori (1987), were represented by three populations each, the populations being assigned to ecotypes based on their geographical location (littoral zone/interior of continent) and vegetation type (restinga/cerrado-caatinga transition).

The analyses sought to quantify how well the individuals would partition into the two ecotypes and six populations using cross-validation procedures applied to classification algorithms and also establish the relative similarities of the six populations using cluster analysis of population mean vectors. Diagnostic morphological characters of the two ecotypes were determined using univariate methods.

Field collection authorization

This research is registered in the National System of Genetic Patrimony and Associated Traditional Knowledge (SIS-GEN) with the registration code A8B44BB. The collecting license was authorized by the System of Authorization and Information on Biodiversity (SISBIO) of the Chico Mendes Institute for the Conservation of Biodiversity—ICMBIO (n° 64340–1).

Geography and ecology of sample sites

The association of the cerrado and restinga vegetation types with the division of *A. occidentale* into two ecotypes was clearly intended by Mitchell and Mori (1987) to be broadly interpreted; thus their Table IV (p. 40) uses the terms "Coastal Brazil" for the restinga ecotype and "Central Brazil" for the cerrado ecotype. Our interpretation of the ecotypes follows this looser terminological meaning. Both cerrado and restinga are very heterogeneous vegetationally and are understood by ecologists as variable domains rather than highly specific formations (e.g. Castro 2013; Santos-Filho 2009; Santos-Filho et al. 2010; Santos-Filho and Zickel 2013). Andrade et al. (2019) experimented with further subdivisions by categorizing one set of coastal populations as "dune" because of their occurrence on Quaternary dune systems and their highly characteristic dwarfed physiognomy, but there are other coastal populations located on

Tertiary geology which are taller and yet exposed to maritime influences. We have termed all these as restinga ecotype populations. Similarly, in the case of the cerrado ecotype, we have grouped under this heading populations from typical cerrado and those from vegetation showing clear transitional features to caatinga.

The distance between study sites varied from a minimum of 13 km to a maximum of 250 km. The three restinga sites are located in the municipalities of Araiões (Maranhão state) and Parnaíba and Cajueiro da Praia (both in Piauí state). The cerrado study areas are located in the municipalities of Campo Maior, José de Freitas and in the Sete Cidades National Park (Fig. 1, Table 1), all in Piauí state. The climate over the study region is Aw type in the Köppen-Geiger classification (Kottek et al. 2006), characterised by a hot rainy summer season, a dry winter season, highest rainfall between January and June and mean annual rainfall varying between 1000 and 1600 mm (Lima et al. 2017).

All the restinga sites lie within the Environmental Protection Area of the Rio Parnaíba Delta, which extends over 311,731 hectares with a perimeter of 1,473 km and includes coastal municipalities of the states of Maranhão, Piauí and Ceará, with altitudes from 0 to 45 m (Ivanov 2020). The geomorphology of the Canárias and Labino areas consists of Quaternary formations of quartzitic sands (Jacomine et al. 1986), whereas that of the region further east, where Cajueiro da Praia is situated, is dominated by the Tertiary formation of Barreiras sandstones (MME 2006). Besides restinga, the vegetation of the Delta includes caatinga, cerrado, mangrove and tabuleiro plant communities (Fernandes et al. 1996; Castro 2007) and vegetation transitional between cerrado and caatinga which is typical of a large part of the northern region of Piauí (Santos-Filho et al. 2010).

The cerrado localities lie in the continental interior and cover a wider area, but all within the northern part of Piauí state (Fig. 1). The Sete Cidades National Park has an area of 6,221 ha and a perimeter of 36 km and is a Federal Conservation Unit of Total Protection, located between the municipalities of Piracuruca and Brasileira (04°02'–08' S, 41°40'–45' W) (ICMBIO 2018), between altitudes of 100 and 290 m. The vegetation is a mosaic of various types of cerrado, and the sample site lies within cerrado in the strict sense, which occupies the greatest area within the Park (Oliveira et al. 2010). The Campo Maior site lies within the Campo Maior microregion, which has an area of 1,657 km² with vegetation transitional between cerrado and caatinga and including gallery forests with carnaúba (*Copernicia prunifera* (Mill.) H.E. Moore), periodically flooded caatinga, hyperxerophilous caatinga, sub-deciduous cerrado, sub-deciduous forest and carrasco (Farias and Castro 2004). The sampled area was in a caatinga-cerrado transitional vegetation with greater similarity to typical cerrado (Sousa et al. 2021). The José de Freitas site lies



Fig. 1 Sampling localities for wild *Anacardium occidentale* (cajuí): restinga and cerrado ecotypes

Table 1 Collecting localities, sample sizes and voucher specimens for each study population

Ecotype	State	Municipality, locality	Latitude and longitude	N	HDELTA
Restinga	Maranhão	Araioses, Ilha das Canárias	02° 45' 55.503" S, 41° 51' 15.814" W	30	5829
	Piauí	Parnaíba, Labino	02° 50' 39.981" S, 41° 45' 45.0793" W	30	7007
	Piauí	Cajueiro da Praia, Barrinha	02° 54' 58.124" S, 41° 23' 28" W	30	7008
Cerrado	Piauí	Parque Nacional de Sete Cidades	04° 5' 31.692" S, 41° 40' 49.158" W	30	7010
	Piauí	Campo Maior	04° 59' 0.340" S, 42° 14' 0.062" W	30	6670
	Piauí	José de Freitas	04° 48' 49.815" S, 42° 36' 4.987" W	30	7009

N number of individuals sampled. HDELTA specimen registration number at HDELTA herbarium, Parnaíba, Piauí

within the eponymous municipality in the Teresina micro-region, which extends over an area of 1,538 km² (IBGE 2018; Aguiar and Gomes 2004) with vegetation similar to that in the Campo Maior region; the sample site was similarly selected in an area of more typical cerrado.

Field sampling

Field collections were carried out mainly in the dry season (July to December) of the years 2018, 2019, 2020 and 2021 when this species flowers and fruits. Each population was

visited twice for field observation and collection of flowering and mature fruiting material. The same marked trees were sampled on successive visits.

For the morphometric analyses, thirty individuals from each of the six populations were sampled by traversing each study area and selecting trees more than 10 m apart for collection of sample material. From each individual, five branches that included leaves, flowers and fruits were selected, distributed as far as possible in different parts of the canopy, making a total of 150 branches from each population. The branches were sealed in plastic bags at the collecting sites and then transferred to a refrigerator at the HDELTA herbarium in the Federal University of the Parnaíba Delta, Parnaíba, Piauí (UFDPAR). Digital images of

bark, leaves, flowers and fruits (Figs. 2, 3) were made in the field.

Voucher herbarium specimens representing each population with leaves, flowers and fruits were collected separately in each sampling area. The specimens were prepared according to the methodology of Mori et al. (1989) and incorporated into the HDELTA herbarium at the Ministro Reis Velloso Campus of the Federal University of the Parnaíba Delta (UFDPAR) in Parnaíba, Piauí, Brazil (Table 1).

Morphometric characters and data collection

Measurements of plant height, crown width, trunk length and perimeter, and recording of internal bark colour were all

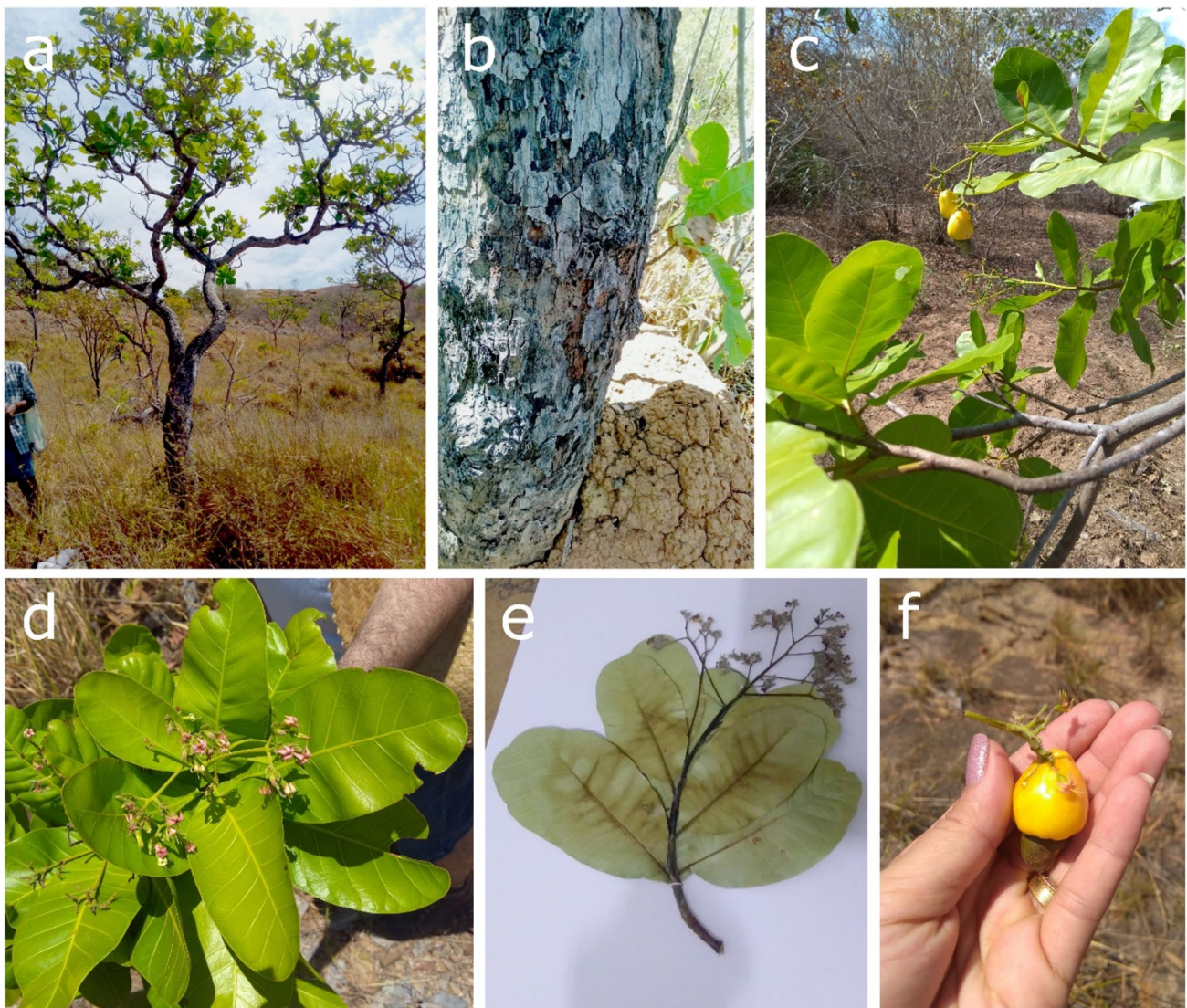


Fig. 2 Vegetative and reproductive organs used in the morphometric study of *Anacardium occidentale* (cajuí). Cerrado ecotype in Serra de Passa-Tempo, Campo Maior, Piauí. **a** General view of habit. **b** Colour

of trunk. **c** Fruits showing yellow hypocarps. **d** Paniculate inflorescence. **e** Preparation of herbarium specimen voucher. **f** Measurement of hypocarp and drupe



Fig. 3 Vegetative and reproductive organs used in the morphometric study of *Anacardium occidentale* (cajuí). Restinga ecotype in Labino, Piauí. **a** General view of habit. **b** Paniculate inflorescences and leaves. **c** Trunk. **d** Colour of trunk and bark slash

made in the field. Plant height was measured with a measuring rod subdivided into metre units (Fig. 3a). A metric measuring tape was used to measure trunk perimeter. Measurement and recording of all other variables and additional photography (Online Resource 1a) were carried out in the

HDELTA herbarium, where a millimetre ruler was used for dimensions of the leaf, flower, inflorescence, hypocarp and drupe and a pachymeter was used for measuring leaf thickness. The leaves and inflorescences were pressed and dried and then measured from dried material. The flowers were

heated in distilled water and measured once rehydrated. The fruit (drupe and hypocarp) was measured using fresh material or material kept in a refrigerator. Each of the five replicate branches of each sampled individual was examined in turn. One mature leaf at the third or fourth node below the inflorescence was selected for measurement. Secondary vein number on the left and right sides of the leaf blade was recorded separately and then combined into the single variable of total number of secondary veins per leaf. Five masculine and five hermaphrodite flowers and five mature compound fruits (drupe and hypocarp) were measured from each individual. The terminology of vegetative and reproductive characters was based on that of Mitchell and Mori (1987). A total of 42 quantitative and five qualitative characters were recorded (Table 2, Online Resource 1g). The qualitative characters were selected based on information from local informants who collect fruits from wild populations, and we added them as they were said to be indicative of different variants. They are internal bark colour (red/brown/red), leaf blade overall shape (elliptic/obovate), apex shape (acute/rounded/emarginate), base shape (acute-convex, acute-cuneate, obtuse) and hypocarp colour (yellow/red).

Data analysis

The univariate analyses, based on all recorded 42 quantitative variables (Online Resource 1g), showed that most variables had non-normal distributions and heterogeneous variances between populations (Online Resource 1b, Online Resource 2). Because bract, style and pistil length showed no significant difference between either populations or ecotypes, they were not used in the data set for multivariate analyses (Online Resource 1g). On the other hand, variables that showed no significant difference between ecotypes but did show significant differences between populations were included in the multivariate analyses (Table 2, Online Resource 2).

The data sets used for multivariate analyses differed according to the method used, each of which has its own limitations. For principal component analysis, linear discriminant analysis of ecotypes and k-nearest neighbour analysis the matrix of 39 quantitative variables was used (Table 2). For linear discriminant analysis of populations a reduced matrix of 29 quantitative variables was used (Table 2 indicates the ten characters not used), in order that the number of variables was less than the sample size of the populations ($N=30$ in all cases). For the classification and regression tree analysis the five qualitative variables were added to the 39 quantitative ones (Table 2). In the data matrices used for multivariate analysis, the mean value of the five replicate measurements was used for quantitative variables, and the modal state for qualitative ones since the character states were non-numerical.

The data analyses were based on those described by Andrade et al. (2019) and implemented using the statistical language R (R Core Team 2018). Univariate analyses were carried out at three levels: entire data set, the two ecotypes and the six populations. These determined extreme values, distribution, mean, median, maximum and minimum values, standard deviation (SD), standard error of the mean (SE) and coefficient of variation (CV). Tests were computed at each level for normality, variance homogeneity and difference of means and medians (t-test, Wilcoxon rank sum, multiple comparisons). The results of these analyses were used to determine the most important diagnostic characters of the predetermined ecotype groups (Online Resource 2).

The major dimensions of variation of the overall data set were determined using principal component analysis (PCA) of the 39 quantitative variables (Table 2), and this showed which variables were most influential on each PC (principal component). The significance of the principal components was computed using the function *evplot* (Borcard et al. 2011: 121–124) which applies the Kaiser-Guttman test and the broken-stick model to provide two views of the number of PCs to be interpreted. The estimation of the minimum number of informative principal components in a PCA is inherently empirical (for further discussion see Borcard et al. 2011; Legendre and Legendre 2012). The Kaiser-Guttman and broken-stick methods usually give different results, the former being generally more generous than the latter, and thus together they serve as a guide to the likely range of interpretable PCs. The PCA was carried out using the function *prcomp* in the *stats* package in R (R Core Team 2018) with scaled data (argument: `scale=TRUE`).

We used three different classification methods (LDA, KNN, CART) in exploratory mode to obtain multiple views of the similarity relationships of the ecotype and population categories. We evaluated the consistency of these a priori categories using nonparametric methods, that is, the degree to which the individuals of each category grouped together in each analysis. For LDA and KNN, the results were quantified using leave-one-out cross-validation (James et al. 2013). This is a procedure in which each individual in turn is removed from the matrix and treated as a test set (James et al. 2013); the classification method (LDA, KNN, CART) is then deployed to classify the reduced matrix and the test individual is assigned to the nearest predicted population or ecotype. For each such category, the percentage of erroneous assignments of its individuals (i.e. to populations or ecotypes different to those to which they belong) indicates its degree of morphological distinctness, and also reveals, in the destinations of the misclassified individuals, the other categories with which it overlaps, according to the classification method used. For CART, the consistency of the population and ecotype categories was evaluated by computing

Table 2 Morphometric analysis of wild *Anacardium occidentale*

Plant part	Character	Type	Code
Whole plant	Height of individual (m)	Quantitative	altura
Tree crown	Width of crown (m)	Quantitative	copa-larg
Stem	Length of trunk to first branch (cm)	Quantitative	tronc-alt
	Circumference of trunk at ground level (cm) §§	Quantitative	tronc-circ-bas §
	Circumference of trunk at breast height (cm)	Quantitative	tronc-circ-peit
	Internal bark colour**	Qualitative	casc-int-cor
	Blade thickness* (mm) §§	Quantitative	lam-espes-med §
Leaf	Blade length* (cm)	Quantitative	lam-com-med
	Blade maximum width* (cm)	Quantitative	lam-lar-med
	Blade shape**	Qualitative	lam-form-mod
	Blade aspect ratio (length/ maximum width)*	Quantitative	lam_asp_med
	Blade apex shape**	Qualitative	lam_apic_mod
	Blade base shape**	Qualitative	lam_bas_mod
	Secondary veins, total number*	Quantitative	nerv_num_tot_med
	Petiole length* (mm) §§	Quantitative	pec_com_med
	Petiole width* (mm)	Quantitative	pec_lar_med
	Ratio blade length/petiole length*	Quantitative	lam_pec-tax_med
Inflorescence	Inflorescence length (cm) §§	Quantitative	infl_com §
	Inflorescence width (cm) §§	Quantitative	inf_larg §
	Rachis length (cm)	Quantitative	raq_com
	Number of branches in the panicle	Quantitative	inf_rams
Male flower	Pedicel length* (mm) §§	Quantitative	fmas_ped_com_med §
	Calyx length* (mm)	Quantitative	fmas_cal_com_med
	Corolla length* (mm)	Quantitative	fmas_coro_com_med
	Corolla diameter* (mm)	Quantitative	fmas_coro_dia_med
	Stamens, total number*	Quantitative	fmas_esta_num_med
	Staminal tube length* (mm)	Quantitative	fmas_tub_esta_com_med
	Short stamen length* (mm)	Quantitative	fmas_estacur_com_med
	Long stamen length* (mm)	Quantitative	fmas_estalon_com-med
Hermaphrodite flower	Pedicel length* (mm) §§	Quantitative	fher_ped_com_med §
	Calyx length* (mm)	Quantitative	fher_cal_com_med
	Corolla length* (mm)	Quantitative	fher_coro_com_med
	Corolla diameter* (mm) §§	Quantitative	fher_coro_dia_med §
	Stamens, total number*	Quantitative	fher_esta_num_med
	Staminal tube length* (mm)	Quantitative	fher_tub_esta_com_med
	Short stamen length* (mm) §§	Quantitative	fher_estacur_com_med
	Long stamen length* (mm) §§	Quantitative	fher_estalon_com_med §
	Ovary length* (mm)	Quantitative	fher_ov_com_med
Fruit	Hypocarp + drupe length (pseudofruit)* (cm)	Quantitative	frutpedu_com_med
	Drupe length* (cm)	Quantitative	cast_com_med
	Drupe width* (cm)	Quantitative	cast_lar_med
	Hypocarp length* (cm)	Quantitative	hipo_com_med
	Hypocarp width* (cm)	Quantitative	hipo_lar_med
	Hypocarp colour**	Qualitative	hipo_cor_mod

Data sets used in multivariate analyses: CART used all 44 variables (5 qualitative, 39 quantitative); PCA, LDA of ecotypes and KNN used the full set of 39 quantitative variables; LDA of populations used only 29 quantitative variables (see text)

* Quantitative variable: analyses used mean value of 5 replicate values from the same individual.** Qualitative variable: analyses used modal value of 5 replicate character states from the same individual. § No significant difference between ecotype medians. §§ Variable not used in LDA of populations

confusion matrices which showed the percentage erroneous assignment of individuals to their original categories.

Linear discriminant analysis is a very widely used technique based on covariance matrices of the data, which seeks the maximum separation of the centroids of each a priori group (populations, ecotypes) while minimizing within-population variance. It uses eigenanalysis to produce axes (the discriminant functions) on which the individuals can be plotted, analogous to PCA. Despite the fact that our data did not meet the LDA assumptions of multivariate normality and homogeneity of covariance matrices, we used it because one result of its model is to show which variables are most influential in separating the groups and thus a useful guide for other studies. Preliminary analyses showed that the populations were rather distinct using the nonparametric techniques (KNN, CART) and because the results of the three techniques were broadly similar, this justifies the use of LDA, while requiring a critical attitude to results which diverge noticeably from other techniques and could thus be artefactual. In the LDA of populations, the 39 quantitative variables were reduced to 29 (Table 2) by deleting the ten variables with the least significant univariate difference between ecotypes (Online Resource 1g); this was to ensure that the number of variables was less than the sample size of the smallest group (30 individuals), following Legendre and Legendre (2012). The homogeneity of the covariance matrices of the populations was tested using functions from the *vegan* package (Oksanen et al. 2018), and the LDAs were carried out using the *lda* function in the MASS package (Ripley et al. 2022). Scaled (as in PCA) data were used to compute the discriminant function axes and unscaled data to carry out the cross-validation analysis (following Legendre and Legendre 2012: 675–676). The function ELLI (Claude 2008: 85) was used to plot 95% confidence ellipses around each population in the ordination plots. For the LDA with ecotypes as categorical variable, the full quantitative data set of 39 variables was used. Leave-one-out cross-validation was carried out using the $CV = TRUE$ argument in the *lda* function. The results of the cross-validation were visualized using a bar chart in which each individual is represented by a vertical bar and the posterior probability of belonging to different populations is indicated by the number and length of the coloured segments of each bar.

The k-nearest neighbour (KNN) analysis used all 39 quantitative variables (Table 2) to classify each individual into populations and ecotypes. The method computes the Euclidean distance between each individual and all others and selects its k -nearest neighbours. The individual is then allocated to the population which has the majority representation among these k neighbours. The data set was scaled, following Varmuzer and Filzmoser (2009). The optimal value of k (the neighbourhood) was first established by testing a range of values ($k = 1$ to 40 or

60) and determining which k number produced the smallest number of allocation errors. This test was carried out separately for the two categorical variables used—populations and ecotypes. Once the optimal k number was established, the result of the KNN analysis was tested by cross-validation. The KNN analyses were carried out using the function *knn.cv* in the *class* package (Ripley and Venables 2022).

The classification and regression tree analysis (CART) data set included five qualitative variables together with the 39 quantitative variables used in the PCA, LDA and KNN analyses (Table 2). The algorithm for classification trees was used; this divides the whole set of individuals (the root node) by each variable and selects that which produces the least impure (heterogeneous) pair of subgroups (nodes); node impurity was measured using the Gini index (Varmuza and Filzmoser 2009; Foulkes 2009). The two subgroups are then each divided in the same way into two further subgroups after a new search for the variable which results in the least impure subgroups and this procedure continues until some stopping point is reached, e.g. terminal groups (leaves) containing an arbitrary minimum number of individuals. Each terminal group is named by the population which is represented by the greatest number of individuals. The final number of nodes in the tree is determined by an optimization procedure called cost complexity pruning (Foulkes 2009), which finds an optimal number of nodes by reconciling maximum node purity (minimum level of heterogeneity) with the minimum number of terminal nodes (maximum tree simplicity). The CART analyses were carried out using the *rpart* (Therneau and Atkinson 2019) and *rpart.plot* (Milborrow 2020) packages.

Population variability was estimated as the average value computed from the coefficients of variation of each of the 39 quantitative variables in each population. The significance of the morphological differences between populations and ecotypes was computed using the *adonis* function in package *vegan* (Oksanen et al. 2018) which carries out a non-parametric multivariate analysis of variance (NPMANOVA). The multiple comparisons of the populations based on the NPMANOVA were computed using the *pairwise.perm.manova* function of the *RVAideMemoire* package (Hervé 2018).

Hierarchical cluster analysis (UPGMA: unweighted pair group method with arithmetic mean) was carried out on a matrix of population mean vectors computed from the data set of 39 quantitative variables. This was to provide a separate view of the similarities of the six populations and to verify if they formed two clusters corresponding to the ecotypes. A Euclidean distance matrix was computed using the function *dist*, and the tree was computed with the function *hclust* using the method argument "average". Both functions belong to the R base package *stats*.

The classification of the populations into the two ecotypes was also tested by running cross-validated KNN analyses on all ten distinct partitions of the six populations into two groups of three and comparing the error rates of the partitions. A range of other two-group partitions were also tested. R scripts were written for these analyses (see Online Resource 3 for all R scripts used in this study).

Results

Univariate diagnostics of the ecotypes

Tests showed significant differences between ecotypes in 31 of the quantitative variables (Online Resource 1g). Twenty-two of the quantitative variables were found to differ between the ecotypes with high significance (Table 3). Of these, the eight vegetative variables showed higher values in the cerrado ecotype and the 14 flower and fruit variables had higher values in the restinga ecotype. To mention only the differences of greatest statistical significance, the cerrado populations had longer trunks, longer leaf blades, wider petioles and higher ratio of leaf blade length to petiole length; the restinga populations had larger hypocarps and drupes and more stamens per flower (in both flower types).

Multidimensional variation

The major univariate differences were reflected in the principal component analysis. The most significant principal components ranged between the first five (broken-stick model) and the first ten (Kaiser-Guttman test). The first ten principal components expressed only 71.37% of the total variance (the first five expressed 53.81%), indicating a relatively low level of correlation among the 39 quantitative variables. The ordination of the first two principal components (Online Resource 1c), representing 32.42% of total variance, suggested that within-population variability is lower in restinga populations (the 95% confidence ellipses are smaller) than in the cerrado. The biplot (Fig. 4) and the loadings (Online Resource 1d) showed that vegetative and floral characters are equally important in the overall variation of the data set.

On PC1, individuals with the greatest positive scores (mostly restinga plants) tend to have longer and wider hypocarps and drupes and in general floral variables with higher values. Individuals with negative scores (predominantly from the cerrado) tend to have longer leaf blades, wider petioles, taller trunks and leaf blades which are relatively longer than wide, among other characters (Online Resource 1d). On the PC2 axis the characters which have positive scores (mostly restinga populations and Sete Cidades) tend to have more stamens in the masculine and hermaphrodite flowers, longer petioles, wider inflorescences

and tree crowns, wider hypocarps and longer inflorescences, while those with negative scores have larger leaf blade length/petiole length ratios, longer leaf blades, greater corolla diameter in hermaphrodite flowers, higher total number of leaf secondary veins and wider petioles.

Classification of the populations and ecotypes

Populations

The distinctness of the populations was evaluated by linear discriminant analysis (LDA), and their separation on the first two discriminant function axes is shown in Fig. 5.

In this two-dimensional space, presenting 80.0% of total variance, the three cerrado populations overlap and together are partially separated from the restinga group, of which Barrinha is the most distinct. The greatest overlap between ecotypes is between the Labino (restinga) and José de Freitas (cerrado) populations. The results of the cross-validation test are shown in Table 4 and Fig. 6 and show a total of 22 misclassifications (12.2% of total) among the populations.

The proportion of correctly classified individuals (bold values in Table 4) is lowest in Campo Maior and Labino, which are the populations showing greatest interpopulation overlap in Fig. 5. However, while Campo Maior (cerrado) has a single individual allocated to a restinga population, Labino (restinga) has misclassifications to all other populations except Sete Cidades, also reflected in the relative positions of the populations in the ordination of Fig. 5.

The most distinct populations are Sete Cidades and Barrinha, in which the mis-allocated individuals are to other populations of the same ecotype (Fig. 6). Table 4 and Fig. 6 also show that the overlap of José de Freitas with Campo Maior is lower than with Labino, contradicting the impression given in Fig. 5. This probably results from distortion introduced by the two-dimensional ordination, whereas the cross-validation uses all five discriminant functions of the LDA. Figure 6 expresses the similarities of individuals in more detail than Table 4 since individual bars with two or more colours reveal the range of populations to which they are most similar.

The KNN cross-validation analysis (Online Resource 1h, e) generated 28 errors overall (15.6%), among the populations. The KNN and LDA analyses were similar at population level regarding the higher consistency of the Sete Cidades and Barrinha populations, but the KNN result differed in the lower consistency of José de Freitas and the higher consistency of Labino. As in the LDA, Campo Maior showed relatively low consistency and Canárias had an intermediate level. In the LDA and KNN analyses, the Labino population as reconstituted by cross-validation showed the greatest mixture of individuals from other populations (Table 4, Fig. 6, Online Resource 1h and 1e).

Table 3 Univariate comparison of the 22 most diagnostic quantitative variables of the cerrado and restinga ecotypes

Variable	Mitchell and Mori (1987)		Present study		Significance of difference between ecotypes
	Cerrado*	Restinga*	Cerrado*	Restinga*	
Trunk length (cm)			164.6	40.9	Kruskal–Wallis: $\chi^2 = 92.183$, $p < 0.001$ *** (1.015e–05)
Leaf blade length (cm)	13.4	11.4	12.18	10.12	ANOVA: $F = 68.34$, $p < 0.001$ *** (3.08e–14)
Petiole width (mm)			2.18	1.91	Kruskal–Wallis: $\chi^2 = 53.826$, $p < 0.001$ *** (2.191e–13)
Leaf blade length/petiole length ratio	35.9	7.4	14.67	10.98	Kruskal–Wallis: $\chi^2 = 45.433$, $p < 0.001$ *** (1.55e–11)
Total number secondary veins			25.9	23.87	Kruskal–Wallis: $\chi^2 = 25.073$, $p < 0.001$ *** (5.521e–07)
Leaf blade width (cm)	7.7	7.5	7.03	6.32	ANOVA: $F = 28.34$, $p < 0.001$ *** (3.04e–07)
Leaf blade aspect ratio (length/width)	1.8	1.7	1.75	1.61	Kruskal–Wallis: $\chi^2 = 19.223$, $p < 0.001$ *** (1.163e–05)
Plant height (m)			5.492	4.35	Kruskal–Wallis: $\chi^2 = 17.886$, $p < 0.001$ *** (2.345e–05)
Ovary, hermaphrodite flower, length (mm)			1.8	2.05	Kruskal–Wallis: $\chi^2 = 13.509$, $p < 0.001$ *** (0.0002375)
Corolla, male flower, length (mm)			8.76	9.59	Kruskal–Wallis: $\chi^2 = 16.992$, $p < 0.001$ *** (3.754e–05)
Staminal tube, male flower, length (mm)			0.62	0.71	Kruskal–Wallis: $\chi^2 = 18.35$, $p < 0.001$ *** (1.839e–05)
Short stamen length, male flower, length (mm)			2.47	2.8	Kruskal–Wallis: $\chi^2 = 18.623$, $p < 0.001$ *** (1.593e–05)
Crown width (m)			5.561	8.0	Kruskal–Wallis: $\chi^2 = 19.482$, $p < 0.001$ *** (1.015e–05)
Corolla, male flower, diam. (mm)			5.35	6.25	Kruskal–Wallis: $\chi^2 = 21.864$, $p < 0.001$ *** (2.926e–05)
Staminal tube, hermaphrodite flower, length (mm)			0.82	0.95	Kruskal–Wallis: $\chi^2 = 22.693$, $p < 0.001$ *** (1.9e–06)
Drupe, length (cm)			1.96	2.20	Kruskal–Wallis: $\chi^2 = 30.909$, $p < 0.001$ *** (2.704e–08)
Hypocarp + drupe, length (cm)			4.64	5.44	Kruskal–Wallis: $\chi^2 = 54.179$, $p < 0.001$ *** (1.83e–13)
Stamen number, male flower			8.02	9.05	Kruskal–Wallis: $\chi^2 = 62.837$, $p < 0.001$ *** (2.245e–15)
Hypocarp, length (cm)			2.68	3.31	Kruskal–Wallis: $\chi^2 = 72.203$, $p < 0.001$ *** (< 2.2e–16)
Drupe, width (cm)			1.56	1.87	Kruskal–Wallis: $\chi^2 = 72.273$, $p < 0.001$ *** (< 2.2e–16)
Stamen number hermaphrodite flower			7.70	8.72	Kruskal–Wallis: $\chi^2 = 82.134$, $p < 0.001$ *** (< 2.2e–16)
Hypocarp width (cm)			2.35	3.13	Kruskal–Wallis: $\chi^2 = 108.77$, $p < 0.001$ *** (< 2.2e–16)

Diagnostic variables of the cerrado ecotype are shown in rows 1–8 in descending order of significance. Those of the restinga ecotype are shown in rows 9–22 in ascending order of significance. See Online Resource 1g for data on all 42 quantitative variables

*Mean values

The CART analysis (Fig. 7, Online Resource 1i) resulted in 26 misclassifications (14.4% of total) among the populations. The first node, i.e. the whole set of individuals, is divided by the colour of the internal bark, a qualitative character. The subgroups are then divided by

quantitative fruit and floral variables (hypocarp width, length of long stamen in male flower, length of staminal tube in male flower, corolla diameter of male flower). Table SM3 shows a high level of correct classification except in Campo Maior (cerrado), where only ten were

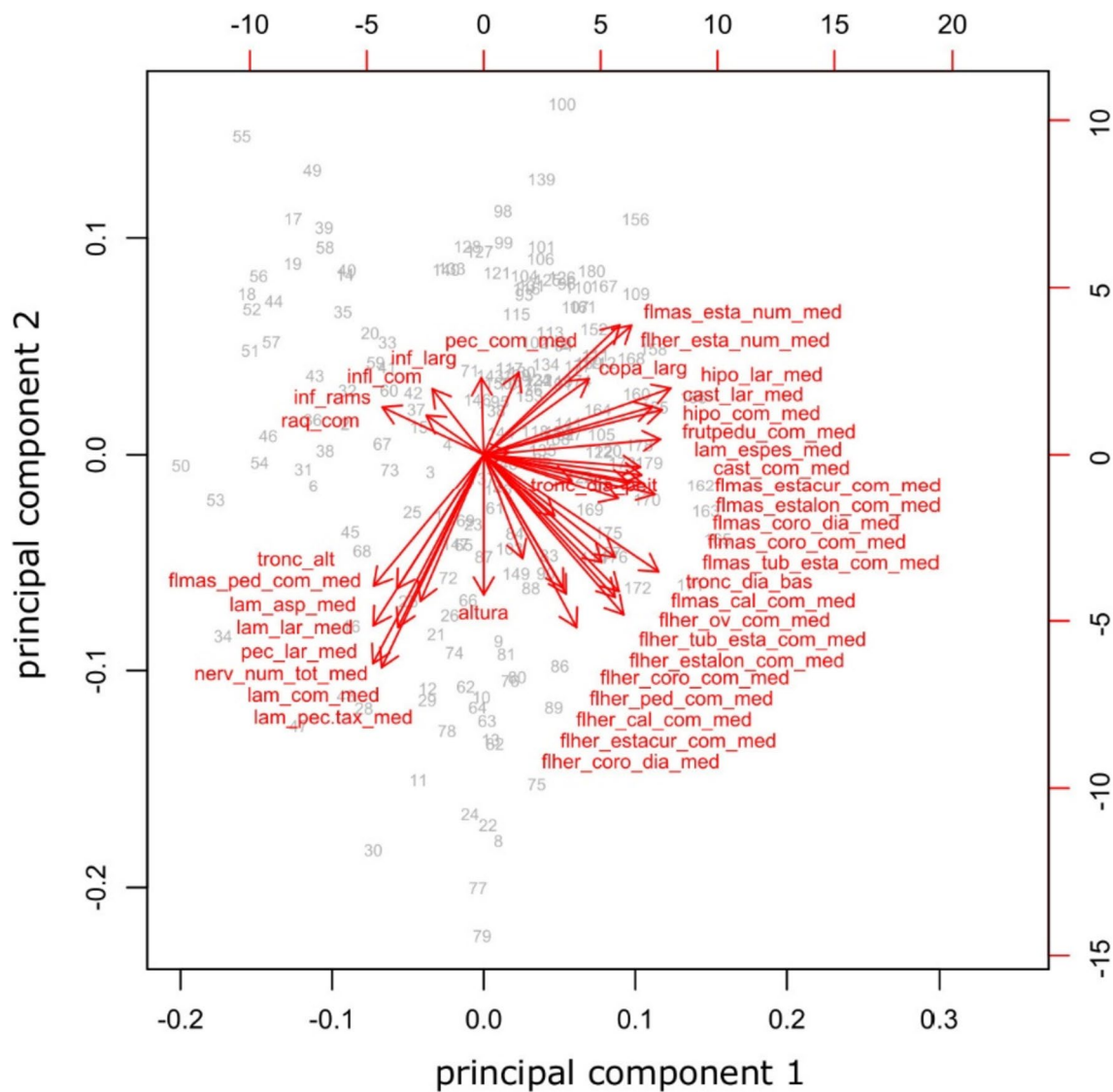


Fig. 4 Biplot superimposed on first two principal components (32.42% of total variance) of PCA using 39 quantitative vegetative and reproductive characters (red arrows) from 180 individuals (grey numbers) in six wild populations of *Anacardium occidentale*. Cer-

rado ecotype: Campo Maior, José de Freitas, Parque Nacional de Sete Cidades. Restinga ecotype: Barrinha, Ilha das Canárias, Labino. Computed with R (R Core Team 2018)

correct and 20 were assigned to other nodes, including Canárias (restinga). According to this analysis, the Campo Maior population is the least distinct in its morphology. In the CART analysis of the two ecotypes the single division was made on the quantitative variable hypocarp width (Online Resource 3).

Overall, the LDA cross-validation generated the least number of errors, followed by CART and then KNN. The CART analysis agreed with the others in showing high consistency in Sete Cidades and Barrinha, but differed from them in a much lower consistency for Campo Maior and showing reconstituted Canárias as the most heterogeneous terminal group.

The cerrado populations were individually more variable than the restinga ones (Online Resource 1j) when the variable trunk height was omitted from estimation of overall population coefficient of variation, Sete Cidades being the most variable and Barrinha the least. The multiple comparisons computed from nonparametric MANOVA (NPMANOVA, Online Resource 1k) showed that the cerrado populations were not significantly different from one another morphologically, but were in all cases significantly different from the restinga populations. On the other hand, the restinga populations differed more from one another, and only the pair Labino and Canárias was not significantly different. NPMANOVA also showed a significant

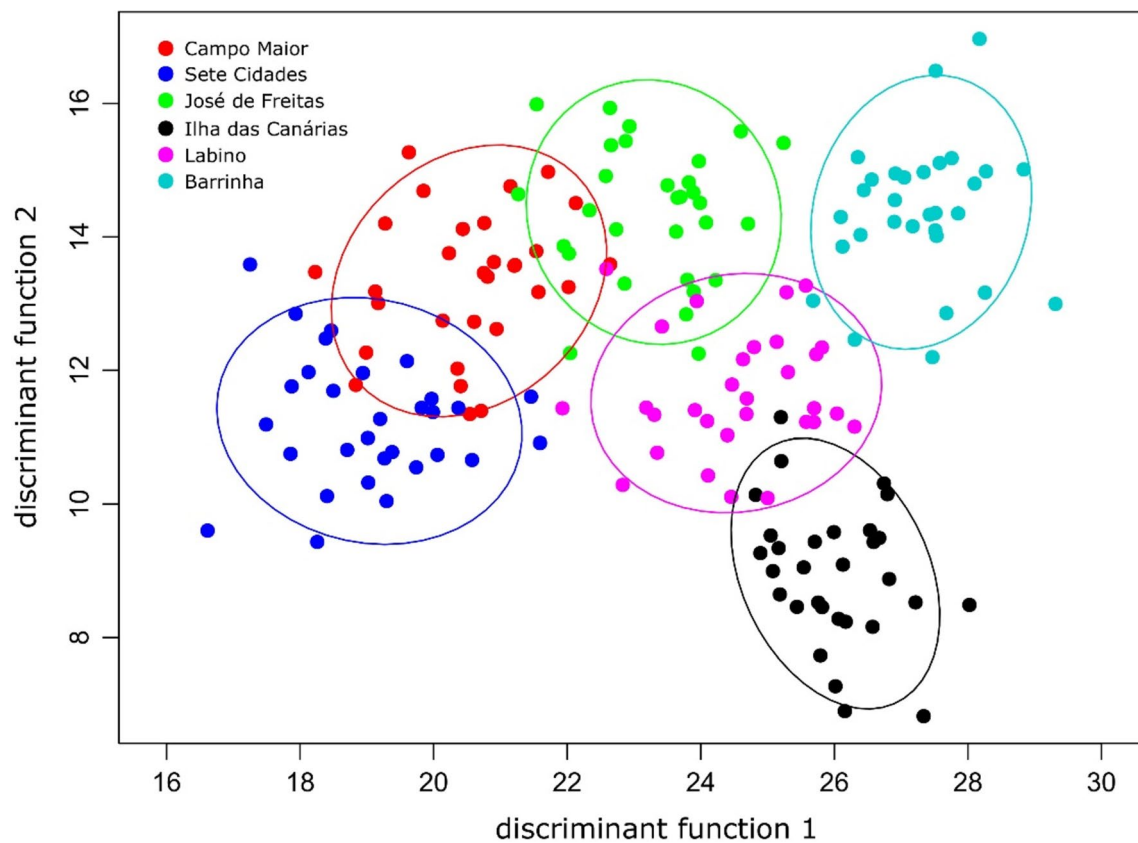


Fig. 5 Linear discriminant analysis (LDA) of 29 quantitative vegetative and reproductive characters from 180 individuals in six wild populations of *Anacardium occidentale*. Ordination on the first two discriminant function axes (80.0% of total variance). Cerrado popu-

lations: Campo Maior, José de Freitas, Parque Nacional de Sete Cidades. Restinga populations: Barrinha, Ilha das Canárias, Labino. Computed with R (R Core Team 2018)

Table 4 Cross-validation test of the LDA using populations as categorical variable and data set of 29 variables

Rows: allocation of individuals from original populations	Columns: populations reconstituted by cross-validation						Total
	Campo Maior	Sete Cidades	José de Freitas	Canárias	Labino	Barrinha	
Campo Maior	24 (80%)	4 (13%)	1 (3%)	0	1 (3%)	0	30
Sete Cidades	2 (7%)	28 (93%)	0	0	0	0	30
José de Freitas	0	0	28 (93%)	0	2 (7%)	0	30
Canárias	0	0	1 (3%)	26 (87%)	3 (10%)	0	30
Labino	2 (7%)	0	1 (3%)	3 (10%)	23 (77%)	1 (3%)	30
Barrinha	0	0	0	0	2 (7%)	28 (93%)	30
Total	28	32	31	29	31	29	180

Cerrado ecotype: Campo Maior, José de Freitas and Parque Nacional de Sete Cidades. Restinga ecotype: Barrinha, Ilha das Canárias and Labino. Computed with R (R Core Team 2018). Bold values show number (% in parentheses) of correctly assigned individuals in each population. Total misclassifications = 22 (12.2%)

morphological difference between the ecotypes (Online Resource 1m, $P < 0.001$).

The UPGMA analysis of population mean vectors showed a clear association, respectively, of the cerrado

and restinga populations into separate clusters and within them showed José de Freitas and Barrinha as distinct from the other two (Fig. 8).

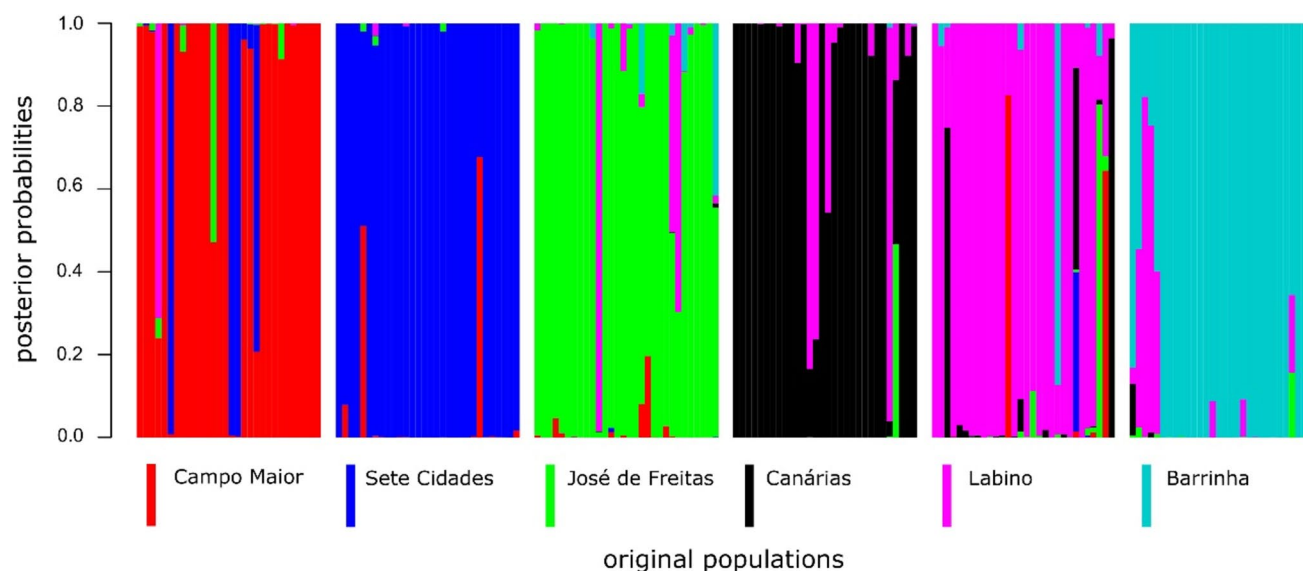


Fig. 6 Cross-validation test of linear discriminant analysis (LDA) of 29 quantitative vegetative and reproductive characters from 180 individuals in six wild populations of *Anacardium occidentale*. Each bar represents an individual; in bars with more than one coloured segment, segment length corresponds to the probability that the individ-

ual belongs to the population of that colour. Cerrado ecotype: Campo Maior, Sete Cidades and José de Freitas. Restinga populations: Barrinha, Ilha das Canárias and Labino. Computed with R (R Core Team 2018)

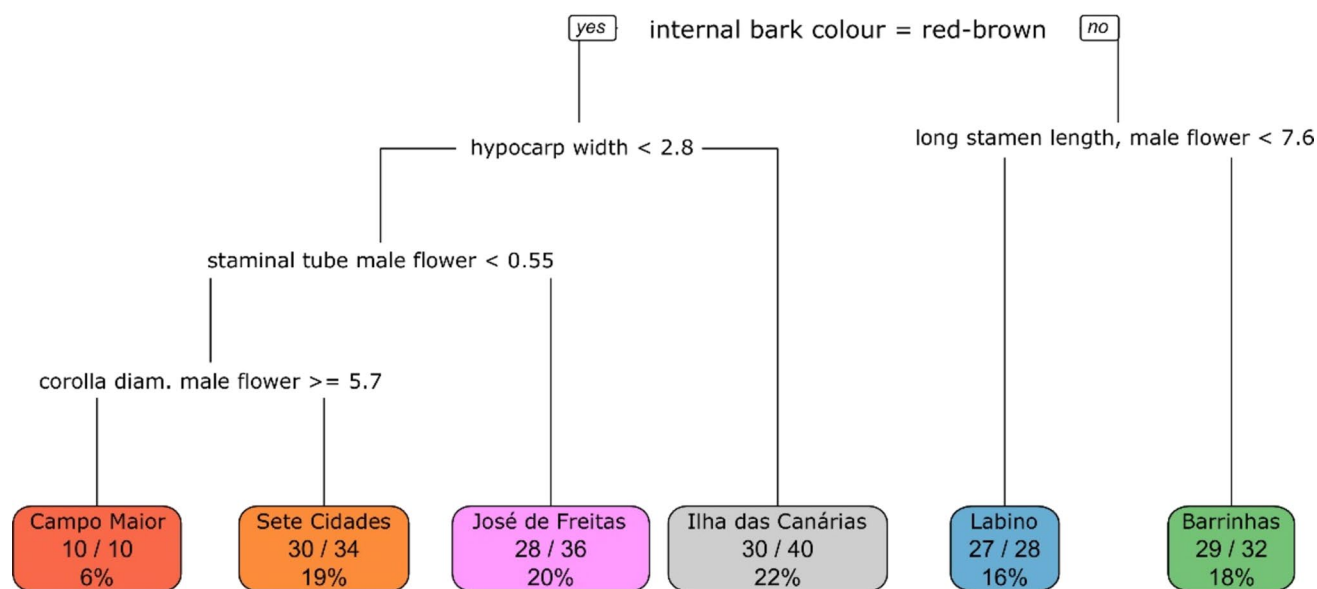


Fig. 7 Optimized classification tree (CART) based on 39 quantitative and five qualitative vegetative and reproductive characters from 180 individuals in six wild populations of *Anacardium occidentale*. Gini index used as impurity measure. Computed with R (R Core Team 2018). Numerical values at nodes are in mm; terminal groups named

for population with most included individuals; paired numbers represent number of correct classifications from named population/total individuals classified in group; percentage represents proportion of total study sample (180)

Fig. 8 Cluster analysis (UPGMA) of the six population mean vectors based on 39 quantitative vegetative and reproductive characters from 180 individuals in six wild populations of *Anacardium occidentale*. Computed with R (R Core Team 2018)

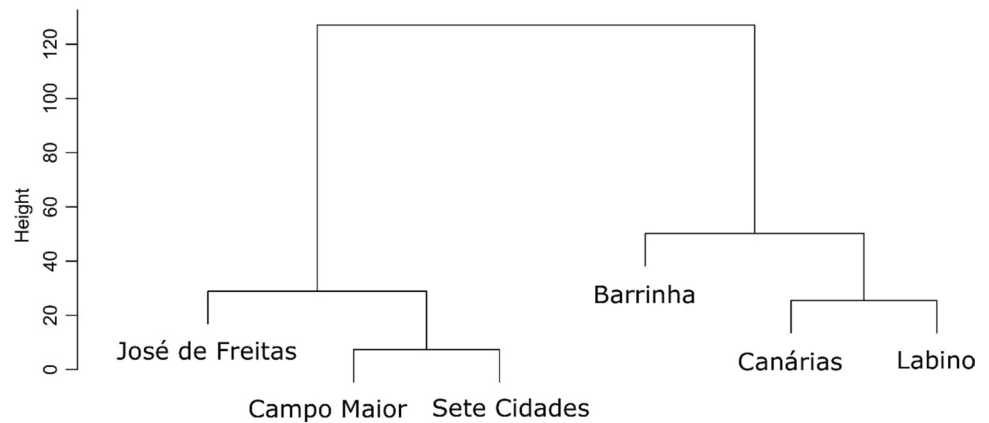
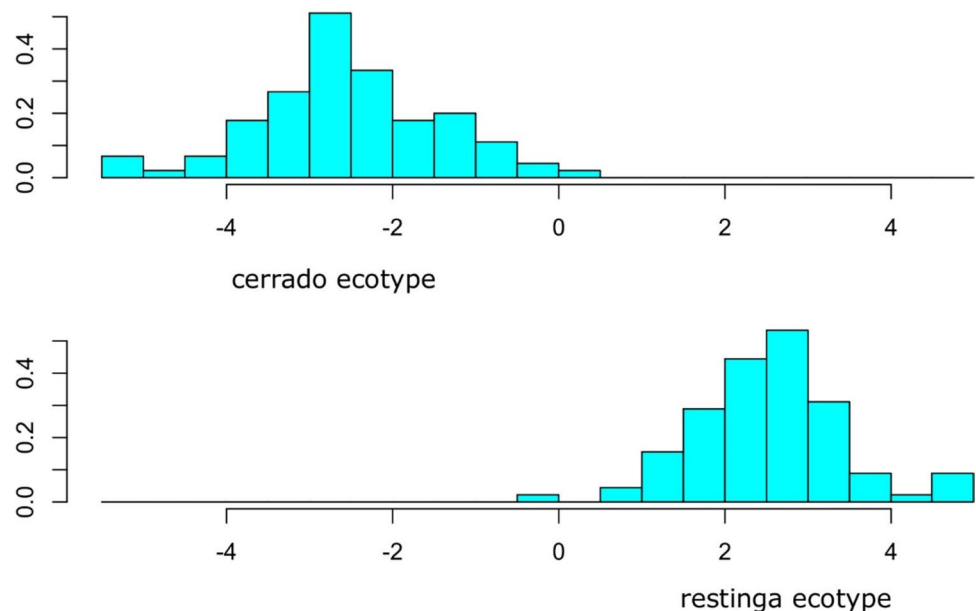


Fig. 9 Linear discriminant analysis of 39 quantitative vegetative and reproductive characters of 180 individuals from six wild populations of *Anacardium occidentale*. Frequency distribution of the scores on the single discriminant function axis using ecotype as the categorical variable. The axis in two diagrams represents the same discriminant function. Individuals from six wild populations of *A. occidentale*. Cerrado ecotype: Campo Maior, José de Freitas and Parque Nacional de Sete Cidades. Restinga ecotype: Barrinha, Ilha das Canárias and Labino. Computed with R (R Core Team 2018)



Ecotypes

The LDA differentiation of the two ecotypes is shown in Fig. 9 as a frequency distribution of the individuals along a single discriminant function axis. Almost all negative scores belong to individuals from the cerrado, and the great majority of the positive scores are from restinga plants; there is almost no overlap between the two ecotypes.

Cross-validation (Table 5) showed that 95.6% of cerrado individuals and 98.9% of restinga individuals were correctly classified, resulting in 2.8% misclassifications.

The KNN analysis correctly classified all restinga individuals, but there were more cerrado misclassifications, resulting in an overall error of 5% (Table 5). When the KNN cross-validation was carried out assigning the individuals of each population to ecotype (Online Resource 1n, f), all individuals of the restinga populations were correctly assigned, and most cerrado misclassifications to restinga

were shown to be from José de Freitas. The CART analysis showed lower differentiation with 11.7% misclassifications (Table 5); there were more misclassifications from the cerrado populations than from the restinga. Overall the three classification approaches agreed in assigning greater consistency to the restinga group. Cross-validated KNN analyses of the ten distinct partitions of the six populations into two groups of three showed that the ecotype partition had a lower percentage misclassification rate than any other such combination (Online Resource 1p). The two-group combination consisting of Barrinhas versus all the other five had an even lower error rate, which reinforces the distinctness previously mentioned for this population. The combination uniting José de Freitas with the restinga populations also had a low error rate, which reflects the LDA and CART cross-validations (Table 4, Online Resource 1i) in which the misclassifications of this population were assigned to restinga populations rather than cerrado ones.

Table 5 Results of cross-validation of LDA, KNN and CART analyses of the two ecotypes of six wild populations of *Anacardium occidentale*

		Cerrado	Restinga	Total errors	Total individuals
LDA	Cerrado (original)	86 (95.6%)	4 (4.4%)		90
	Restinga (original)	1 (1.1%)	89 (98.9%)		90
	Total	87	93	5 (2.8%)	180
KNN ($k = 9$)	Cerrado (original)	81 (90%)	9 (10%)		90
	Restinga (original)	0 (0%)	90 (100%)		90
	Total	81	99	9 (5%)	180
CART	Cerrado (original)	78 (86.7%)	12 (13.3%)		90
	Restinga (original)	9 (10%)	81 (90%)		90
	Total	87	93	21 (11.7%)	180

Cerrado ecotype: Campo Maior, José de Freitas and Parque Nacional de Sete Cidades. Restinga ecotype: Barrinha, Ilha das Canárias and Labino. Computed with R (R Core Team 2018). Bold values show number (% in parentheses) of correctly assigned individuals in each ecotype

Rows: allocation of individuals from original populations. Columns: ecotypes reconstituted by cross-validation

Discussion

The ecotypes

All the classification methods we used are based on the assumption that individuals of each population will resemble each other more than those of other populations. The cross-validation tests demonstrate that this supposition is frequently violated. The variability within each population usually overlaps with at least some of the others (Online Resource 1b). Nevertheless, as stated earlier, a key premise of this study is that the individuals of each population are genetically closer to each other than to any other population, which is to say there is an a priori reason for treating each population as a biologically meaningful group. When the populations are arranged into the sets corresponding to the ecotypes (defined on geography and ecology), they show a stronger morphological difference than in any other arrangement of two sets of three populations each (Online Resource 1n, p, f).

This study is the first to test the distinction between the two ecotypes of *A. occidentale* and differs from previous morphometric studies in using a larger range of morphological variables, sampled exclusively from wild populations. Our results generally corroborate the recognition of the cerrado and restinga ecotypes proposed by Mitchell and Mori (1987). Bearing in mind its limited geographical scope compared to the full range of the species, our study indicated that the restinga ecotype is morphologically somewhat more distinct than the cerrado ecotype and has consistently larger floral parts, drupes and hypocarps (see also CART analysis of ecotypes: Online Resource 3) and more numerous stamens. On the other hand, cerrado plants are taller and larger in most vegetative variables observed, except crown width,

trunk circumference and petiole length (Table 3, Online Resource 1g).

We confirmed that four of the five differentiating variables tabled by Mitchell and Mori (1987: 40) were among the strongest diagnostic characters of the ecotypes: cerrado populations had taller trunks, longer and wider leaf blades with higher length/width ratios and the blade length was relatively greater than the petiole length; secondary vein number was also greater and we confirmed that the petiole was thicker in the cerrado plants (Table 3). We observed that although restinga plants had significantly longer petioles, this difference was weaker than the other characters mentioned here, and there was no significant difference in leaf thickness between ecotypes (Online Resource 1g).

Variability of characters

There were substantial differences in variation measures (standard deviation and coefficient of variation) between our results (Online Resource 1q) and those of Mitchell and Mori (1987: Table IV), probably because their data were most likely obtained from measuring numerous herbarium specimens, each one probably originating from a different population, and gathered over a continent-sized area.

Trunk height varied considerably in restinga populations from zero, that is, the lowest major branches emerging at ground level, to 2 m long. Since these plants are often partly buried because of the action of dunes (Souza et al. 2021), these measurements may not reflect the true length of the trunks and so our estimates of mean population coefficient of variation are better viewed with this variable excluded (Online Resource 1j). In this case it appears that the cerrado populations are consistently more variable overall than the restinga ones. In addition to this, the cerrado populations,

despite their greater geographical spread (Fig. 1), were not significantly different from one another, whereas the geographically more concentrated restinga populations were more distinct from one another as well as less variable individually (Online Resource 1k). This result suggests the possibility that, pending the availability of molecular genetic data, phenotypic plasticity may play a greater role in the differentiation of restinga populations, perhaps linked to fine-grained habitat heterogeneity and dynamic change in dune-related habitats (Lázaro-Nagal et al. 2015; Mata et al. 2022; Schneider 2022; Souza et al. 2021). Unlike the results of Andrade et al. (2019), the Labino population did not emerge as the morphologically least variable population in our study, which contrasts with the relatively low genetic diversity for Labino plants reported by Santos et al. (2019); instead the least variable population was Barrinha (Online Resource 1j).

Certain quantitative variables of the inflorescence (length, width, bract length) and flower (style and pistil length in hermaphrodite flowers) showed no significant differences between ecotypes or populations. Variability in inflorescence length and width was quite high (coefficient of variation 28% and 35%, respectively), and since these traits are important for plant fitness (Pereira et al. 2023), selection may play a role in preventing their divergence in different populations. Qualitative characters were used only in the CART analysis; of these five characters, internal bark colour played a role dividing the data but only in the population analysis. This finding suggests that analyses with qualitative characters, or which include a greater number, may also be useful in differentiating populations.

Heterogeneity in restinga ecotype

The distinctiveness of the Barrinha population and the NPMANOVA results (Online Resource 1j, k) suggests stronger differentiation among restinga than cerrado populations. In the restinga case, this may be because of geological and related ecological factors, since Barrinha occurs over Tertiary sandstones, whereas the other two restinga populations lie on Quaternary sands. This variation within the restinga ecotype category should be followed up in future studies and may be related to adaptations to particular habitat conditions. A strong dominance of *A. occidentale* has been reported in the restinga vegetation of Pará (Silva et al. 2010), Ceará (Johnson 1973) and Piauí (Santos-Filho et al. 2010, 2016), which may be partly attributable to its ability to resist dune burial, among other factors (Souza et al. 2021). Santos-Filho (2009) also noted that in restinga, *A. occidentale* has low tolerance to competition from other species. Future research may be able to confirm that the restinga ecotype has a degree of adaptation to this marine-influenced environment which is analogous to that of cerrado populations, which

occur in a vegetation characterized by a much richer flora (Barros et al. 2003), a wide range of different soil types and are subjected to periodic natural fires, among other environmental factors (Ribeiro and Walter 2008; Mendes et al. 2012; Oliveira et al. 2012; Barros et al. 2020).

Molecular studies

No molecular genetic study has yet directly compared the two ecotypes of *A. occidentale*. The three previous published studies on wild populations were all carried out in Piauí state using ISSR markers. Borges et al. (2018) studied two populations in coastal Piauí representing the restinga ecotype (Ilha Grande, Parnaíba) and found high within-population variability and very little differentiation between them. Santos et al. (2019) compared wild and domesticated populations over a wider area of northern Piauí and reported higher within-population diversity in wild plants, significant between-population structure but only weak distinction between the wild and domesticated groups. Gomes et al. (2021) studied three wild cerrado populations in northern Piauí and found no significant genetic differentiation between them. Santos et al. (2019) concluded that morphological patterns reported by Andrade et al. (2019) for the same region were not a reliable guide to genetic diversity, implying that environmental factors may play a significant role in morphological differentiation. The results of these genetic investigations are consistent with knowledge of the reproductive biology of wild cashew, which is an outcrossing, bee-pollinated (Freitas and Paxton 1996, 1999; Freitas et al. 2014, 2002; Holanda-Neto et al. 2002) and predominantly bat-dispersed species (Mitchell and Mori 1987) in which significant inter-population gene flow is to be expected. Cota et al. (2017), in a study of the related cerrado species *A. humile*, reported significant inbreeding between naturally clumped plants within a population, which may also be factor in explaining the genetic structure of populations, especially in strongly clumped restinga populations (Andrade et al. 2019; Santos et al. 2019).

Floral biology

Floral characters can be influenced by different kinds of pollinator, which may differ between populations. Studies of the floral biology of cashew have focused more on domesticated plants (Masawe et al. 1996), but there is also work on wild populations notably by Freitas and Paxton (1998) and Freitas et al. (2014). *Anacardium occidentale* is andromonoecious with a predominance of male flowers and is pollinated mainly by ants, bees and wasps. Within the natural range of *A. occidentale* in northeastern Brazil, the common honey bee *Apis mellifera* L. is the major pollinator in commercially cultivated orchards, whereas the solitary native bee *Centris*

[*Hemisiella*] *tarsata* has been described as potentially the pollinator of wild cashew. Both wild and domesticated trees are thought to have very similar floral attributes (Freitas and Paxton 1998). Most fruits arise from cross-fertilization since there is a selective abortion mechanism that rejects self-pollinated flowers (Holanda-Neto et al. 2002). Linked to this, the proximity of forest fragments increases the number of native pollinator visits and these are essential for improving productivity (Freitas et al. 2014). Maintenance of nearby native forests is thus recommended for domesticated plantations in order to encourage cross-fertilization, and this stimulates gene flow between native and domesticated forms, producing hybrids between them. Our study suggests the possibility of a difference in the pollination ecology of cerrado plants, since we observed their floral variables had consistently and significantly lower values.

Dispersal biology

Dispersal vectors such as mammals are said to have fed on cashew hypocarps as early as the Eocene, more than 45 million years ago, and are still important biotic dispersal agents (Manchester et al. 2007). Mitchell and Mori (1987) state that frugivorous bats are dispersal agents of *A. occidentale*, carrying the drupes long distances. Humans have been an important agent of seed dispersal (Mitchell and Mori 1987), because of the economic and nutritional importance of the species, and have distributed the fruit over wide areas and long distances, bringing about new forms of gene exchange with local populations (Johnson 1973).

Domesticated and wild forms

There are many publications on the morphological variation of domesticated and wild populations of cashew, and various names have been applied to the plants studied; these have been reviewed up to 2017 by Andrade et al. (2019, including their Supplementary Table 1) from the viewpoint of systematics, and difficulties in taxon recognition and nomenclature are discussed there. Many of these studies were carried out in germplasm collections and their purpose was more to identify variation in characteristics for plant breeding than to explore the biosystematics of wild forms. Some, however, are either field-sampled from wild populations (Lucena 2006; Rufino et al. 2008; Gomes et al. 2011; Andrade et al. 2019; Carneiro et al. 2019; França et al. 2020) or based on multiple samples from wild populations cultivated in germplasm collections (Pessoni 2007; Maia et al. 2012; Gomes et al. 2013; Borges 2015; Belo et al. 2019; Matos Filho et al. 2019) and thus provide information on natural within-population variability. Most focus on drupe and hypocarp variables because of their importance for commercial products; the drupe is the main cashew export product around the

world (Aliyu and Awopetu 2011; Rossetti et al. 2019) and drupe size and weight are the most important characters in selecting plants for genetic improvement of cultivated cashews (Rossetti et al. 2019).

The research programme on wild cashew based in northern Piauí of which the present study forms a part has been previously focused mainly on differences between wild restinga populations and domesticated populations (Vieira et al. 2014; Andrade et al. 2019; Santos et al. 2019; Souza et al. 2021; Ribeiro et al. 2021). The domesticated plants sampled were very heterogeneous in status, varying from plantations of agronomic clones to traditional orchards which probably consisted of mixtures of informally planted trees and naturalized individuals resulting from local introgression with nearby wild populations. Nevertheless, these studies showed some differences between the two groups. Vieira et al. (2014), in a geometric morphometric study of leaf blade outline shape (see also Andrade et al. 2010; Nascimento et al. 2021), showed a statistically significant difference between wild and domesticated populations in coastal Piauí. Andrade et al. (2019) compared wild and domesticated populations from coastal Piauí and domesticated populations at locations further inland and found that coastal individuals had shorter, relatively broader leaves, smaller drupes and hypocarps and fewer leaf secondary veins, making them similar vegetatively to the restinga ecotype studied here. However, since they were compared to only domesticated plants from non-coastal locations, it was to be expected that the fruits of the latter would be larger. The present study compared only wild populations and showed that wild cerrado plants have even smaller fruits than those of the restinga, and that there is a much clearer distinction between the restinga and cerrado ecotypes than between the former and domesticated plants.

Ecotypes and taxonomy

The ecotype concept is not one of the formal taxonomic categories of the Nomenclatural Code for plants (Turland et al. 2018), but has been used mainly in biosystematic studies to differentiate groups of populations that are interfertile, phylogenetically close and partially differentiated ecologically, physiologically or morphologically (Stuessy 2009). The concept of cryptic species might also be applicable to the present case (Monro and Mayo 2022), once there is sufficient molecular evidence available. For the present however, we conclude that the ecotype concept continues to be useful until a more detailed picture of intra-specific diversity has been attained across its extensive geographical range. From the taxonomic standpoint, this situation continues to be less than ideal, but the lack of a more general view of the intra-specific variation in *A. occidentale* hampers the development of a formal classification of those taxa considered to belong to this species, as currently delimited in the wide sense by

taxonomists such as Mitchell and Mori (1987) and Silva-Luz et al. (2023). One example is provided by the contrasting views on recognizing the species *Anacardium othonianum*, accepted by various modern researchers (e.g. Belo et al. 2019), especially those working on the cerrado taxa of central Brazil, whereas for Mitchell and Mori (1987) this taxon represents the cerrado ecotype of *A. occidentale*. Other taxonomic problems worthy of note are the continuing confusion over the status of *A. microcarpum* Ducke (1922), an Amazonian species which has still not been sufficiently investigated in its type locality (Andrade et al. 2019; Pereira et al. 2021), and the fact that the type of *A. occidentale* itself is from Sri Lanka (Mitchell and Mori 1987). Since the type is not from a wild population, there remains the possibility of restricting this name to domesticated forms once a more comprehensive picture of the systematics of *A. occidentale* sensu lato is available, but this cannot be achieved without the perspective provided by molecular data. Morphometric techniques are a good way to classify phenotypes, and important for breeding programmes and genetic improvement, but detailed knowledge of the evolutionary lineages can only come from genetic studies, and a general understanding of the species requires both approaches. There is a clear need for more morphological, anatomical (e.g. Ramos et al. 2016), ecological and molecular phylogeographic research on wild forms of *A. occidentale*, not only in northeastern Brazil where its greatest natural diversity of forms occurs (Paiva et al. 2003), but also elsewhere throughout its range.

Conclusions

This study successfully differentiated the two ecotypes proposed by Mitchell and Mori (1987), although this result should be treated with caution since we studied only a small number of populations. The stronger distinction of the restinga populations from one another and their greater internal uniformity suggest that conservation strategies should maximize the number of protected populations in the coastal regions, where they are probably under the greatest human pressure. However, it seems likely that in both cerrado and restinga domains much greater population diversity can be expected when more wild populations have been investigated. Whether the two-ecotype hypothesis will survive more extensive study remains to be seen; patterns of diversity across the whole species range may be more mosaic-like than this hypothesis allows. Future studies should, if possible, target more remote and undisturbed populations, before they shrink or disappear or become genetically mixed with domesticated forms resulting from rapid expansion of human occupation.

The morphological divergence of the ecotypes may well reflect ecological and environmental factors as much

as genetic ones, and it is here that longer-term studies of germplasm collections are of great importance for investigating whether the differentiating phenotypic traits we have discussed have a significant genetic basis, as well as their role in generating cultivars with phenotypic traits of potential ecological, economic and agronomic significance (Pessoni 2007; Borges et al. 2018; Andrade et al. 2019; Gomes et al. 2021). Vegetative characters such as height, crown architecture and leaf attributes are likely to be influenced by nutrient levels and by wind in the restinga, where salt levels are higher and the soils more oligotrophic. Fruit variables may be influenced by gene flow from domesticated cashews, especially in more populated areas, but also by animal and human dispersal of the compound fruits (drupe and hypocarp).

In addition, a much broader program of molecular genetic studies of wild cashew populations could at least provide indexes of within- and between-population genetic diversity across the species range and thus build more rapidly a general framework for conservation strategies of this very important species. In the face of growing human pressure on natural resources, a coordinated approach is needed which combines research on morphological, ecological and molecular markers in wild populations with cultivar development in order to ensure a wider genetic basis for cashew agronomics, given the future impact of climate and habitat changes. Characterizing the morphological diversity of wild cashew populations is thus one of the fundamental elements to understanding the natural variation of the species and its adaptive capacity in the light of future climate change given the continuing human need for its sustainable use. The present study is a small but significant contribution to this end.

Information on Electronic Supplementary Material

Online Resource 1. Supplementary Figures and Tables.

Online Resource 2. Results of univariate tests and analyses.

Online Resource 3. R scripts for multivariate analyses.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00606-023-01887-7>.

Acknowledgements The authors are grateful for essential support given to this study by the Postgraduate Course in Development and Environment (PRODEMA) of the Federal University of Piauí (UFPI), Campus Petrônio Portela, Teresina, Piauí. We also thank the HDELTA herbarium and the Plant Molecular Laboratory (LAMOPE) of the Federal University of Delta do Parnaíba (UFDPar), Parnaíba, Piauí, for the use of their facilities and support for field studies. The second author thanks the Royal Botanic Gardens Kew, UK, for institutional support.

Author contributions All authors wrote the main manuscript text and reviewed the manuscript. Fieldwork, herbarium studies and laboratory

work were carried out primarily by JCF, with assistance from IRV and IMA. Data analysis was carried out by JCF and SJM.

Funding The research foundations Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES Foundation) and Fundação de Amparo à Pesquisa do Estado do Piauí (FAPEPI) provided research grants to the first author.

Declarations

Conflict of interest The authors declare there are no conflicts of interest.

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 RB, Gomes JRC (2004) Projeto cadastro de fontes de abastecimento por água subterrânea, estado do Piauí: diagnóstico do município de José de Freitas. CPRM - Serviço Geológico do Brasil, Fortaleza
- Aliyu OM, Awopetu JA (2011) Variability study on nut size and number trade-off identify a threshold level for optimum yield in cashew (*Anacardium occidentale* L.). Int J Fruit Sci 11:342–363. <https://doi.org/10.1080/15538362.2011.630297>
- Andrade IM, Mayo SJ, Kirkup DW, Berg CVD (2010) Elliptic Fourier Analysis of leaf outline shape in forest fragment populations of *Anthurium sinuatum* and *A. pentaphyllum* (Araceae) from Northeast Brazil. Kew Bull 65:3–20
- Andrade IM, Nascimento JDO, Sousa MVS, Santos JO, Mayo SJ (2019) A morphometric study of the restinga ecotype of *Anacardium occidentale* (Anacardiaceae): wild coastal cashew populations from Piauí, Northeast Brazil. Feddes Repert 2:89–116. <https://doi.org/10.1002/fedr.201800024>
- Archak S, Gaikwad AB, Swamy KRM, Karihaloo JL (2009) Genetic analysis and historical perspective of cashew (*Anacardium occidentale* L.) introduction into India. Genome 52:222–230. <https://doi.org/10.1139/G08-119>
- Barreto SSB, Ferreira RA (2011) Aspectos morfológicos de frutos, sementes, plântulas e mudas de Leguminosae Mimosoideae: *Anadenanthera colubrina* (Vell.) Brenan e *Enterolobium contortisiliquum* (Vellozo) Morong. Revista Brasil Sementes 2:223–232. <https://doi.org/10.1590/S0101-31222011000200004>
- Barros JS, Haridasan M, Farias RRS, Mendes MRA, Castro AAJF, Sousa SR, Lopes RN (2020) Influência da geologia e fertilidade do solo nas fitofisionomias e estrutura da vegetação em área de transição edafoclimática da bacia do Parnaíba-Piauí, Nordeste do Brasil. Revista Acad Ci Piauí 1:6–19. <https://doi.org/10.29327/261865.1.1-1>
- Barros LM, Crisóstomo JR (1995) Melhoramento Genético do Cajueiro. In: Araújo PP, Silva VV (eds) Cajucultura: Modernas Técnicas de Produção. EMBRAPA/CNPAT, Fortaleza
- Barros LM, Paiva JR, Cavalcanti JJV (2003) Recursos genéticos de cajueiro: situação atual e estratégias para o futuro. Recursos Genéticos e Melhoramento de Plantas para o Nordeste Brasileiro, Embrapa Agroindústria Tropical, Fortaleza
- Barros LM, Paiva JR, Cavalcanti JJV, Alves RE, Lima AC (2002) Clone de cajueiro anão BRS 189 cultivar. Crop Breed Appl Biotechnol 1:157–158
- Belo APM, Souza ERB, Camilo YMV, Naves RV, Vieira MC (2019) Fenologia, biometria e precocidade de plantas de caju arbóreo do cerrado (*Anacardium othonianum* Rizz.). Ci Florest 29:1672–1684. <https://doi.org/10.5902/1980509818841>
- Borcard D, Gillet F, Legendre P (2011) Numerical ecology with R. Springer, New York
- Borges ANC (2015) Caracterização genética em germoplasma de cajuí (*Anacardium* spp.) por meio de marcadores morfoagronômicos e moleculares ISSR. MSc Thesis, Universidade Federal do Piauí, Teresina
- Borges ANC, Lopes ACA, Britto FB, Vasconcelos LFL, Lima PSC (2018) Genetic diversity in a cajuí (*Anacardium* spp.) germplasm bank as determined by ISSR markers. Genet Molec Res 17:18212
- Brainer MSCP (2021) Cajucultura: O proveito do pedúnculo. Caderno Setorial Escritório Técnico de estudos Econômicos do Nordeste-ETENE, Fortaleza
- Brainer MSCP, Vidal MF (2018) Cajucultura nordestina em recuperação. Caderno Setorial-Escritório Técnico de estudos Econômicos do Nordeste-ETENE, Fortaleza
- Carneiro LA, Silva LS, Gomes MFC, Santos MF, Valente SES, Gomes RLF, Costa MF (2019) Morphological characterization and genetic divergence of a cashew population in Floriano, Piauí, Brazil. Genet Molec Res 3:1–8. <https://doi.org/10.4238/gmr18348>
- Carvalho RS, Pinto JFN, Reis EF, Santos SC, Dias LAS (2012) Variabilidade genética de cajuzinho-do-cerrado (*Anacardium humile* St.Hil.) por meio de marcadores RAPD. Revista Brasil Frutic 34:227–233
- Castro AAJF (2007) Unidades de planejamento: uma proposta para o Estado do Piauí com base na dimensão diversidade de ecossistemas. Publ Avulsas Conservação Ecosist, Univ Fed Piauí 18: 1–28
- Castro AAJF (2013) Cerrados do Nordeste do Brasil e ecótonos associados: Sítio 10 – Sítio Ecocem. In: Tabarelli M, Rocha O, Lacerda LD (eds) PELD-CNPq: Dez anos do Programa de Pesquisas Ecológicas de Longa Duração no Brasil: Achados, Lições e Perspectivas. Editora Universitária UFPE, Recife, pp 299–336
- Castro AAJF, Castro ASF, Farias RRS, Sousa SR, Castro NMCF, Silva CGB, Mendes MRA, Barros JS, Lopes RN (2009) Diversidade de espécies e de ecossistemas da vegetação remanescente da Serra Vermelha, área de chapada, municípios de Curimata, Redenção do Gurgéa e Morro Cabe no Tempo, sudeste do Piauí. Publ Avulsas Conservação Ecosist, Univ Fed Piauí 23:1–72
- Claude J (2008) Morphometrics with R. Springer, New York
- Cota LG, Moreira PA, Brandão MM (2017) Structure and genetic diversity of *Anacardium humile* (Anacardiaceae): a tropical shrub. Genet Molec Res 16:1–13. <https://doi.org/10.4238/gmr16039778>
- de la Mata R, Zas R, Bustingorri G, Sampedro L, Rust M, Hernandez-Serrano A, Sala A (2022) Drivers of population differentiation in phenotypic plasticity in a temperate conifer: a 27-year study. Evol Appl 15:1945–1962. <https://doi.org/10.1111/eva.13492>
- Ducke A (1922) Plantes nouvelles ou peu connues de la region Amazonienne (Ile Partie). Arch Jard Bot Rio de Janeiro 3:3–269
- Farias JC, Vieira IR, Figueirêdo LS, Mayo SJ, Andrade IM (2020) Cosmovisión en el contexto del extractivismo de cajuí (*Anacardium*

- occidentale* L.) en el Área de Protección Ambiental del Delta del Parnaíba, Piauí, Brasil. *Revista Mex Etnobiol* 3:3–19
- Farias RRS, Castro AAJF (2004) Fitossociologia de trechos da vegetação do Complexo de Campo Maior, Campo Maior, PI, Brasil. *Acta Bot Brasil* 18:949–963. <https://doi.org/10.1590/S0102-33062004000400025>
- Fernandes AG, Lopes AS, Silva EV, Conceição GM, Araújo MFV (1996) IV. Componentes Biológicos: Vegetação. In: Fundação CEPRO (eds) Macrozoneamento costeiro do Estado do Piauí. Fundação Rio Parnaíba, Teresina, pp 43–72
- Foulkes AS (2009) Applied statistical genetics with R: for population-based association studies. Springer, New York
- França KMA, Rocha LFCS, Sousa LFC, Melo RSS, Dantas ACA (2020) Caracterização morfológica de cajuí (*Anacardium* sp.) do cerrado sul maranhense. *Acta Tecnol* 14:79–91. <https://doi.org/10.35818/acta.v14i1.899>
- Freitas BM, Paxton RJ (1996) The role of wind and insects in cashew (*Anacardium occidentale*) pollination in NE Brazil. *J Agric Sci* 126:319–326
- Freitas BM, Paxton RJ (1998) A comparison of two pollinators: the introduced honey bee *Apis mellifera* and an indigenous bee *Centris tarsata* on cashew *Anacardium occidentale* in its native range of NE Brazil. *J Appl Ecol* 35:109–121. <https://doi.org/10.1046/j.1365-2664.1998.00278.x>
- Freitas BM, Pacheco Filho AJF, Andrade PB, Lemos CQ, Rocha EEM, Pereira NO, Bezerra ADM, Nogueira DS, Alencar RL, Rocha RF, Mendonça KS (2014) Forest remnants enhance wild pollinator visits to cashew flowers and mitigate pollination deficit in NE Brazil. *J Pollinat Ecol* 4:22–30. [https://doi.org/10.26786/1920-7603\(2014\)10](https://doi.org/10.26786/1920-7603(2014)10)
- Freitas BM, Paxton RJ, Holanda-Neto JP (2002) Identifying pollinators among an array of flower visitors, and the case of inadequate cashew pollination in NE Brazil. In: Kevan P, Imperatriz Fonseca VL (eds) Pollinating bees—the conservation link between agriculture and nature. Ministry of Environment, Brasília, pp 229–244
- Gomes JG, Oliveira NCC, Lopes PSN, Ballesteiro M (2011) 12630 – Componentes principais no estudo da diversidade de cajuí na sub-bacia do Rio dos Cochós. *Cad Agroecol* 6:1–5
- Gomes MFC, Borges ANC, Batista GSS, Luz GA, Oliveira MEA, Lopes ACA, Araújo ASF, Gomes RLF, Britto FB, Lima PSC, Valente SES (2021) Genetic diversity and structure in natural populations of cajuí from Brazilian cerrado. *Biosci J* 37:e37080. <https://doi.org/10.14393/BJ-v37n0a2021-53974>
- Gomes SO, Souza VAB, Costa MPSD, Silva CCP, Vale EM, Sousa M, Brito JP (2013) Avaliação da qualidade física e química de cajuí (*Anacardium* spp.) na região Meio-Norte. *Rev GEINTEC* 3:139–145. <https://doi.org/10.7198/S2237-07222013000300013>
- Hervé M (2018) RVAideMemoire: Testing and Plotting Procedures for Biostatistics. – R package, version 0.9-69
- Holanda-Neto JP, Freitas BM, Bueno DM, Araújo ZB (2002) Low seed/nut productivity in cashew (*Anacardium occidentale*): Effects of self-incompatibility and honey bee (*Apis mellifera*) foraging behaviour. *J Hortic Sci Biotechnol* 2:226–231. <https://doi.org/10.1080/14620316.2002.11511484>
- IBGE (2018) Conheça Cidades e Estados do Brasil. Available at: <https://cidades.ibge.gov.br/>. Accessed 26 Jan 2023
- ICMBIO (2018) Parque Nacional das Sete Cidades. Available at: <http://www.icmbio.gov.br/portal/visitacao/1/unidades-abertas-a-visitacao/208-parque-nacional-das-sete-cidades>. Accessed 21 May 2018
- Ivanov MMM (2020) Unidades de Conservação do estado do Piauí. EDUFPI, Teresina
- Jacomine PKT, Cavalcanti AC, Pessoa SCP, Burgos N, Melo Filho HFR, Lopes OF, Medeiros LAR (1986) Levantamento exploratório-reconhecimento de solos do Estado do Piauí. Embrapa/SNLCS/Sudene, Rio de Janeiro
- James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning with applications in R. Springer, New York
- Johnson DV (1973) The cashew of Northeast Brazil: A geographical study of a tropical tree crop. PhD thesis, University of California, Los Angeles
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of Köppen-Geiger climate classification updated. *Meteorol Z* 15:259–263
- Lázaro-Nogal A, Matesanz S, Godoy A, Pérez-Trautman F, Gianoli E, Valladares F (2015) Environmental heterogeneity leads to higher plasticity in dry-edge populations of a semi-arid Chilean shrub: insights into climate change responses. *J Ecol* 103:338–350
- Legendre P, Legendre L (2012) Numerical Ecology, 3rd English. Elsevier, Amsterdam
- Lima MG, Salviano AAC, Santana FF, Feitosa RMR (2017) Secas de 2010 a 2016 no Piauí: impactos e respostas do Estado em relação com os programas nacionais. *Parcer Estratég* 44:155–180
- Lucena VMX (2006) Diversidade genética entre genótipos de cajueiro (*Anacardium occidentale* L.) e qualidade do fruto e pseudofruto. MSc thesis, Universidade Federal de Roraima, Boa Vista
- Maia MCC, Oliveira LC, Almeida AS, Araújo LB, Vasconcelos LFL, Lacerda MN (2012) Divergência genética e correlações entre variáveis agroindustriais da população de trabalho de cajuí da Embrapa Meio-Norte. – Anais II Congresso Brasileiro de Recursos Genéticos, 24 a 28 de setembro de 2012, Belém, Pará, Trabalho 543
- Manchester SR, Wilde V, Collinson ME (2007) Fossil cashew nuts from the eocene of europe: biogeographic links between Africa and South America. *Int Pl Sci* 8:1199–1206. <https://doi.org/10.1086/520728>
- Marhold K (2011) Multivariate morphometrics and its application to monography at specific and infraspecific levels. Chapter 6 In: Stuessy, TF, Lack, HW (eds) Monographic plant systematics: fundamental assessment of plant diversity. A. R. G. Gantner Verlag, Ruggell, pp 75–101
- Masawe PAL, Cundall EP, Caligari PDS (1996) Distribution of cashew flower sex-types between clones and sides of tree canopies in Tanzania. *Ann Bot (Oxford)* 78:553–558. <https://doi.org/10.1006/anbo.1996.0160>
- Matos Filho CHA, Nunes JAR, Lopes ACA, Gomes RLF (2019) Selection of common cashew tree genotypes in commercial growing areas in municipalities of Piauí, Brazil. *Crop Breed Appl Biotechnol* 3:245–252. <https://doi.org/10.1590/1984-70332019v19n3a35>
- Mendes MRA, Munhoz CBR, Silva-Junior MC, Castro AAJF (2012) Relação entre a vegetação e as propriedades do solo em áreas de campo limpo úmido no Parque Nacional de Sete Cidades, Piauí, Brasil. *Rodriguésia* 4:971–984. <https://doi.org/10.1590/S2175-78602012000400014>
- Milborrow S (2020) Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'. R package, version 3.0.9
- MME (Ministério de Minas e Energia) (2006) Mapa Geológico do Estado do Piauí. Brasília -DF. Available at: http://rigeo.cprm.gov.br/jspui/bitstream/doc/2923/1/mapa_piaui.pdf. Accessed 16 Apr 2020
- Mitchell JD, Mori SA (1987) The cashew and its relatives (*Anacardium*: Anacardiaceae). *Mem New York Bot Gard* 42:1–76
- Monro AK, Mayo SJ (2022) Cryptic Species: morphological stasis, circumscription, and hidden diversity. The Systematics Association Special Volume 89. Cambridge University Press, Cambridge
- Mori AS, Silva LAM, Lisboa G, Coradin L (1989) Manual de manejo do herbário fanerogâmico. 2ª edição. CEPLAC, Ilhéus

- Nascimento MGP, Mayo SJ, Andrade IM (2021) Distinguishing the Brazilian mangrove species *Avicennia germinans* and *A. schaueriana* (Acanthaceae) by elliptic Fourier analysis of leaf shape. *Feddes Repert* 2:77–107. <https://doi.org/10.1002/fedr.202000025>
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2018) *Vegan: Community Ecology Package*. R package, version 2.4–6. Available at: <https://CRAN.R-project.org/package=vegan>. Accessed 22 May 2020
- Oliveira ACP, Penha AS, Sousa RF, Loiola MIB (2012) Composição florística de uma comunidade savânica no Rio Grande do Norte, Nordeste do Brasil. *Acta Bot Brasil* 3:559–569. <https://doi.org/10.1590/S0102-33062012000300006>
- Oliveira MEA, Castro AAJ, Martins FR (2010) Classificação e caracterização dos tipos vegetacionais do Parque Nacional de Sete Cidades, Piauí, Brasil. In: Castro AAJF, Castro NMCF, Arzabe C (eds) *Biodiversidade e Ecótonos da Região Setentrional do Piauí*. 5ª edição. EDUFPI, Teresina, pp 66–89
- Paiva JR, Crisostomo JR, Barros LM (2003) Recursos Genéticos do cajueiro: coleta, conservação, caracterização e utilização. EMBRAPA CNPAT, Fortaleza
- Pereira AMC, Matos Neto JD, Figueiredo RW, Carvalho JDG, Figueiredo EAT, Menezes NVS, Gaban SVF (2020) Physicochemical characterization, antioxidant activity, and sensory analysis of beers brewed with cashew peduncle (*Anacardium occidentale*) and orange peel (*Citrus sinensis*). *Food Sci Technol* 3:749–755. <https://doi.org/10.1590/fst.17319>
- Pereira ASS, Barbosa CVO, Silva EF, Guimarães JTF, Pantoja JPSF, Teixeira LA, Félix-da-Silva MM (2021) Flora of *Anacardium* (Anacardiaceae) in the state of Pará, Brazil. *Rodriguésia* 72:1–12. <https://doi.org/10.1590/2175-7860202172112>
- Pereira JS, Do Vale JC, Machado IP, Melo JWS, Vidal Neto FC, Melo DS, Barros LM (2023) Flowering time and harvest on floral morphology and production of hermaphrodite flowers in the cashew tree *Anacardium occidentale* (Anacardiaceae). *Rev Biol Trop* 71:e52092. <https://doi.org/10.15517/rev.biol.trop.v71i1.52092>
- Pessoni LA (2007) Estratégias de análise de diversidade em germoplasma de cajueiro (*Anacardium spp.*). PhD thesis, Universidade Federal de Viçosa, Minas Gerais
- Picó FX, Abdelaziz M, Castilla AR (2021) Introduction to the special issue: the ecology and genetics of population differentiation in plants. *AoB PLANTS* 13:6. <https://doi.org/10.1093/aobpla/plab0157>
- R Core Team. 2018: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available at: <https://www.R-project.org>. Accessed 21 May 2018
- Ramos GQ, Cotta EA, Fonseca Filho HD (2016) Análise morfológica das folhas de *Anacardium occidentale* L. *Biota Amazônia* 1:16–19. <https://doi.org/10.18561/2179-5746>
- Ratter JA, Bridgewater S, Ribeiro JF, Fonseca-Filho J, Rodrigues da Silva M, Milliken W, Pullan M, Pott A, Oliveira-Filho A, Durigan G, Pennington RT (2017) Analysis of the floristic composition of the Brazilian cerrado vegetation IV: Presentation of a Revised Data-Base of 367 Areas. Conservation and Management of Biodiversity of the Cerrado Biome. Available at: <https://cerrado.rbg.org.uk/download/download.php>. Accessed 7 Jun 2021
- Ribeiro JF, Walter BMT (2008) Fitofisionomias do Bioma Cerrado. In: Sano SM, Almeida SP, Ribeiro JF (eds) *Cerrado: ecologia e flora*, vol. 1. EMBRAPA, Brasília
- Ribeiro NC, Lima Neto FEM, Nobre ARA, da Silva DA, Mayo SJ, de Andrade IM (2021) Potential antioxidant and antibacterial bioactivity of leaf and stem bark extracts in wild cashew (*Anacardium occidentale* L.) populations from coastal Piauí, northeastern Brazil. *Feddes Repert* 132:141–157. <https://doi.org/10.1002/fedr.202000021>
- Ripley B, Venables W (2022) *Class: Functions for Classification*. R package, version 7.3–20. Available at: <https://CRAN.R-project.org/package=class>. Accessed 3 Apr 2022
- Ripley B, Venables W, Bates DM, Hornik K, Gebhardt A, Firth D (2022) *MASS: Support functions and datasets for Venables and Ripley's MASS*. R package, version 7.3–56. Available at: <https://CRAN.R-project.org/package=MASS>. Accessed 3 Apr 2022
- Rossetti AG, Vidal Neto FC, Barros LM (2019) Sampling of cashew nuts as an aid to research for the genetic improvement of cashew tree. *Pesq Agrop Brasil* 54:1–8. <https://doi.org/10.1590/S1678-3921.pab2019.v54.00962>
- Rufino MSM, Corrêa MPF, Alves RE, Barros LM, Leite LAS (2007) Suporte técnico para a exploração racional do cajueiro. Empresa de Pesquisa Agropecuária do Ceará, Fortaleza
- Rufino MSM, Alves RE, Aragão FAS (2008) Análise multivariada de genótipos de cajueiro em áreas nativas da região Meio-Norte do Brasil. *Proc Interamer Soc Trop Hort* 52:140–143
- Santos JO, Mayo SJ, Bittencourt CB, Andrade IM (2019) Genetic diversity in wild populations of the restinga ecotype of the cashew (*Anacardium occidentale*) in coastal Piauí, Brazil. *Pl Syst Evol* 305:913–924. <https://doi.org/10.1007/s00606-019-01611-4>
- Santos SC, Carvalho RS, Davide LMC (2020) Genetic polymorphism among natural populations of *Anacardium humile* A.St.-Hil. *Revista Brasil Frutic* 42:e-476. <https://doi.org/10.1590/0100-29452020476>
- Santos-Filho FSS (2009) Composição florística e estrutural da vegetação de restinga do Estado do Piauí. PhD Thesis, Programa de Pós-Graduação em Botânica, Universidade Federal Rural de Pernambuco, Recife
- Santos-Filho FS, Zickel CS (2013) Origem e estrutura da costa e a vegetação de restinga: o caso do litoral do Piauí. – In: Santos Filho FS, Soares AFCL, Almeida Jr EB (eds) *Biodiversidade do Piauí: pesquisas e perspectivas*, vol. 2. Editora CRV, Curitiba, pp 11–36
- Santos-Filho FS, Almeida-Junior EB, Soares CJRS, Zickel CS (2010) Fisionomias das restingas do Delta do Parnaíba, Nordeste, Brasil. *Revista Brasil Geogr Fís* 3:218–227. <https://doi.org/10.26848/rbgf.v3i3.232605>
- Santos-Filho FS, Mesquita TKS, Almeida EB Jr, Zickel CS (2016) A flora de Cajueiro da Praia: uma área de Tabuleiros do litoral do Piauí, Brasil. *Revista Equador* 5:21–35. <https://doi.org/10.26694/equador.v5i2.4460>
- Schneider HM (2022) Characterization, costs, cues and future perspectives of phenotypic plasticity. *Ann Bot (Oxford)* 130:131–148
- Serrano LAL, Pessoa PFAP (2016) Dados Sistema de Produção. Embrapa Agroindústria Tropical Sistema de Produção, 2ª edição
- Silva RM, Mehlig U, Santos JUM, Menezes MPM (2010) The coastal restinga vegetation of Pará, Brazilian Amazon: a synthesis. *Revista Brasil Bot* 4:563–573. <https://doi.org/10.1590/S0100-84042010000400005>
- Silva-Luz CL, Pirani JR, Pell SK, Mitchell JD (2023) Anacardiaceae. In: *Flora e Funga do Brasil*. Jardim Botânico do Rio de Janeiro. Available at: <https://floradobrasil.jbrj.gov.br/FB77852>. Accessed on 24 Jan 2023
- Silveira ES, Machado AA, Favero S, Roel AR, Cereda MP (2019) Líquido da casca da castanha-de-caju (LCC) como repelente do caruncho-do-bambu *Bambusa vulgaris*. *Ci Florestal* 3:1389–1397. <https://doi.org/10.5902/1980509825276>
- Sousa FEL, Ribeiro KV, Silva MP, Santos KPP (2021) Diversidade florística do entorno das piscinas naturais da serra de Campo Maior (PI), Nordeste do Brasil. *Rede Rev Eletrôn PRODEMA* 1:105–114. <https://doi.org/10.22411/rede2021.1501.09>
- Souza RTB, Mayo SJ, Andrade IM (2021) Migrating dunes and restinga vegetation in Piauí, northeastern Brazil: The dominance of wild cashew trees (*Anacardium occidentale*). *Feddes Repert* 3:1–24. <https://doi.org/10.1002/fedr.202100012>

- Stuessy TF (2009) Plant taxonomy, 2nd edn. Columbia University Press, New York
- Therneau T, Atkinson B (2019) Rpart: Recursive Partitioning and Regression Trees. R package, version 4
- Turland NJ, Wiersema JH, Barrie FR, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Kusber WH, Li DZ, Marhold K, May TW, McNeill J, Monro AM, Prado J, Price MJ, Smith GF (eds) (2018) International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Regnum Vegetabile, vol. 159. Koeltz Botanical Books, Glashütten
- Varmuza K, Filzmoser P (2009) Introduction to multivariate statistical analysis in chemometrics. CRC Press, Boca Raton
- Vieira IR, Oliveira JS, Loiola MIB (2016) Effects of harvesting on leaf production and reproductive performance of *Copernicia prunifera* (Mill.) H. E. Moore. Rev Árvore 40:117–123. <https://doi.org/10.1590/0100-67622016000100013>
- Vieira M, Mayo SJ, Andrade IM (2014) Geometric morphometrics of leaves of *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the coastal region of Piauí, Brazil. Revista Brasil Bot 3:315–327. <https://doi.org/10.1007/s40415-014-0072-3>

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