Feddes Repertorium 2019, *130,* 89–116 DOI: 10.1002/fedr.201800024

RESEARCH PAPER

A morphometric study of the restinga ecotype of *Anacardium occidentale* (Anacardiaceae): wild coastal cashew populations from Piauí, Northeast Brazil

Ivanilza Moreira de Andrade[1], *Joana D'Arc de Oliveira Nascimento*[1], *Márcia Vieira de Sousa*[1], *Juelina Oliveira dos Santos*[2], *Simon Joseph Mayo*[3]

[1] Herbário Delta da Parnaíba-HDELTA, Universidade Federal do Piauí, Campus Ministro Reis Velloso,
Av. São Sebastião, 2819, Parnaíba, Piauí, 64202-020, Brazil
[2] Programa de Pós-Graduação Biotecnologia (PPGBiotec), Universidade Federal do Piauí, Campus de Parnaíba,
Av. São Sebastião 2819, Parnaíba, Piauí, 64202-020, Brazil
[3] Identification and Naming Department, Royal Botanic Gardens Kew, Richmond, Surrey, TW9 3AE, U.K.

The infraspecific taxonomy of wild *Anacardium occidentale* is little studied. We investigated whether wild populations on coastal dunes in Piauí, Brazil differed from non-wild populations. Ten populations were sampled and twenty one morphological variables were measured. Variation within and between populations was investigated with univariate and multivariate statistics. Dune populations were mostly more similar to one another than to domesticated ones. There was significant correlation between inter-population geographical distance and morphological dissimilarity. Classification methods showed 96.4% successful assignment to the dune category and 86% to 100% to dune populations individually, but dune and non-dune populations overlap morphologically. Dune populations had shorter, broader leaves, shorter drupes and fewer secondary veins. Non-dune coastal populations showed strongest similarity to dune populations. Populations distant from the coast were most divergent. The population from the *cerrado* region was most distinct, with thicker leaf blades and narrower petioles. The dune populations are recognised as the "restinga ecotype" of *A. occidentale*. Correlation of dissimilarity and distance may result from gene flow and/or non-inherited environmental effects. Ecology and nomenclature (including the vernacular "cajuí") of the restinga ecotype are reviewed. Further comparison of restinga populations is needed along the Brazilian coastline and with natural cerrado populations.

Submitted: July 15, 2018
Revised: December 31, 2018
Accepted: January 6, 2019

Keywords:
caju, cajuí, *Anacardium microcarpum*, infraspecific taxonomy, restinga, cerrado

Introduction

Anacardium occidentale L., the cashew, is an important economic plant, known mainly for its edible nut that is marketed and enjoyed by people around the world. Although grown commercially and naturalised throughout the tropics today (Archak et al. 2009, Masawe et al. 2013, Nayak et al. 2014), this species, like its congeners, is native only to the neotropics. There, eleven species are currently recognised by taxonomists (Mitchell & Mori 1987, Mitchell 1992), eight of which are recorded for Brazil (Luz & Pirani 2017) (Table 1).

A. occidentale has been a domesticated plant for longer than the European colonization of the New World (Johnson 1973), and the associated selection and transfer of forms with desirable qualities has made it difficult to establish its natural distribution and infraspecific

---

Corresponding author: Dr Simon Joseph Mayo, Identification and Naming Department, Royal Botanic Gardens Kew, Kew Green, Richmond, Surrey, TW7 6AE, U.K.
Email: simonjosephmayo@gmail.com.

Table 1. Currently recognized species of *Anacardium* L. and their geographical distributions.

Species	Geographical distribution
A. amapaense J.D.Mitch.	Brazil, French Guiana
A. corymbosum Barb.Rodr.	Brazil
A. excelsum (Bertero & Balb. ex Kunth) Skeels	Colombia, Costa Rica, Cuba, Ecuador, Honduras, Nicaragua, Panama, Venezuela
A. fruticosum J.D.Mitch. & S.A.Mori	Guyana
A. giganteum W.Hancock ex Engl.	Brazil, Colombia, Guyana, Peru, Surinam, Venezuela
A. humile A.St-Hil.	Bolivia, Brazil, Paraguay
A. microsepalum Loes.	Brazil
A. nanum A.St.-Hil.	Brazil
A. occidentale L.	Native to Brazil. Cultivated and naturalized throughout tropics.
A. parvifolium Ducke	Brazil
A. spruceanum Benth. ex Engl.	Bolivia, Brazil, French Guiana, Surinam, Venezuela

Sources: Mitchell & Mori 1987, Mitchell 1992, Luz & Pirani 2017

variation patterns (Mitchell & Mori 1987: 39, Map 2, Paiva et al. 2003). According to Mitchell & Mori (1987) there are two wild ecotypes of *A. occidentale*, one occurring in coastal *restinga* vegetation (Araujo et al. 2018) of eastern and north-eastern Brazil, and the other in *cerrado*, a savanna-like vegetation of central and Amazonian Brazil, Colombia, Venezuela and the Guianas (Ratter et al. 2017). They argued that cultivated genotypes grown and naturalised in the palaeotropics today are probably derived from the Brazilian restinga ecotype. However, their revision did not attempt a further clarification of the infraspecific taxonomy of this species and the current picture remains unclear, despite an extensive literature that has since accrued on the morphological and genetic variation of *A. occidentale sensu lato*, based largely on the study of germplasm collections, and focussing on the genetic diversity of domesticated cashews (e.g. Barros et al. 1999, Pessoni 2007, Callado 2008, Vidal Neto et al. 2013).

The cerrado ecotype has received more attention from botanists in its natural habitat than the restinga ecotype. Taxonomists have described a number of presumed wild taxa from Brazilian cerrado vegetation types in Goiás state: *A. curatellaefolium* A.St.-Hil. (1831), *A. rondonianum* Machado (1949), *A. amilcarianum* Machado (1949), *A. kuhlmannianum* Machado (1949) and *A. othonianum* Rizzini (1969), all of which Mitchell & Mori (1987) assigned to the cerrado ecotype. *A. occidentale* var. *gardneri* Engler (1883) from Ceará and *A. microcarpum* Ducke (1922) from Pará also seem to have been based on natural populations, from caatinga vegetation and "campo coberto" respectively. All these names are now treated as synonyms of *A. occidentale* by Mitchell & Mori (1987) and Luz & Pirani (2017).

The distinctness of the cerrado ecotype has been further emphasised by the investigations of Brazilian plant scientists interested primarily in the genetic resources of the domesticated cashew (e.g. Paiva et al. 2003, Pessoni 2007) and some have clearly described *A. othonianum* as a good species (e.g. Agostinho-Costa et al. 2006). Studies targeting the diversity of natural field populations include those of Gomes et al. (2011, as *A. othonianum*) and Ferreira et al. (2015, as *A. spp*). Cerrado vegetation scientists on the other hand, e.g. Barros et al. (2016) and Ratter et al. (2017), follow the taxonomists in accepting a broad concept of *A. occidentale* (*sensu lato*).

In contrast, few studies have yet been published of wild restinga populations of *Anacardium occidentale sensu lato*. Johnson (1972, 1973, 1974) reported primarily ecological and physiognomic details of a range of coastal populations in north-eastern Brazil, particularly in Ceará state. Mitchell & Mori (1987) provided quantitative data contrasting the leaf characters of the restinga and cerrado ecotypes based on herbarium specimens. More recent detailed work from genetic diversity or taxonomic perspectives includes Pessoni (2007), whose morphometric and molecular study included a subspontaneous population from the restinga of northern Rio de Janeiro state. Leaf shape variation in the restinga ecotype in Piauí has been studied by Vieira et al. (2014), and Borges (2015) studied morphometric and ISSR molecular markers. Both these authors sampled a range of populations similar to those investigated in the present study. Further work, as yet unpublished, has been carried out on these populations and others in the coastal region of Piauí by V.M.A. dos Santos (pers.comm. 2016) on cytogenetics, N.C. Ribeiro (pers. comm. 2016) on the chemistry and biological activity of leaf and bark extracts and J.O. dos Santos (pers. comm. 2017) on genetic diversity in these populations using ISSR molecular markers.

Many studies investigating genetic diversity among genotypes of *A. occidentale* have used multivariate analysis of morphological data, often in combination with molecular data (Supplementary Table S1). Most used individuals cultivated in germplasm collections, e.g. Barros (1991), Archak et al. (2003), Samal et al. (2003), Aliyu & Awopetu (2007), Pessoni (2007), Desai (2008), Archak et al. (2009), Chipojola et al. (2009), Dasmohapatra et al. (2014), Frota Júnior et al. (2014), Sethi (2015), Jena et al. (2016), Paikra (2016), Matos Filho

(2017). A smaller number of studies used data sampled directly from natural populations in the field, e.g. *A. occidentale*: Lucena (2006), Gomes et al. (2011, as *A. othonianum*); *A. humile*: Carvalho (2011), Santos & Santos Junior (2015). Previous multivariate morphometric studies of plants from restinga ecotype populations of Piauí are: Rufino et al. (2008a, as "cajuizeiro"), Maia et al. (2012, 2016 as *A. spp.*), Vieira et al. (2014, as *A. microcarpum*) and Borges (2015, as *A. spp.*).

The present study takes a taxonomic approach to the infra-specific variation of *A. occidentale*, and uses multivariate analysis of morphometric data sampled directly from field populations. The primary aim was to investigate whether natural coastal populations of *A. occidentale* occurring on stabilised dune (restinga) habitats in Piauí state form a recognizable taxonomic group when compared to populations in other habitats which have been transformed to a greater or lesser extent by human activity. We started from the null hypothesis that the populations studied were all independent of one another and subdivided this into three study propositions: 1) The natural (old dune) populations do not form a distinct taxonomic group, i.e. there is no single variable or combination of variables which clearly diagnoses them. 2) There is no evidence to support the view that geographical proximity and morphological similarity are correlated. 3) Populations are more consistently classified by their habitat (dune or non-dune) than by their proximity to the coast (coast or non-coast).

Material and Methods

Access authorization – Access to genetic patrimony for this project has been officially registered by the Universidade Federal do Piauí on the data base of the Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado (Ministério do Meio Ambiente, Conselho de Gestão do Patrimônio Genético) under the registration code A8B44BB.

Sampling and Data Sets – Morphological data from the unpublished studies of M.V. Sousa (pers. comm. 2014) and J.D'A.O. Nascimento (pers. comm. 2016), were analysed using univariate and multivariate methods. The samples studied were made from ten populations in two separate surveys (Table 2, Fig. 1). The populations were chosen to contrast the restinga ecotype – plants of clearly natural ecological status on coastal stabilised sand dune systems – with those from a diversity of other environments, all more or less influenced by human activity. The five natural populations (Baixão, Cal, Labino, Pedra do Sal, Tatus) are located in Quaternary dune fields in the municipalities of Ilha Grande and Parnaíba, in the extreme northwest of the state (Castro 2007, Santos-Filho & Zickel 2013). These populations belong to dune vegetation communities (Fernandes et al. 1996 as "Vegetação Subperenifolia de Dunas", Santos-Filho et al. 2010 as "Fruticetos abertos em áreas inundáveis") and are subject to extractive fruit collection by local people (Rufino 2004, Rufino et al. 2007, 2008b). Three others, at Cajueiro da Praia, Floriópolis and Rosápolis, are situated in extensively urbanised areas on Tertiary Barreiras sandstone formations near coastal Quaternary sediments. Their original vegetation would probably have been some mixture of "tabuleiro" vegetation (Fernandes et al. 1996, "Vegetação Estacional do Tabuleiro) and restinga (Castro 2007) and the *Anacardium* populations today appear to be a mixture of domesticated and sub-spontaneous plants. The population at Cocal (65 km from the coast) consisted of an orchard of cultivated stocks supplied by the agronomic industry in an agricultural area originally covered by deciduous thorn forest ("caatinga"). The population at Cabeceiras (189 km from the coast) was located in an agriculturally modified landscape within the cerrado area of Piauí (Castro 2007, 2013).

Two models were used to investigate whether ecology or geography better determined the classification of the populations. The ecological model consisted of assignment either to "dune" or "nondune" categories, based on the habitat. Dune populations occurred on stabilised dunes and are considered as natural populations, while nondune populations occur in other habitats and are considered to consist, partially or completely, of domesticated forms. The geographical model assigned populations to "coast" or "noncoast" categories. Coast populations were defined as situated within 10 km from the shoreline, and noncoast as beyond. This limit was based on the approximate maximum width of the stabilised dune systems of the littoral zone of Piauí.

The two data sets differed in the the populations sampled (Table 2), the variables recorded (Table 3), and the seasons in which the fieldwork was carried out. For these reasons they were analysed separately. Field collections for data set 1 (Nascimento, pers. comm. 2016) were made between August and October in 2014 and 2015, during the flowering and fruiting period, and in data set 2 (Sousa, pers.comm. 2014) between June and December in 2012 and 2013. A total of 21 quantitative vegetative and reproductive morphological variables (Table 3) were observed. Three fertile branches were collected from each tree, stored in plastic bags and the study organs then measured in the laboratory while still fresh. Wherever possible, three separate measurements were made for each variable from each individual and their values averaged in the final matrix used for analysis.

Data Analysis – All computations were carried out using the R statistical language, version 3.5.0 (R Core Team 2018). The scripts for the analyses, including

Table 2. Populations studied in the state of Piauí, Brazil.

Population locality	Population code name	Sample (*N*) *		Municipality	Herbarium vouchers †		Longitude	Latitude	Local name	Ecology	Geography ***
		Data Set 1**	Data Set 2**		Data Set 1	Data Set 2					
Baixão	baixao	-	20 (18)	Ilha Grande	-	M.V. Sousa 114	−41.81222	−2.84547	cajuí	dune	coast
Cabeceiras	cabeceiras	12 (11)	-	Cabeceiras do Piauí	HDELTA 5050	-	−42.30727	−4.47768	caju	non-dune	noncoast
Cajueiro da Praia	cajueiro	15 (12)	15 (9)	Cajueiro da Praia	J.D.O. Nascimento 2577	M.V. Sousa 260	−41.34504	−2.93482	cajuí	non-dune	coast
Cal	cal	15 (15)	20 (19)	Ilha Grande	J.D.O. Nascimento 2580	M.V. Sousa 155	−41.81894	−2.83769	cajuí	dune	coast
Cocal	cocal	15 (15)	20 (17)	Cocal da Estação	J.D.O. Nascimento 2575	M.V. Sousa 223	−41.53000	−3.48867	caju	non-dune	noncoast
Floriópolis	floriopolis	15 (14)	19 (15)	Parnaíba	J.D.O. Nascimento 2578	M.V. Sousa 186	−41.71961	−2.89492	cajuí	non-dune	coast
Labino	labino	15 (14)	17 (17)	Parnaíba	J.D.O. Nascimento 2574	M.V. Sousa 1	−41.76303	−2.84433	cajuí	dune	coast
Pedra do Sal	pedra	15 (12)	18 (18)	Parnaíba	J.D.O. Nascimento 2582	M.V. Sousa 52	−41.72952	−2.82148	cajuí	dune	coast
Rosápolis	rosapolis	15 (13)	20 (20)	Parnaíba	J.D.O. Nascimento 2576	M.V. Sousa 344	−41.78434	−2.96317	caju	non-dune	noncoast
Tatus	tatus	15 (15)	19 (15)	Ilha Grande	J.D.O. Nascimento 2573	M.V. Sousa 31	−41.82401	−2.82976	cajuí	dune	coast

*number of individuals: three measurements were made per variable per individual and averaged.

**values are: original sample size (sample size after removing outliers).

***coast populations are defined as within 10 km of the shoreline; noncoast populations are more distant.

†specimens deposited at herbarium HDELTA, Universidade Federal do Piauí, Parnaíba.

Fig. 1. Location of study populations. A. South America; rectangle shows area in B. B. Northern region of Piauí, Maranhão and Ceará states; the Rio Parnaíba marks the boundary between Maranhão and Piauí; the Lençois Maranhenses is a large area of active dunes in Maranhão state; rectangle shows area in C. C. Area of the Rio Parnaíba Delta in coastal Maranhão and Piauí; rectangle shows area in D which includes Ilha Grande. D. Area of Ilha Grande and Parnaíba municipalities in Piauí state, showing Parnaíba city and the mouth of the Rio Parnaíba; note presence of active dunes – white crescent shapes. All base images from Google Earth (Google Earth 2018; downloaded 1 July 2018). Best viewed in electronic version.

graphics, are available from S.J. Mayo. R packages and non-base functions are listed in Table 4 with reference citations. Missing data were imputed where necessary (only in data set 2) using a random forest algorithm (Stekhoven 2013); as this procedure involves random modelling, the *set.seed* function was used to ensure repeatable results. Various exploratory data analysis operations (EDA) were carried out on the individual variables prior to multivariate analysis. Extreme values were checked by comparing with the ranges given in the species description of *A. occidentale* by Mitchell & Mori (1987) and the variable distributions were explored using index plots, exploratory data analysis graphics and boxplots for each population. This was followed by univariate tests for normality (Shapiro-Wilk) and homogeneity of variances among populations (Bartlett, Fligner-Killeen), followed by tests for differences in their means and medians (Anova, Kruskal-Wallis). Multiple comparison and significance estimates of differences between population means were carried out graphically with Tukey HSD. Hierarchical cluster analysis and dotcharts of the mean population values of each variable were used to graphically display population similarities. The complete set of values of each variable were graphed as boxplots using the ecological and geographic two-group models to visualise the separation of the groups. Multivariate outliers were detected for each population separately using procedures described by Varmuza & Filzmoser (2009, p. 437) and Filzmoser & Gschwandtner (2018). The outliers were deleted to form a reduced data set. Boxplots of the scaled (standardised, i.e. converted to z scores) version of this reduced data were computed for all variables in all populations.

The replicate data vectors for each individual plant of the unscaled reduced data set were then averaged to produce the data set used for multivariate analysis; this

Table 3. Quantitative measured variables used in Data sets 1 and 2.

Variable code Data set 1	Variable code Data set 2	Variable explained
LLAMLEN	LLAMLEN	Leaf blade length (cm)
LLAMWID	LLAMWID	Leaf blade width (cm)
LLAMTHIK		Leaf blade thickness (mm)
NLATV	NLATV	Total number of secondary veins (sum of number on each side)
PETLEN	PETLEN	Petiole length (mm)
PETWID	PETWID	Petiole width (mm)
INFLEN	INFLEN	Inflorescence length (cm)
	INFPEDLEN	Inflorescence peduncle length (cm)
	RACLEN	Rachis length (cm)
RACINTNUM		Number of internodes in inflorescence rachis
	NRAMIF	Number of branches in inflorescence
	PEDICLEN	Pedicel length (mm)
	CALXLEN	Calyx length (mm)
	COROLEN	Corolla length (mm)
	STAMNUMB	Number of stamens
	STAMSHORLEN	Shortest stamen length (mm)
	STAMLONLEN	Longest stamen length (mm)
DRUPLEN		Drupe length (cm)
DRUPDIAM		Drupe diameter, measured at centre (cm)
HYPLEN		Hypocarp length (cm)
HYPDIAM		Hypocarp width, measured at centre (cm)

data set is referred to henceforth as the individual means data set. Principal component analysis (PCA) was carried out on the individual means data set using the scaling option of the *prcomp* function and the *biplot* function, both in the *stats* package (R Core Team 2018). The number of significant principal components was computed by using the *evplot.R* function created by Borcard et al. (2011) and convex hulls (Takagi 2011) and 95% confidence ellipses (Claude 2008) were computed for each population.

Before carrying out linear discriminant analysis (LDA), the homogeneity of the population covariance matrices was tested using the function *betadisper* from package *vegan* (Oksanen et al. 2018), as described by Borcard et al. (2011, p. 208). The individual means data set was tested in four conditions, 1) untransformed, 2) transformed to natural logarithms, 3) scaled but not logged and 4) scaled and logged.

The data set with the most homogeneous population covariance matrices was then used in a linear discriminant analysis carried out with the *lda* function (Ripley 2018), with the populations as the categorical variable, and using the cross-validation option. Proportions of correct assignments to each population were computed and bar-plotted. Barplots of the contributions of each variable to the discriminant axes were made from scaled versions of the data, following Legendre & Legendre (2012: 675).

The individual means data set was also analysed by the non-parametric k-nearest-neighbours (KNN) technique using the *knn.cv* function in the *class* package (Ripley 2015), using scaled data as recommended by Varmuza & Filzmoser (2009). This technique avoids the need for consistency of covariance matrices. The algorithm classifies each individual plant into the population to which the majority of its *k* nearest neighbours belong *a priori*, their proximity being determined by a distance matrix computed from the data set of measured variables input to the function. The function carries out a cross-validation test and the proportions of correct assignments to each population were computed and bar-plotted. The optimal value for the number of nearest neighbours (*k*) was found by finding the average of 999 replications for each *k* number from 1 to 100.

Differences between populations were tested with non-parametric multivariate analysis of variance (np-manova) carried out on the individual means data set using the function *adonis* from the *vegan* package (Oksanen et al. 2018). To obtain a multiple comparison table we used the function *pairwise.perm.manova* in package *RVAideMemoire* (Hervé 2018), with 999 permutations.

A Mantel test was carried out comparing geographical and morphological distances between population centroids, using the *mantel* function of the *vegan* package (Oksanen et al. 2018). The geographical distance matrix was computed from an initial matrix of longitude and latitude of the population localities using the function *distm* from the package *geosphere* with the distance argument *fun = distVincentyEllipsoid* (Hijmans et al. 2017) which computes the great-circle distance between point pairs using the Vincenty (ellipsoid) method. The morphological distances were represented as a matrix of Mahalanobis distances between the population mean scores. Following Claude (2008, p. 115) these distances were

Table 4. R packages and non-base functions used in statistical analyses with R (R Core Team 2018).

R package	version	functions	authors	Used for
-	-	evplot.R	F. Gillet in Borcard et al. (2011)	compute number of significant principal components
-	-	Plot_ConvexHull	Takagi (2011)	plot convex hulls around points in a scatterplot
-	-	ELLI	Claude (2008)	plot confidence ellipses in a scatterplot
car	3.0-0	bcPower	Fox et al. (2018)	univariate exploratory data analysis
class	7.3-14	knn.cv	Ripley (2015)	computing k-nearest-neighbour classification with cross-validation
geosphere	1.5-7	distm	Hijmans (2017)	compute geographical distance matrix
Hmisc	4.1-1	minor.tick	Harrell (2018)	univariate exploratory data analysis
lattice	0.20-35	barchart	Sarkar (2017)	plot multiple barcharts
MASS	7.3-49	lda	Ripley (2018)	compute linear discriminant analysis
missForest	1.4	missForest	Stekhoven (2013)	compute missing data
mvoutlier	2.0.9	pcout	Filzmoser & Gschwandtner (2018)	detection of multivariate outliers
RgoogleMaps	1.4.1	GetMap	Loecher (2016)	finds required map tile from Google Maps
RgoogleMaps	1.4.1	PlotOnStaticMap	Loecher (2016)	overlays locality points onto image of map tile
rpart	4.1-13	rpart	Therneau & Atkinson 2018	compute classification tree
RVAideMemoire	0.9-69	pairwise.perm.manova	Hervé (2018)	compute multiple comparison table from non-parametric multivariate analysis of variance
StatDA	1.6.9	edaplot	Filzmoser (2018)	univariate exploratory data analysis
stats	3.6.0	prcomp, biplot	R Core Team (2018)	principal component analysis
vegan	2.4-6	betadisper	Oksanen et al. (2018)	test for multivariate homogeneity of variances
vegan	2.4-6	adonis	Oksanen et al. (2018)	compute non-parametric multivariate analysis of variance
vegan	2.4-6	mantel	Oksanen et al. (2018)	Mantel test for correlation between matrices
vegan	2.4-6	spantree	Oksanen et al. (2018)	compute and display a minimum spanning tree

computed by projecting the population mean vectors onto the eigenvectors of the previously computed linear discriminant analysis and then using the resulting scores matrix as input to the *dist* function (with default arguments) from the *stats* package (R Core Team 2018), producing a triangular matrix of Mahalanobis distances between all pairs of population mean vectors. The Mantel statistic was based on Pearson's product-moment correlation with 999 permutations.

To show the similarity relationships of the populations more clearly, an LDA of the population mean vectors was carried out and bivariate ordinations of the discriminant scores plotted. This representation was supplemented by computing and plotting a minimum spanning tree with the *spantree* function of the *vegan* package (Oksanen et al. 2018), which differs from the LDA ordination in that the dendrogram takes account of all dimensions of the discriminant space; the Mahalanobis distance matrix used was obtained using the *dist* function (with default arguments, *stats* package, R Core Team 2018) on the matrix of discriminant scores of the population mean vectors.

To compare the variability of the populations the individual means data set was used. The coefficient of variation (CV) of each variable in each population was computed and plotted in barcharts using the *lattice* package (Sarkar 2017). The CV's of the variables were then averaged for each population and for each variable across all populations and graphed in barplots. Population localities were mapped provisionally with the package *RgoogleMaps* (Loecher 2016) using the functions *GetMap* and *PlotOnStaticMap* but final maps were produced using downloads from Google Earth (Google Earth 2018) as base map images.

KNN analysis was applied to the ecological and geographical two-category models for classifying the populations. The purpose of these analyses was to see how consistently plants labelled *a priori* (based on their population membership), for example as “dune” or “nondune”, would be so classified by the KNN algorithm based on the values of their vectors of measurements. Comparison of the results of the two models would indicate whether one yielded more consistent results than the other, and hence better explained the data set. The data set of scaled individual means was used with the k parameter argument set to k = 17. The results of the cross-validation visualised the proportion of individuals in each population that were correctly classified in the ecological or geographical category assigned *a priori* to that population.

Classification and Regression Tree (CART) analysis (Varmuza & Filzmoser 2009, Crawley 2013), as implemented in the *rpart* package (Therneau & Atkinson 2018), was used to find the variables which provided the optimal partitioning of the two categories of each of the classification models, ecological and geographical. This is a data-mining method that can predict the membership of individuals (the rows of the data set) to the classes (e.g. ecological category) based on the values of the variables (columns of the data set). It works by finding the value of the variable which splits the individuals into the two classes in which there is the least misclassification of individuals according to their original categories (e.g. dune or nondune), and then the subclasses are recursively divided in the same way, using whichever variables provide the optimal splits in each case. Varmuza & Filzmoser (2009), Crawley (2013) and especially Foulkes (2009) provide detailed explanations of the method.

LDA was also applied to the ecological and geographical models using the scaled version of the individual means data set, in order to obtain estimates of the variables with most influence in the discrimination of the two categories of each model.

The distinctness of the natural populations as a group was tested under the two-category ecological model classification – the "dune" category represented these populations treated as a single group. The individual means data set was used for univariate tests (normality: Shapiro-Wilk; variance homogeneity: F test, Fligner-Killeen; group differences: Student t, Wilcoxon rank sum), and multivariate discriminant analysis (LDA, KNN) with cross-validation.

Results

The first data set had greater geographical scope (Table 2) and included leaf thickness and fruit and hypocarp characters (Table 3) highlighted for their taxonomic importance in the monograph by Mitchell & Mori (1987). The most detailed presentation of results is thus focussed on this data set. The second data set included a fifth coastal dune population at Baixão, but lacked data from the Cabeceiras population, the cerrado region sample site furthest from the coast. The variables included floral characters absent from data set 1, but lacked those of the fruit and hypocarp (Table 3). The analyses of data set 2 were used as a check on the results from data set 1, and although less well-marked the patterns corroborate those from the latter.

Data set 1 – Unless otherwise stated, the results presented and illustrated in the following sections are based on data set 1. Results and discussion based on data set 2 are given in separate paragraphs.

Overall and intra-population variation – Boxplots showed considerable variation in variables among populations (Fig. 2). Univariate tests showed that variables in the raw data sets were almost all non-normally distributed in the overall data set, and variance in almost all variables was significantly different among populations (Supplementary Table S2.A). In the individual means data set, i.e. after population multivariate outliers were removed and the replicate data vectors of each individual were averaged, the majority of variables had similar variance among populations. In this processed data set (Supplementary Table S2.B), significant differences between population pairs were detected in all variables except number of internodes in the inflorescence rachis (RACINTNUM).

The mean coefficient of variation across all variables was found to be highest in the Rosápolis and Cocal populations (Table 5A) and lowest in Labino and Cajueiro. Mean CV varied considerably among the natural (dune) populations, from 12.6 in Labino to 17.8 in Cal. Among the variables, mean CV across all populations was highest in hypocarp length and inflorescence length, and lowest in total number of secondary leaf veins and drupe length (Table 5B).

The principal component analysis (PCA) of the scaled individual means showed the main trends of variation in the data. There were five significant principal component axes (out of 12), which together accumulated 76.9% of total variance. On the first principal component axis (PC1), the ordination (Supplementary Fig. S1) showed partial separation of the three noncoast populations (Cabeceiras, Cocal, Rosápolis) from the coast ones. The coast populations (which include all the dune populations) were highly superposed, but with the two nondune coast populations (Cajueiro, Floriópolis) showing some overlap with the noncoast Cocal and Rosápolis populations. On the second axis, Cabeceiras was clearly separated from Cocal and Rosápolis, and the coast populations were all highly superposed. The biplot (Fig. 3) showed that lower values on the first axis were determined by lower values particularly of leaf length and width, petiole length, number of leaf secondary veins, drupe length and drupe diameter. High values on the second axis (Cabeceiras population) were due to thicker leaves, larger hypocarps and narrower petioles.

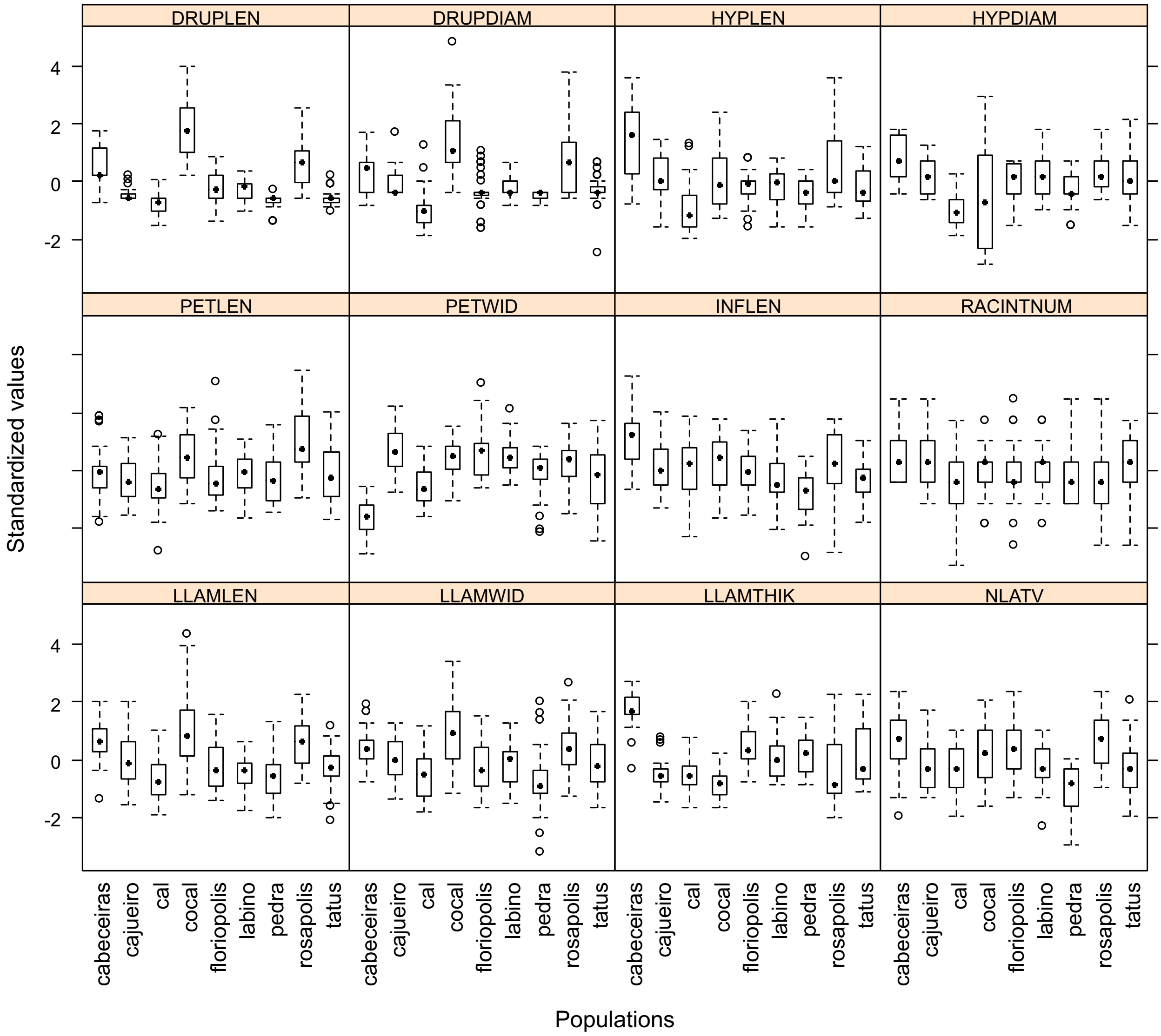

Fig. 2. Boxplots of variables within populations using the scaled version of data set 1 after removal of population outliers, but before averaging replicated measurements of each individual. Black dots indicate median values.

Similarity relations between populations – Data set 1 was found to have homogeneous population covariance matrices and so the linear discriminant analysis (LDA) could be carried out on untransformed data. Analyses of the scaled data set produced identical ordinations and cross-validation results and are figured here. The first two axes of the ordination (Fig. 4) represented 75.8% of the between-group variance. The pattern was similar to that seen in the PCA ordination. The cluster of strongly overlapping populations in the central part of the first discriminant axis consisted of all the coast populations. The three noncoast populations (Cabeceiras, Cocal, Rosápolis) were separated into two groups on this axis, the Cabeceiras population characterised mainly by its thicker leaf blades and longer hypocarps, and the Cocal and Rosápolis populations by their longer drupes and wider petioles (Supplementary Fig. S2.A). The cluster of coast populations is characterised by mostly positive values on the second axis which correspond mainly to wider petioles, shorter and broader drupes and thinner, shorter leaf blades (Supplementary Fig. S2.B). Of the noncoast populations, Rosápolis shows the most overlap with coast ones. The cross-validation results (Table 6, Supplementary Fig. S3) showed that the noncoast populations had generally higher consistencies with Cabeceiras the most (91% correct) and Rosápolis the least (54% correct), and also that consistency varied widely in the coast populations, from very low in Tatus (27% correct) to relatively high in Cal (73%).

Table 5. Coefficients of variation (CV). A. Average values across all variables in each population. B. Average values for variables across all populations. C. Average values for variables across four natural (dune) populations (Cal, Labino, Pedra, Tatus).

A.

populations	Labino	Cajueiro	Cabeceiras	Pedra	Floriópolis	Tatus	Cal	Cocal	Rosápolis
CV	12.6	13.6	14.7	15.7	16.3	16.3	17.8	18.9	19.3

B.

variables	NLATV	DRUPLEN	LLAMWID	LLAMLEN	DRUPDIAM	PETWID	RACINTNUM	LLAMTHIK	HYPDIAM	PETLEN	INFLEN	HYPLEN
CV	9.6	10.4	12.1	12.6	13.0	13.8	16.0	18.0	18.7	22.9	23.0	23.5

C.

variables	DRUPLEN	NLATV	DRUPDIAM	LLAMLEN	LLAMWID	PETWID	RACINTNUM	HYPDIAM	LLAMTHIK	INFLEN	PETLEN	HYPLEN
CV	6.9	9.1	9.5	12.5	14.2	14.8	15.4	15.6	21.2	22.1	22.4	23.2

The non-parametric k-nearest-neighbours (KNN) analysis of the scaled individual means data set was carried out using the *k* value of 17 nearest neighbours. Although the optimal value was $k = 1$, we opted for the second best result ($k = 17$) because of the risk of underfitting noted by Varmuza & Filzmoser (2009: 215). The cross-validation results (Table 6, Fig. 5) agreed with those of the LDA in the higher consistency of Cabeceiras (91% correct), Cocal (73%) and Cal (73%) and low consistency of Tatus (7% correct) and Cajueiro (17% correct).

In the LDA of population mean vectors, comparison of the minimum spanning tree (Fig. 6) with the ordination of the population centroids on the first two discriminant axes (Supplementary Fig. S4) emphasised the similarity of the coastal populations (both dune and nondune) on the one hand and the separation of Cocal and Rosápolis from Cabeceiras; the Cal population appeared more distinct in the minimum spanning tree.

Non-parametric multivariate analysis of variance (NPMANOVA) using untransformed data produced significant differences in most population pairs. The five population pairs that were not significantly different consisted of coast populations (Table 7), three involving Cajueiro (nondune, coast) and one comprising only dune populations (Tatus-Pedra).

Correlation of geographical and morphological distance – The Mantel test showed a strong correlation between geographical and morphological distance (Mantel statistic based on Pearson's product-moment correlation, $r = 0.805$, $p = 0.007$, 999 permutations).

Classification of populations by habitat or geography – In boxplots (Fig. 7), variable values in natural populations (category "dune", ecological classification model) were clearly lower than in the others ("nondune"), with shorter and narrower leaves, fewer leaf secondary veins and smaller drupes; drupe length and diameter were much less variable in dune populations. Univariate tests (Supplementary Table S2.C) showed that in all variables except leaf thickness (LLAMTHIK) and inflorescence rachis internode number (RACINTNUM) the dune category was significantly different from the nondune category. Boxplots using the geographical model (Fig. 8) showed strongest differences in leaf size, number of leaf secondary veins, petiole and inflorescence length, drupe size and hypocarp length, and lower variability in drupe length and width in the coast category.

The cross-validation results of k-nearest-neighbour (KNN) analyses of the two-group models of population classification showed greater overall consistency of assignment in the geographical model (Table 8B, coast 98.8% correct; noncoast 69.2% correct) than the ecological model (Table 8A, dune 94.6% correct; nondune 69.2% correct). This was also true when consistency of assignments to populations was considered. Under the geographical model, five out of six coastal populations

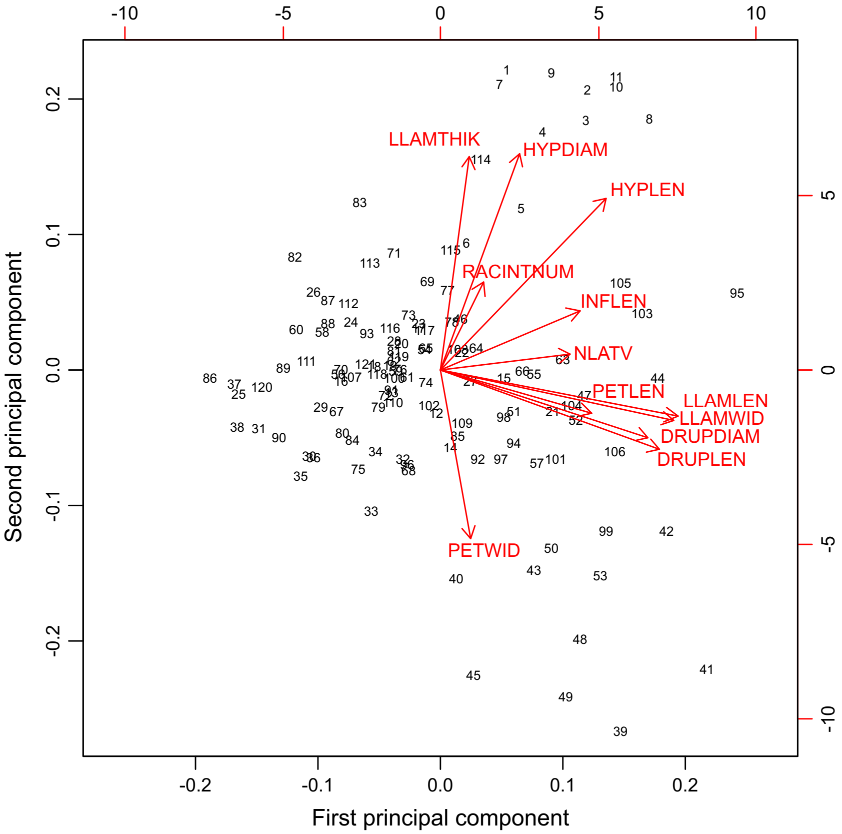

Fig. 3. Principal component analysis of individual means data set (data set 1) using scaled variables. Biplot: angle and length of arrows show relative contributions of variables to each axis. See Table 3 for explanation of variable codes. Small numbers represent the individuals in the row order in which they appear in the data matrix.

had 100% correct assignment, including all the natural (dune) populations; here the least consistent population was Rosápolis with 54% correct assignments (Table 8D). With the ecological model (Table 8C) only three populations showed 100% correct assignment – two natural populations (Pedra, Tatus) and Cocal. In this model Florіópolis was the least consistent (57% correct), and Cajueiro was 83% incorrect.

Classification and regression tree analysis (CART) prioritised leaf blade length (LLAMLEN) as the first partitioning variable and resulted in 83.8% correct classification for dune individuals and 73.8% correct for nondune. With two successive dividing variables, leaf length and drupe length (DRUPLEN) in that order, the proportions of correctly classified individuals rose to 94.6% for dune individuals but fell to 70.8% in nondune. When the geographical classification was used, however, drupe length was prioritised as first dividing variable, and proportions of correctly classified individuals were higher in both categories: 98.7% for coast individuals and 79.5% for noncoast.

The variables with most influence in a linear discrimination (LDA) of the dune and nondune categories were LLAMLEN, DRUPLEN, NLATV and LLAMWID, in that order of importance (Table 9A). The dune category (natural populations grouped together) was discriminated mainly by shorter, broader leaf blades, shorter drupes and fewer secondary leaf veins. In the geographical model the most influential variables were DRUPLEN, PETWID and LLAMLEN (Table 9B), the coast category (coastal populations grouped together) therefore having shorter drupes, broader petioles and shorter leaf blades.

The mean coefficient of variation (CV) of the categories of the two population models showed that the populations of the dunes category were less variable than those of the nondune (17.3 vs 21.6), and the coast population category less variable than the noncoast (17.3 vs. 22.0). In four natural (dune) populations (Table 5C), the mean CVs of the variables were lowest for drupe length (6.9), number of secondary veins (9.1), drupe diameter (9.5) and leaf blade length (12.5), and highest for hypocarp length (23.2), petiole length (22.4), inflorescence length (22.1) and leaf blade thickness (21.2).

Data set 2 – In this data set the most variable populations were Labino and Cal. PCA and LDA analyses

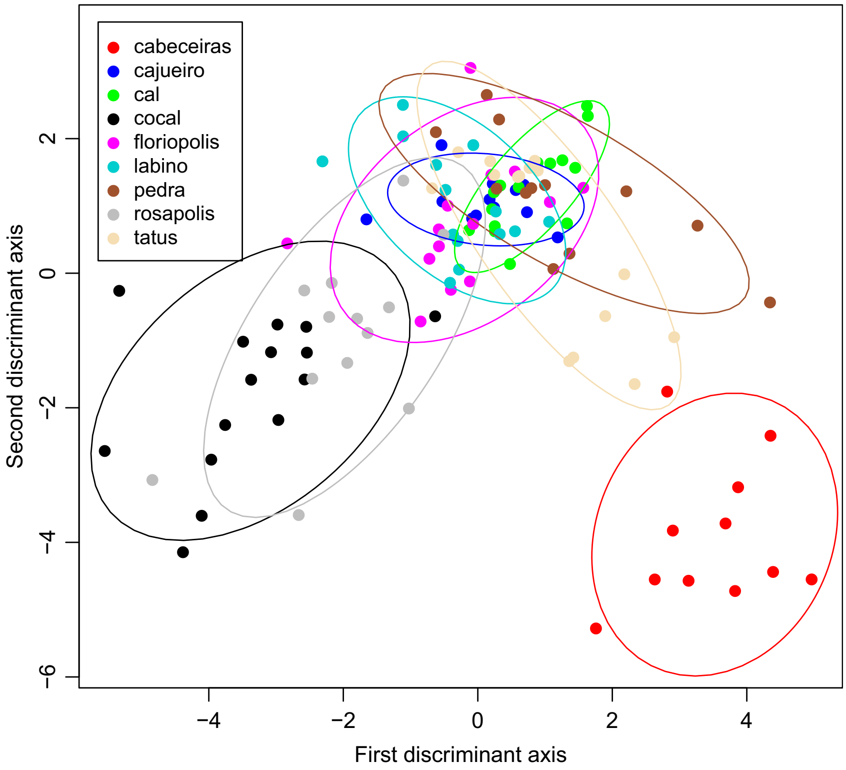

Fig. 4. Ordination of a linear discriminant analysis of the scaled individual means data set (data set 1), using populations as the categorical variable. Best viewed in electronic version.

Table 6. Cross validations of linear discriminant (LDA) and k-nearest-neighbour (KNN) analyses expressed as proportions of correct assignments. Populations in order of increasing values in LDA.

	tatus	cajueiro	pedra	floriopolis	labino	rosapolis	cal	cocal	cabeceiras
LDA	0.27	0.33	0.42	0.43	0.5	0.54	0.73	0.73	0.91
KNN	0.07	0.17	0.33	0.21	0.57	0.38	0.73	0.73	0.91

showed no clear separation of the populations; ordination of the population mean vectors of discriminant scores showed weak clustering of most of the coast populations (Baixão, Floriópolis, Labino, Pedra, Tatus). Cocal differed somewhat from the other populations in its larger leaves and floral structures and Cal emerged as the most divergent of the dune populations. In cross-validation tests of the LDA and KNN analyses, Cocal (55.6%, 44.4% correct) was the most consistent population overall. Non-parametric MANOVA showed non-significant differences between four dune population pairs and in two cases between dune populations and non-dune (Cal-Cocal, Tatus-Floriópolis). There was a highly significant correlation between geographical distance and morphological difference (Mantel $r = 0.712$, $p = 0.006$). KNN classification of individuals under the ecological model produced less clear-cut results, with only 75.6% correct assignment of dune individuals and 65.6% correct for nondune. The greatest proportion of misclassification was found in the Floriópolis population. Classification under the geographical model was somewhat better for coast populations (88.2% correct) but worse for noncoast (48.6% correct). All the dune coast populations Baixão (100% correct), Labino (94% correct), Pedra (94% correct), Cal (89% correct) and Tatus (88%) had high consistency of assignment. Overall coefficients of variation of coast (CV= 25.4) and dune (25.8) categories were slightly greater that noncoast (24.7) and nondune (23.9).

Discussion

Differentiation between natural and domesticated populations – Our working definition of the natural populations of the restinga ecotype corresponds to the

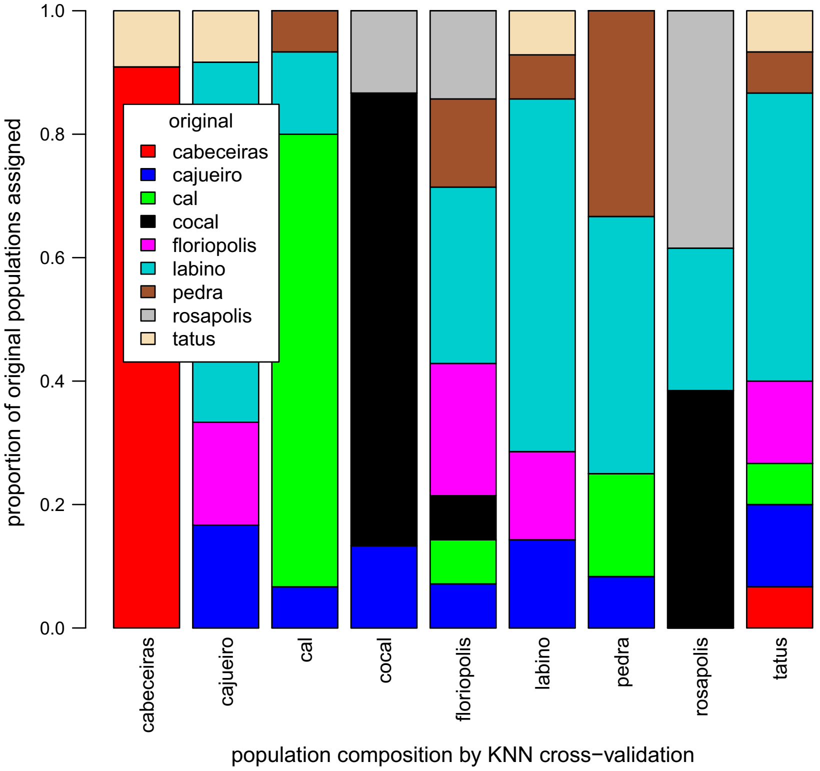

Fig. 5. Cross-validation results from the KNN analyses of the individual means data set using scaled data (Varmuza & Filzmoser 2009: 215). Each bar shows a population as constituted by the KNN cross-validation test, in which the bar name is that of the population whose individuals comprise its largest component. The more dominant one such component is, the more consistent and distinct the original population may be inferred to be as regards their morphological variation. Best viewed in electronic version.

"dune" category of the ecological model. In this case the univariate analyses showed that from a statistical standpoint the dune populations were significantly different from the others (nondune) in most variables measured (Supplementary Table S2.C). There was, however, considerable overlap in the values of all variables (Fig. 7), making it impossible to diagnose reliably the natural populations as a whole using any single character.

Cross-validation tests of the KNN analyses using the ecological model showed that most natural (dune) individuals could be correctly assigned to that category (Table 8A, 94.6% success). Under linear discriminant analysis (LDA), dune individuals were discriminated from non-dune ones most strongly by their shorter and relatively broader leaves, shorter drupes and fewer leaf secondary veins and these were also the variables with the least overall variation in the study (Table 9A). However, a high proportion of nondune individuals (Table 8A, 30.8% error) were misclassified as belonging to the dune category. When the cross-validation results were subdivided into their respective populations, it became evident that most of the misclassification between dune and nondune categories emanated from the Cajueiro, Floriópolis and Rosápolis populations (Table 8C), which are the semi-domesticated populations geographically nearest to those we have defined as natural. Cross-validation under the ecological model effectively classified (83% of individuals) the Cajueiro population, the semi-domesticated population nearest the coast, as a dune population, indicating a strong similarity to that category (Table 8C).

The geographical model yielded better discrimination (Table 8B) and this appears to be a consequence of the fact that under this model, the Cajueiro and Floriópolis populations were grouped *a priori* with the natural populations into the same category (Table 8D). These results therefore point to the conclusion that distance from the coast is a more effective explicator of inter-population similarity than the ecological distinction between natural (dune) and domesticated (nondune) status. Support for this idea is also provided by the significant correlation found between inter-population geographical and morphological distance, resulting from the Mantel test.

The CART analyses showed that the group of natural populations (dune category) was distinguished most effectively from domesticated ones (nondune) by shorter

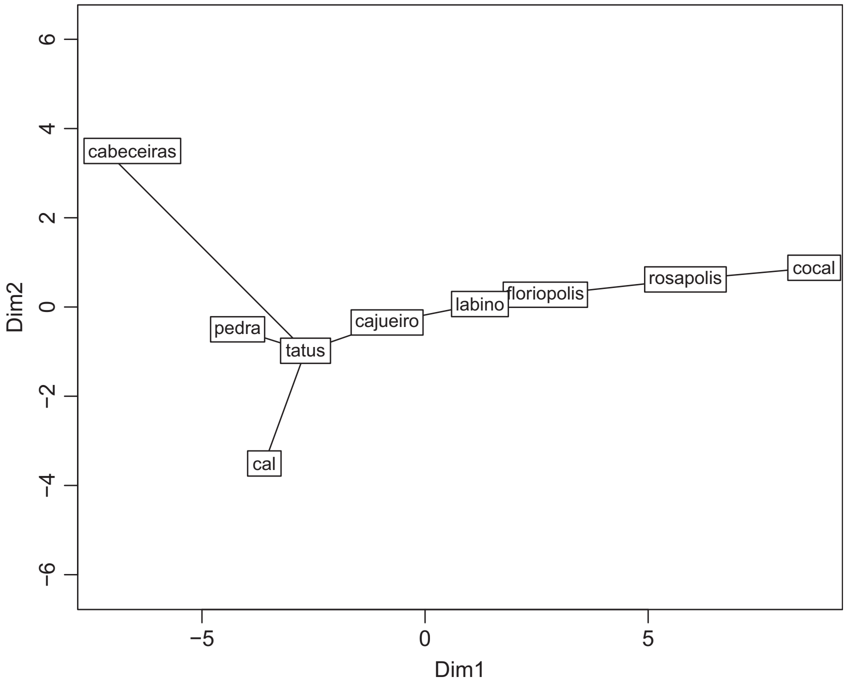

Fig. 6. Linear discriminant analysis of mean vectors of populations. Minimum spanning tree (MST) computed from a matrix of Mahalanobis distances between population means.

Table 7. Multiple comparison of population pairs using permutation multivariate analysis of variance (NP-MANOVA), showing the *p*-values of the Pillai test results. Computed in R using the *RVAideMemoire* package and the function *pairwise.perm.manova* with untransformed data.

	cabeceiras	cajueiro	cal	cocal	floriopolis	labino	pedra	rosapolis
cajueiro	0.002	-	-	-	-	-	-	-
cal	0.002	0.002	-	-	-	-	-	-
cocal	0.002	0.002	0.002	-	-	-	-	-
floriopolis	0.002	0.022	0.002	0.002	-	-	-	-
labino	0.002	0.138 n.s.	0.002	0.002	0.165 n.s.	-	-	-
pedra	0.002	0.070 n.s.	0.007	0.002	0.016	0.003	-	-
rosapolis	0.002	0.003	0.002	0.028	0.028	0.003	0.002	-
tatus	0.002	0.323 n.s.	0.026	0.002	0.016	0.006	0.323 n.s.	0.002

n.s. : not significant

leaf blade length, whereas populations within 10km of the coastline (coast category) were discriminated most from the others (noncoast) by shorter drupes. Furthermore, the discrimination provided by drupe length in the geographical model was greater, in terms of the proportion of correct assignments, than that achieved by leaf length under the ecological model. In linear discriminant analysis, these two variables were the most influential discriminators in the ecological model, but in the geographical model petiole width was the second-most influential variable (Table 9).

The restinga ecotype, at least in Piauí and Ceará, is also characterised by its habit and architecture, although we did not include these features in our study. Johnson (1972) described and illustrated the low stature and spreading, partially prostrate growth, which we also observed (Fig. 9B, D-E); major branches are often partially buried in the sand and can maintain the growth of

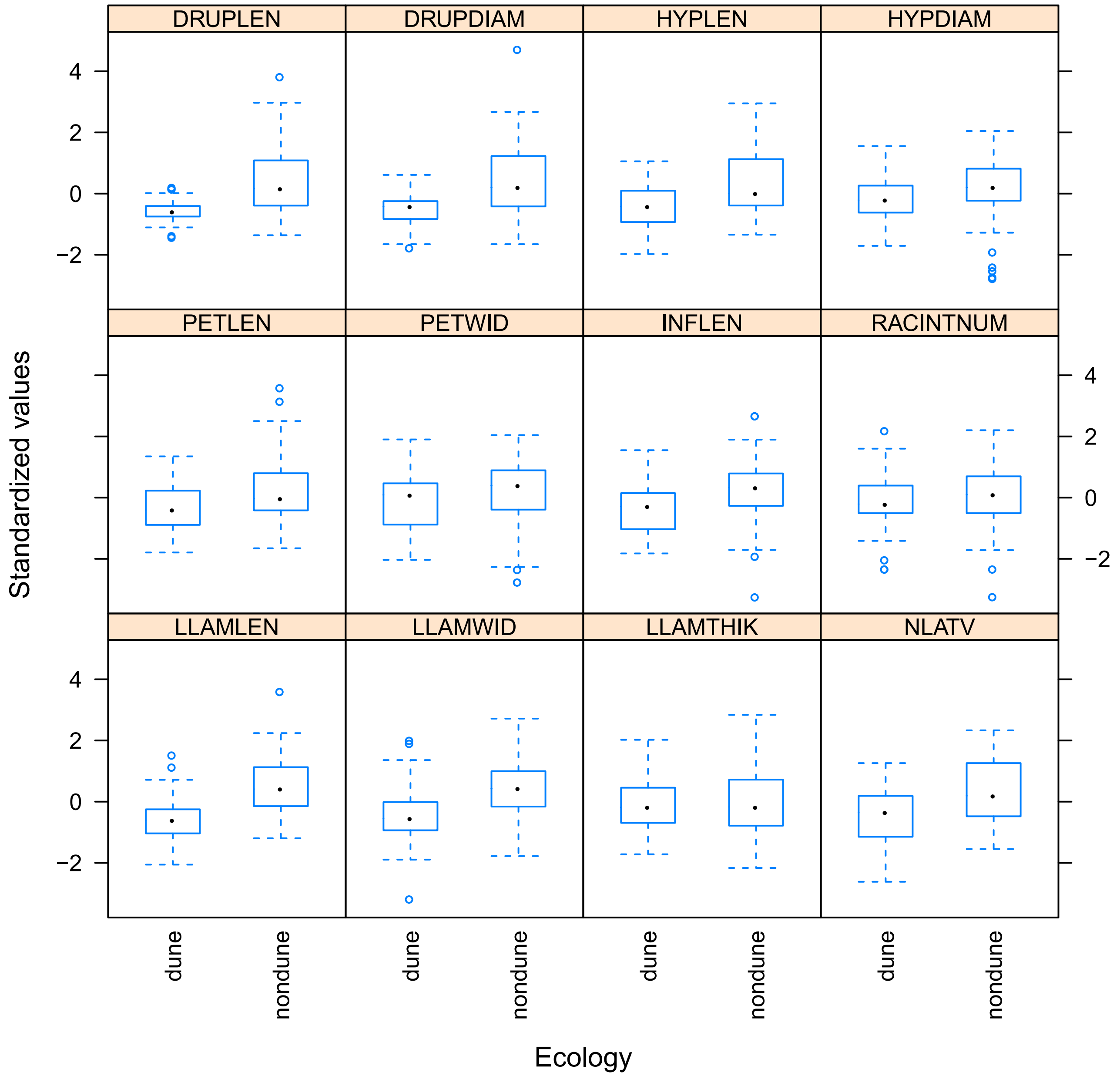

Fig. 7. Boxplots of the scaled individual means data set comparing variables in the ecological model; "dune" represents the category of natural populations.

terminal shoots above the surface of an advancing dune (Fig. 9C). Johnson (1972) also stated that when growing further inland on more fertile soils the cashew tree is more upright, but the major branches still emerge low down on the trunk and spread widely. Callado (2008) refers to the extensive lateral root system (Fig. 9B). All these features suggest the possibility of adaptation to the stabilised dune habitat, but no studies are known to us which provide evidence for their genetic determination.

Various causal factors for these results can be adduced, including: i) the influence of windborne salt, which could be expected to increase towards the shoreline and reduce growth rates; ii) growth reduction produced by increasingly nutrient-poor soils when passing from the Tertiary continental geology to the Quaternary sands of the coast; iii) gene flow between natural dune populations and neighbouring domesticated or semi-domesticated ones. Experimental design could also have contributed. Thus, the nondune and noncoast

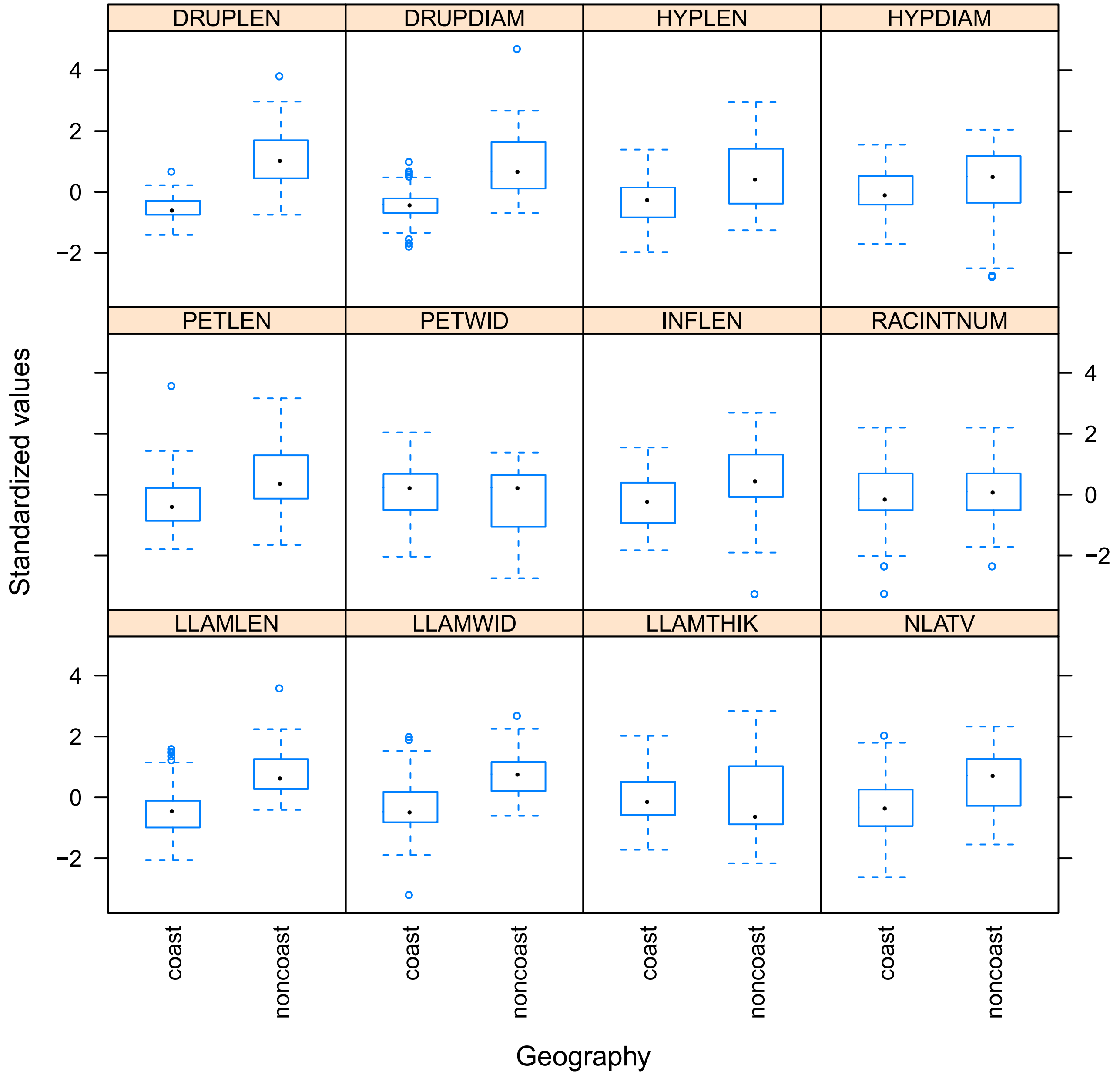

Fig. 8. Boxplots of the scaled individual means data set comparing variables in the geographical model; "coast" represents the category of populations situated within 10 km from the shoreline.

groups were much more heterogeneous, in terms of their much wider geographical dispersion and their differences in genetic status, ranging from semi-naturalised stands, as in Rosápolis, Floriópolis, Cajueiro and Cabeceiras, to cultivated orchards raised from agronomically developed stocks in the case of Cocal. These differences are reflected in the greater morphological variability of the nondune and noncoast categories, as shown by their mean coefficients of variation, and in the results of multivariate analyses using the populations as grouping categories. Borges (2015), in a study using morphometric and molecular markers which included restinga ecotype plants identified as "cajuí" (*Anacardium spp.*), found no significant difference between samples from the municipalities of Ilha Grande and Parnaíba. She suggested that fruit bat dispersal (Mitchell & Mori 1987), bee pollination (Freitas & Paxton 1998) and high germination rates (Corrêa et al. 2002) would promote considerable inter-population gene flow, high within-population genetic diversity, and hybridization with domesticated forms of *A. occidentale.*

Table 8. Cross-validation results of k-nearest-neighbour analyses (KNN) of two models of population classification expressed as percentages. A. Ecological categories as grouping factors. B. Geographical categories as grouping factors. C. Cross-validation assignments of ecological categories plotted by population. D. Cross-validation assignments of geographical categories plotted by population.

A.	assigned to dune	assigned to nondune
dune	94.6	5.4
nondune	30.8	69.2

B.	assigned to coast	assigned to noncoast
coast	98.8	1.2
noncoast	30.8	69.2

C.	Population *a priori* category	cabeceiras nondune	cajueiro nondune	cal dune	cocal nondune	floriopolis nondune	labino dune	pedra dune	rosapolis nondune	tatus dune
assigned to:	dune	9	83	93	0	43	86	100	23	100
assigned to:	nondune	91	17	7	100	57	14	0	77	0

D.	Population *a priori* category	cabeceiras noncoast	cajueiro coast	cal coast	cocal noncoast	floriopolis coast	labino coast	pedra coast	rosapolis noncoast	tatus coast
assigned to:	coast	27	100	100	20	93	100	100	46	100
assigned to:	noncoast	73	0	0	80	7	0	0	54	0

Inter-population similarity relationships – When the focus shifts to the interrelations between the individual populations, principal component analysis and linear discriminant analysis showed similar patterns. The coast populations form a single cluster and the noncoast populations form two further groups. One of the latter groups is linked by the overlapping Rosápolis and Floriópolis populations to the coast cluster, and the other, consisting of Cabeceiras (geographically the most distant population) is quite distinct from all the rest (Fig. 4). These relationships are shown more clearly by the minimum spanning tree and ordination of the LDA of population mean vectors (Fig. 6, Supplementary Fig. S4). In Table 7, all non-significant pairwise differences were between coast populations, which is in concordance with the overall significant positive correlation between inter-population geographical distance and morphological dissimilarity. Together, the results agree with the analysis of the coast-noncoast model of population classification in pointing to geography as the main explanatory factor for the patterns observed.

The natural (dune) populations, despite their similarity, were also heterogeneous in other respects. They differed from one another in their consistency almost as much as the domesticated and semi-domesticated ones (nondune). The KNN cross-validation results (Table 6, Fig. 5) showed a range of consistency within four natural populations, from 7% correct assignment in Tatus to 73% in Cal, while in the nondune populations the range was 17% in Cajueiro to 91% in Cabeceiras. A similar wide range was shown in mean population coefficients of variation (CV), i.e. overall within-population morphological variability (Table 5A). CV varied in the natural populations between 12.6 (Labino) and 17.8 (Cal), and in domesticated ones from 13.6 (Cajueiro) to 19.3 (Rosápolis). The relatively high variability of the natural populations is consistent with that previously reported by authors from natural and spontaneous populations in coastal areas of north-eastern Brazil (e.g. Paiva et al. 2003, Pessoni 2007, Borges 2015). Studies of wild cerrado populations have also reported high within- and between-population variability, e.g. Gomes et al. (2011) for *A. othonianum*, Carvalho (2011), Carvalho et al. (2012) and Santos & Santos Junior (2015) for *A. humile*, and Ferreira et al. (2015) for *A. spp*.

The second data set showed less structure overall both in inter-population similarity and between the categories of the two classification models. This is likely due to a more unequal balance in populations representing the categories (more dune and coast populations and less nondune and noncoast), the absence of the very divergent and distant Cabeceiras population, and the lack of leaf thickness, drupe and hypocarp variables, which were important discriminators in data set 1. However, there are also similarities in the two results. Cocal, the most agronomic population from the interior, is the

Table 9. Loadings of the variables on the discriminant axis from linear discriminant analyses (LDA) using the two classification models of the populations as categorical variables. A. Ecological classification: scores were predominantly negative for dune individuals and positive for nondune. B. Geographical classification: scores were predominantly negative for coast individuals and positive for noncoast.

A. Ecological classification (dune/nondune) variables	LLAMWID	RACINTNUM	PETLEN	LLAMTHIK	DRUPDIAM	HYPDIAM	INFLEN	HYPLEN	PETWID	NLATV	DRUPLEN	LLAMLEN
loadings	−0.40	−0.05	0.05	0.02	0.12	0.15	0.22	0.25	0.30	0.44	0.58	0.58
B. Geographical classification (coast/noncoast) variables	**PETWID**	**LLAMTHIK**	**LLAMWID**	**HYPLEN**	**RACINTNUM**	**DRUPDIAM**	**INFLEN**	**HYPDIAM**	**PETLEN**	**NLATV**	**LLAMLEN**	**DRUPLEN**
loadings	−0.60	−0.19	−0.17	−0.02	0.05	0.06	0.12	0.16	0.25	0.28	0.47	1.28

most divergent population overall, Cal the most divergent of the dune populations, Florióplis the most intermediate population between dune and nondune categories, the geographical (coast-noncoast) classification has the highest consistency, and there is a significant correlation between inter-population geographical and morphological distances. This supports the conclusion that in addition to fruit size, leaf and floral variables also play a part in differentiating coastal and natural populations from those of the interior and of domesticated origin.

Critique of the ecological status of restinga ecotype populations – Various authors have concluded that natural populations of *A. occidentale* from coastal regions in north-eastern Brazil are unlikely to exist today other than in restinga/stabilised dune habitats where they form part of distinct vegetational associations (Fig. 9A). Johnson (1972) was of the opinion that the only natural populations of *Anacardium occidentale* in the coastal region of north-eastern Brazil were those from the restinga. He commented (Johnson 1972: 105): "The previous chapter dealt with the natural growth of cashew in the littoral. In the remainder of the coastal region, contemporary natural occurrence of the plant is questionable, and it is the writer's opinion that none exists. Cashews may have been present in the past, but natural vegetation has been modified to such an extent that it is difficult if not impossible to find an undisturbed site today to resolve the question with certainty."

This interpretation is at least partly supported by scientists working on breeding and genetic improvement of cashew in north-eastern Brazil. Barros et al. (1999), for example, stated that [translated from Portuguese]: "It is important to emphasise that although the cashew is found most abundantly in the coastal region of north-eastern Brazil, where it grows apparently in a spontaneous state, above all in the vegetation of the beaches and dunes, the cashew cannot tolerate competition with other species, which is different from what is observed in other species of the genus, such as those of the cerrado and the Amazon forest, which grow normally as components of the local floras." Paiva et al. (2003) stated that the area of natural distribution of *A. occidentale* is completely confused by the dispersal of the cultivated form and that its greatest diversity is concentrated in the coastal zones of north-eastern Brazil where great variability in the characters of most agronomic interest, such as productivity, drupe and hypocarp weight, is observed in genotypes from the restinga and dune vegetation.

This supports our observations that the "non-dune" category of *A. occidentale* is a heterogeneous mixture of planted and subspontaneous orchards derived from domesticated *A. occidentale* genotypes, some of which were acquired as commercially marketed improved

Fig. 9. Ecology of wild populations of *Anacardium occidentale* restinga ecotype. A. Restinga thicket with *A. occidentale* on the right (arrowhead) and individuals of *Cereus jamacaru* and Myrtaceae. B. *A. occidentale* on a stabilised dune which is undergoing erosion, showing superficial roots (arrowheads). C. Active dune engulfing *A. occidentale* with branch tips maintaining growth above dune surface (arrowhead). D. *A. occidentale* on stabilised dune in background (arrowhead) with dune slack in foreground on right, during the dry season. E. *A. occidentale* showing semi-prostrate growth on low sand hillocks. A, B, D, E, Labino population; C, Tatus population. Best viewed in electronic version.

varieties (e.g. Cocal, according to the owner). The dune populations may with more reason be viewed as a natural group, given their specialised ecology and status as components of characteristic vegetational associations (discussed below, Fig. 9A, Fig. 10C).

If a stronger morphological characterization of the restinga ecotype existed in the past, it is likely to have been blurred by intercrossing with nearby domesticated populations and their spontaneous derivatives in the coastal region. The restinga ecotype is known to have out-crossing floral biology (e.g. Freitas & Paxton 1998, Freitas et al. 2002, Ribeiro et al. 2008, Freitas et al. 2014) and the practice of planting new individuals in orchards from locally produced seed (Paiva et al. 2003, Matos Filho 2017) may result in trees which are genetic mixtures of natural and domesticated genotypes, especially in localities closest to the dune habitat. There is also the possibility of gene flow from domesticated forms into natural dune populations.

The characteristic ecology of the dune populations (Figs. 9, 10C) is an important part of the argument for recognizing them as a natural entity, e.g. as an ecotype. The populations growing on stabilised dunes in the region of the Rio Parnaíba Delta form part of natural vegetation assemblages that have been studied by Fernandes et al. (1996) and more recently and in greater detail by Santos-Filho (2009) and Santos-Filho et al. (2010). The latter authors described the structure and floristics of the plant communities of which *A. occidentale* is a major component. They described the restinga vegetation as occurring on sandy substrates over Quaternary geological surfaces and highlighted the importance of *A. occidentale* as a core species of a vegetation component termed “Fruticetos abertos inundáveis” – reticulated or isolated thickets of woody vegetation occurring on raised ground, between which lie depressions subject to flooding in the rainy season (Figs. 9A, B, D, 10C). They noted (translation from

Fig. 10. Wild populations of *Anacardium occidentale.* A. Developing fruits at Labino. B. Mature fruits and flowers at Pedra do Sal. C. Image showing stabilised dune fields at Ilha Grande, Piauí, during the rainy season, showing their regular orientation in the direction of the prevailing trade winds, the flooded dune slacks, and on the higher relief the mosaic of restinga thickets in which wild *A. occidentale* is a prominent component. D. Fruiting individual at Pedra do Sal. C, image from Google Earth 2018; downloaded 1 July 2018. Best viewed in electronic version.

Portuguese): "Based on our field observations, the most frequently observed facilitating species in this facies is the cajueiro (*Anacardium occidentale* L.). In the restinga of Ilha Grande these vegetation islands almost invariably have a core association of a cajueiro and some associated species such as mandacaru (*Cereus jamacaru* DC.), puçá (*Mouriri pusa* Gardner), carnaúba (*Copernicia prunifera*), tucum (*Astrocaryum vulgare* Mart.), muricis (*Byrsonima* spp.) and *Ouratea fieldingiana* (Gardner) Engl." Santos-Filho (2009: 43) reported that the woody plants of these thickets do not exceed 3m in height. Fernandes et al. (1996) described this vegetation type as semi-deciduous dune vegetation ("Vegetação Subperénifolia de Dunas") on stabilised dunes ("dunas fixas"), noting that its greatest extent was in Ilha Grande, where these dunes achieve their greatest extent in Piauí. These authors recorded the following as the most common species: the windward side is occupied mainly by shrubs up to 4 m tall: murici (*Byrsonima verbascifolia*), cajuí (*Anacardium microcarpum*), guajirú (*Chrysobalanus icaco*), puçá (*Mouriri guianensis*), ubaia (*Eugenia sp.*), batiputá (*Ouratea fieldingiana*), *Coccoloba sp.*, and angélica (*Guettarda angelica*); and the leeward side mainly by trees up to 6 m tall: cajueiro (*Anacardium occidentale*), jatobá (*Hymenaea courbaril*), pau d'óleo (*Copaifera langsdorffii*) and pereiro (*Aspidosperma pyrifolium*).

In the neighbouring state of Ceará, Johnson (1972, 1973) described populations of *A. occidentale* and associated vegetation which strongly resemble those of the stabilised dunes of Parnaíba and Ilha Grande municipalities in Piauí (see our Figs. 9, 10). Johnson's study included seven coastal sites in north-eastern Brazil, and his most detailed observations are from Ceará. He characterised natural populations of *A. occidentale* as components of restinga vegetation on dunes or beach ridges growing in association with both stabilised and active sand dunes. Features of these populations were constant exposure to salt-laden winds, a low, semi-prostrate surface growth habit with the main branches appearing low on the trunk, often becoming prostrate and even rooting adventitiously, and the capacity to resist burial by sand dunes by the subterranean development of the trunk and branch system so that the leafing shoots remain exposed above the dune surface (see our Fig. 9C). He described the woody plant associations in which natural restinga populations of *A. occidentale* occur, characterising them as thickets usually up to 2 m high, occurring as discrete units separated by more open ground (our Fig. 9E). He listed 29 species in 16 families as woody elements of this community, including *Byrsonima crassifolia*, *Cereus jamacaru*, *Coccoloba*, and species of Myrtaceae, Dilleniaceae, and Melastomataceae, among other families.

More general discussions in the literature record that the cashew requires open habitats such as restinga because it is a heliophyte (Johnson 1972: 73), that it is a salt-tolerant species (halophilous, Decker 1936, Johnson 1972: 77) or that it is intolerant of competition from other woody species (Paiva et al. 2003).

Recent work by Brazilian geomorphologists in the northern coastal region has offered a historical-ecological scenario in which the restinga ecotype of *A. occidentale* could have evolved. Guedes et al. (2017) identified a vast area (ca. 16,000 km^2) of north-eastern Maranhão, including the Lençóis Maranhenses active dune field and parts of the Rio Parnaíba Delta, as a Quaternary dune field system, one of the largest in South America (see our Fig. 1B). During the last glacial maximum (ca. 12,000 years B.P.) the area was probably even larger due to lower sea levels. Apart from the Lençóis Maranhenses (Hilbert et al. 2016) and smaller areas of active dunes (e.g. in Ilha Grande, Piauí, Fig. 1D), this area consists today of ancient dune fields which these authors, using optically stimulated luminescence quartz dating, estimated to have stabilised between 19,000 and 14,000 years ago as a result of increase in rainfall leading to reduction in wind-transported sand drift and consequent growth of vegetation. The implication of these findings for restinga biology is that the area of restinga habitat resembling that shown in Fig. 10C would have been enormously greater in the Quaternary of Maranhão and Piauí than it is now. The evolution of the restinga ecotype of *A. occidentale* could therefore have occurred as a pioneer form of the species adapted to exploiting new habitats created by dune formation, exhibiting specialised characteristics such as tolerance to sand burial, salt-laden winds, oligotrophic soils and intolerance of competition from other woody species.

Nomenclature of the restinga ecotype – In the present study we have adopted "restinga ecotype" as the name for the natural populations of *A. occidentale* in the coastal region. This is a working solution to the currently confused nomenclature of these plants, discussed in more detail below.

Strictly speaking, the name *A. occidentale* may be regarded as applicable to one of the domesticated forms of the species, since its type specimen is a plant from Sri Lanka grown in the seventeenth century (Mitchell & Mori 1987) and presumably derived ultimately from Brazilian genotypes introduced by Portuguese colonists in the 16th century (Archak et al. 2009). For many authors, however, especially ecologists and taxonomists, the taxon concept (Berendsohn 1995, 2003) of *A. occidentale* includes both domesticated and wild forms.

The ecological studies of Santos-Filho (2009) and Santos-Filho et al. (2010) on the restinga vegetation of Piauí used the name *A. occidentale*. Johnson (1972, 1973) named the restinga ecotype he studied in coastal north-eastern Brazil as *A. occidentale*, as did e.g. Fernandes et al. (1996), Freitas & Paxton (1998), Barros et al. (1999), Freitas et al. (2002), Paiva et al. (2003), Freitas et al. (2014). Wild forms in the cerrado biome are also named as *A. occidentale* by taxonomists and ecologists (Mitchell & Mori 1987, Oliveira 2004, Castro et al. 2009, Castro 2013, Silva et al. 2013, 2016, Barros et al. 2016, Luz & Pirani 2017, Ratter et al. 2017).

Using the same name, without further qualification, for natural and cultivated forms of economically important species is undesirable, an argument elaborated in detail by Spooner et al. (2003) and Hetterscheid & Brandenburg (1995) for plants in general. It makes it more difficult to distinguish the characteristics of domesticated plants and natural populations, a distinction of real importance, both for understanding the existing range of genetic variation, and for conservation purposes.

One alternative that has been adopted is to recognise natural forms of *A. occidentale*, both from the restinga and cerrado, as distinct species or varieties. The name *Anacardium occidentale* var. *nanum* occurs occasionally in applied literature on cashew genetic improvement (e.g. Agostini-Costa et al. 2004, Lucena 2006) and refers to the early dwarf cashew ("caju anão precoce") variety, but this name appears never to have been validly published (The International Plant Names Index 2018). Four other varieties have been validly described; *A. occidentale* var. *americanum* DC. from the Caribbean, var. *indicum* DC. from Indonesia and var. *longifolium* Presl from Mauritius are most likely domesticated plants. However, *A. occidentale* var. *gardneri* Engl. from Crato, Ceará probably represents a natural population since the label on Gardner's type specimen at the Kew herbarium reads: "Anacardium sp. This seems to differ from A. occidentale. The leaves are in no ways emarginate. The fruit, or rather thickened peduncle is not larger than a cherry – and the tree has a more upright habit than the species on the coast. . . . campo and in the Catingas of Crato – a large tree. Sept. 1838." Müller et al. (2017) studied collections made in the municipality of Crato more recently, using the name *A. microcarpum* Ducke.

Many other studies have used the name *A. microcarpum* Ducke, both for the restinga ecotype in Piauí (Vieira et al. 2014), as well as for plants from other regions and vegetation types in Brazil (e.g. Barros et al.

1999, Crisóstomo et al. 2002, Ribeiro & Tabarelli 2002, Paiva et al. 2003, Agostini-Costa et al. 2004, Müller et al. 2013, Barbosa Filho et al. 2014, Barbosa Filho 2015, Müller et al. 2017). A review of the literature suggests that whatever its correct application may be, the geography, stature and ecology of the original *A. microcarpum* appear to differ significantly from the restinga ecotype in Piauí. Ducke's original plants were from the lower region of the Rio Amazonas in Pará state and were described by him as trees of 4-6 m height in "campos cobertos" vegetation, with the vernacular names "cajú do campo" or "cajú-y" (Ducke 1922). He cited specimens occurring from Santarem in north-western Pará to the Ilha do Marajó at the mouth of the Amazonas and distinguished *A. microcarpum* from *A. occidentale* by its smaller stature, harder coriaceous leaves, more flattened petiole, thicker midrib, laxer inflorescence, smaller drupes, and yellow to red hypocarps no longer than the drupe and with a more acid taste. There is no mention of dunes and the habitat description is closer to cerrado vegetation than restinga. Many studies linked to genetic improvement used accessions determined as *A. microcarpum* and cultivated at the major cashew germplasm collection of the Campo Experimental de Pacajus in Ceará, belonging to Embrapa Agroindústria Tropical (e.g. Barros et al. 1999, Crisóstomo et al. 2002, Paiva et al. 2003, Agostini-Costa et al. 2004, Almeida 2009). However, the published information about these accessions suggests that they are from wild cerrado populations in north-eastern Brazil, rather than those of coastal restingas. Our conclusions are that this scientific name has been used most widely for a cerrado ecotype of north-eastern Brazil rather than the restinga ecotype and that its correct application will remain unclear until the Amazonian populations described by A. Ducke have been restudied.

Other authors have avoided assigning a species name (*Anacardium spp.*) but this leads to imprecision. Studies including material of both restinga and cerrado ecotypes have adopted this solution (e.g. Rufino 2004, Rufino et al. 2007, Maia et al. 2012, Gomes et al. 2013, Borges 2015, Ferreira et al. 2015, Gomes 2017).

At the present early stage of our understanding of intra-specific variation in natural coastal populations of *A. occidentale* an informal name seems most appropriate, and the ecotype terminology already coined by Mitchell & Mori (1987) serves the purpose. These authors first used ecotype terminology for different natural forms of *A. occidentale* and distinguished restinga and cerrado ecotypes. Other authors (e.g. Pessoni 2007) have continued to use this terminology, which is a useful way to frame natural variation of the species complex until the delimitation of formally named taxa (e.g. subspecies) may become justifiable.

Mitchell & Mori's restinga ecotype includes records from restingas not only from coastal north-eastern Brazil, but also from as far south as coastal São Paulo state, a geographical distribution that includes a great diversity of restinga habitats and floras. Santos-Filho (2009) and Santos-Filho et al. (2010), for example, draw attention to the different climate and floristic composition of the restingas of the northern coast of Brazil. As far as we are aware there are no studies of the habit and ecology of restinga populations of *A. occidentale* throughout this range. However, that by Johnson (1972, 1973, 1974), especially his treatment of the natural restinga populations of Ceará, is sufficiently detailed to make us confident that the Piaui dune populations represent the same taxonomic entity.

Mitchell & Mori (1987: p. 40, Table IV) distinguished the cerrado ecotype by its "undulate, thickly coriaceous leaves with short, stout petioles" and usually smaller and more acid hypocarps, whereas the restinga ecotype has chartaceous leaves and longer petioles. The strongest difference shown in their table is the much longer relative length of leaf blade in relation to the petiole in the cerrado ecotype. Although our study did not sample wild populations of the cerrado ecotype, our results agreed to some extent in that the semi-domesticated population at Cabeceiras had the highest blade-petiole length ratio and the thickest leaf blades. However absolute petiole length was greatest in the majority of non-dune populations and the dune populations we studied tended to have low values of the leaf variables tabled by Mitchell & Mori (Supplementary Table S3).

The vernacular name "Cajuí" - The natural restinga ecotype populations of Piauí are known locally as "cajuí" (e.g. Rufino et al. 2002, 2007, 2008a, 2008b, Rufino 2004, Maia et al. 2012, Vieira et al. 2014, Borges 2015). Rufino et al. (2002, 2007) and Rufino (2004) studied the fruits and hypocarps of these populations from the standpoint of their potential as a source of food and income for the local rural people who harvest the plants in a traditional extractive use pattern (Figs. 10A, B, D). Different genotypes collected from natural populations were evaluated for a range of food-related chemical characteristics, focussing on the hypocarp (cashew apple), and a survey was made of local usage (Rufino 2004). Rufino et al. (2007) made detailed recommendations for a more organised agricultural production from plants derived from these natural populations. Rufino et al. (2008b) presented the results of a survey of local people in the municipalities of Ilha Grande and Parnaíba involved in the extractive exploitation of the natural populations of the Dunes ecotype (cajuí). Comissão Ilha Ativa (2014) published an illustrated guide to sustainable extractivism of the same populations in the municipalities of Parnaíba and Ilha Grande that we studied. Gomes et al. (2013) studied non-domesticated genotypes of cajuí (determined as *Anacardium spp.*) from the cerrados of Piauí and Gomes (2017) used the name cajuí (also determined as *Anacardium spp.*) in a study of natural cerrado populations from Piauí (Sete Cidades National Park).

Cajuí, however, is also the vernacular name of a wide range of *Anacardium* species, from small shrubs with underground stems in cerrado vegetation to huge Amazonian trees, the only common characteristic of which is that their drupes or hypocarps are of smaller size than the domesticated cashew ("caju" in Brazil). Mitchell & Mori (1987), Rufino (2004, after Lima 1988) and Agostini-Costa et al. (2006) cited "cajuí" or "cajuhy" as one of the vernacular names of *A. corymbosum* Barb.Rodr., *A. giganteum* J.Hancock ex Engl., *A. humile* A.St.-Hil., *A. microsepalum* Loes., *A. nanum* A.St.-Hil., *A. occidentale,* and *A. spruceanum* Benth. ex Engl., and also *A. amilcarianum* Machado, *A. microcarpum* Ducke, *A. othonianum* Rizzini, which are currently regarded as synonyms of *A. occidentale*. The Brazilian cashew agricultural industry is more precise; drupes that weigh less than 3.3 g after harvesting are classified as cajuí (Rufino et al. 2007). Pontes & Ribeiro (2006) define cajuí as smaller-sized "fruits" (drupe and hypocarp) of either cerrado plants or domesticated cashew trees.

The cerrado ecotype in Piauí – *Anacardium occidentale* also occurs naturally in the cerrado biome of Brazil (Mitchell & Mori 1987, Luz & Pirani 2017). Ratter et al. (2017) cite it as one of the dominant species of the core cerrado flora, occurring in twelve Brazilian states, including southern Piauí and southern Ceará. There are many records of the species from vegetation studies carried out in Piauí which indicate that this species is a natural component of more than one type of cerrado vegetation. Castro (2013) provided details for Piauí, showing that this vegetation extends to the littoral region of the state. In the northern (but not littoral) sector of Piauí, Oliveira (2004) cited *A. occidentale* for the Sete Cidades National Park as a prominent species of the forest types "cerradão", and "floresta estacional decídua" and the mainly grassland vegetation type "campo graminóide cespitoso médio". Silva et al. (2013, 2016) also recorded *A. occidentale* (determined as "cajuí") for the semi-deciduous seasonal forest of this national park. Barros et al. (2016) recorded *A. occidentale* for various vegetation types within the vegetational complex at Campo Maior, Piauí (ENE of Teresina), where it was best represented in the "Capões Florestados", discontinuous woodlands that indicate a transition between drier and damper ground, and also in cerrado *sensu stricto*. The records by Ribeiro & Tabarelli (2002) of *A. microcarpum* in three types of cerrado *sensu stricto* in a vegetation study southeast of Teresina can also probably be ascribed to the cerrado ecotype of *A. occidentale*. In southern Piauí, Castro et al. (2009) cited *A. occidentale*, with the vernacular name "cajuí", for the vegetation of Serra Vermelha, classified as "Floresta Estacional Semidecidual de Transição", but since the authors did not include it in their sampling it must have been rare there.

Conclusions

- The natural (dune) populations are in general more similar to one another than to domesticated or semi-domesticated (nondune) populations, but there is morphological overlap with the nondune populations geographically nearest to them such that no single variable separates the dune from the nondune category. Nevertheless, statistical classification methods showed a high (96.4%) success rate of assignment of individuals to the dune category overall and to dune populations individually (86% to 100%).
- Of the variables studied, the most statistically discriminatory for the natural (dune) populations were their shorter, broader leaf blades, shorter drupes and fewer leaf secondary veins.
- In the natural dune populations, the least variable characters were drupe length and number of secondary veins in the leaf blade, and the most variable were hypocarp and petiole length.
- The range of within-population morphological variation in the natural populations studied is similar to, if slightly less than, that of the semi-domesticated ones.
- There is a significant correlation between interpopulation geographical distance and their morphological dissimilarity.
- The similarity relations of the populations sampled are explained best by their proximity to the coast. Non-natural (nondune) populations nearest the coast show strongest similarity to the natural (dune) populations. This could be due either to gene flow between wild and domesticated genotypes, or to non-inherited effects of the coastal environment or both.
- Populations distant from the coast are most divergent from one another. The Cabeceiras population, in a region where cerrado is the natural vegetation type, is the most distinct of all populations studied, notably by its distinctly thicker leaf blades and narrower petioles.
- The morphometric differentiation of the populations growing on stabilised dunes, together with their occurrence as components of a characteristic restinga vegetation community, supports their recognition as a natural form of *A. occidentale*, i.e. as an informal taxonomic group; we recommend the use of the

previously proposed term “restinga ecotype” as the working name for the natural populations of the coast.

- Further progress in understanding taxonomic structure within *A. occidentale* requires wider comparison of natural populations of the restinga ecotype along the Brazilian coastline in other states, using both morphological and molecular data, and between these and natural cerrado ecotype populations.

The authors are grateful to the Piauí State Funding Agency FAPEPI (Fundação de Amparo à Pesquisa do Estado do Piauí) for providing grants and financial support to M.V. Sousa and to the Parnaíba Municipal Prefecture (Prefeitura Municipal de Parnaíba) EDITAL N. 001/2014 – Edital Mac-Doubles Fernandes do Nascimento de Apoio a Ciência, Tecnologia e Inovação, Termo 037/2015, for funding. We also thank Research Assistant Pedro Neves of EMBRAPA Meio-Norte, Parnaíba for generously sharing with us agronomic information on Cajuí plants. S. Mayo thanks the Royal Botanic Gardens Kew for logistic support.

References

AGOSTINI-COSTA, T.S.; JALES, K.J.; GARRUTI, D.S.; PADILHA, V.A.; LIMA, J.B.; AGUIAR, M.J. & PAIVA, J.R. 2004: Teores de ácido anacárdico em pedúnculos de cajueiro *Anacardium microcarpum* e em oito clones de *Anacardium occidentale var. nanum* disponíveis no Nordeste do Brasil. – Ciência Rural, Santa Maria **34**(4): 1075–1080.

AGOSTINI-COSTA, T.S.; FARIA, J.P.; NAVES, R.V. & VIEIRA, R.F. 2006: Cajus do cerrado. – In: VIEIRA, R.F.; AGOSTINI-COSTA, T.S.; SILVA, D.B.; FERREIRA, F.R. & SANO, S.M. (eds.), Frutas nativas da Região Centro-Oeste do Brasil. – Embrapa Recursos Genéticos e Biotecnologia, Brasília, pp. 136–151.

ALIYU, O.M. & AWOPETU, J.A. 2007: Multivariate analysis of cashew (*Anacardium occidentale* L.) germplasm in Nigeria. – Silvae Genetica **56**(3-4): 170–179.

ALMEIDA, A.S. 2009: Qualidade, compostos bioativos e atividade antioxidante total de pedúnculos de cajuizeiros e frutos de umbuzeiros nativos do semi-árido do Piauí. – PhD thesis, Universidade Federal Rural do Semi-Árido, Mossoró, Brazil.

ARAUJO, D.S.D.; SÁ, C.F.C.; FONSECA-KRUEL, V.S.; PEREIRA, M.C.A.; MACIEL, N.C.; SÁ, R.C.; ARAUJO, A.D.; KRUEL, G.; ANDRADE, L.R. & PEREIRA, O.J. 2018: Restinga net. – Available at: http://www.restinga.net. Accessed 3 June 2018.

ARCHAK, S.; GAIKWAD, A.B.; GAUTAM, D.; RAO, E.V.V.B.; SWAMY, K.R.M. & KARIHALOO, J.L. 2003: Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. – Genome **46**: 362–369.

ARCHAK, S.; GAIKWAD, A.B.; SWAMY, K.R.M. & KARIHALOO, J.L. 2009: Genetic analysis and historical perspective of cashew (*Anacardium occidentale* L.) introduction into India. – Genome **52**: 222–230.

BARBOSA FILHO, V.M. 2015: Avaliação do potencial antioxidante e antibacteriano de Anacardium microcarpum e sua toxicidade em diferentes modelos biológicos in vitro. – PhD thesis, Programa de Pós-Graduação em Ciências Biológicas, Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brazil.

BARBOSA FILHO, V.M.; WACZUK, E.P.; KAMDEM, J.P.; ABOLAJI, A.O.; LACERDA, S.R.; COSTA, J.G.M.; MENEZES, I.R.A.; BOLIGON, A.A.; ATHAYDE, M.L.; ROCHA, J.B.T. & POSSER, T. 2014: Phytochemical constituents, antioxidant activity, cytotoxicity and osmotic fragility effects of Caju (*Anacardium microcarpum*). – Industrial Crops and Products **55**: 280–288.

BARROS, J.S.; HARIDASAN, M.; CASTRO, A.A.J.F.; FARIAS, R.R.S. & MENDES, M.R.A. 2016: Uma visão geológica sobre a florística e fitossociologia de áreas ecotonais: complexo vegetacional de Campo Maior, Piauí. – In: ALMEIDA, E.B. & SANTOS-FILHO, F.S. (eds.), Biodiversidade do Meio Norte do Brasil: Conhecimentos ecológicos e aplicações. – Editora CRV, Curitiba, pp. 45–66.

BARROS, L.M. 1991: Caracterização morfológica e isoenzimática do cajueiro (*Anacardium occidentale* L.) tipos comum e anão-precoce, por meio de técnicas multivariadas. – PhD thesis, Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Piracicaba, Brazil.

BARROS, L.M.; PAIVA, J.R. & CAVALCANTI, J.J.V. 1999: Recursos genéticos de cajueiro: situação atual e estratégias para o futuro. – In: QUEIROZ, M.A.; GOEDERT, C.O. & RAMOS, S.R.R. (eds.), Recursos genéticos e melhoramento de plantas para o Nordeste brasileiro. – Embrapa Semi-Árido, Petrolina, pp. 1–22. – Available at: http://www.ceinfo.cnpat.embrapa.br/arquivos/artigo_2561.pdf . Accessed on 9 Dec. 2017.

BERENDSOHN, W.G. 1995: The concept of “potential taxa” in databases. – Taxon **44**: 207–212.

BERENDSOHN, W.G. 2003: MoReTax - Handling factual information linked to taxonomic concepts in biology. – Schriftenreihe für Vegetationskunde. Vol. **39**. – Bundesamt für Naturschutz, Bonn.

BORCARD, D.; GILLET, F. & LEGENDRE, P. 2011: Numerical ecology with R. – Springer, New York.

BORGES, A.N.C. 2015: Caracterização genética em germoplasma de cajuí (Anacardium spp.) por meio de marcadores morfoagronômicos e moleculares ISSR. – MSc thesis, Programa de Pós-graduação em Genética e Melhoramento, Universidade Federal do Piauí, Teresina, Brazil.

CALLADO, S.M.G. 2008: Environmental Sustainability Analysis of Cashew Systems in North-east Brazil. – PhD thesis, Hohen Landwirtschaftlichen Fakultät der Rheinischen-Wilhelms-Universität zu Bonn. – Available at: http://hss.ulb.uni-bonn.de/2009/1676/1676.htm. Accessed 3 June 2018.

CARVALHO, R.S. 2011: Estudos de acessos de *Anacardium humile* A.St.Hil. por meio de caracterização morfológica e de marcadores RAPD. – MSc thesis, Universidade Federal de Goiás, Jataí, Brazil.

CARVALHO, R.S.; PINTO, J.F.N.; REIS, E.F.; SANTOS, S.C. & DIAS, L.A.S. 2012: Variabilidade genética de cajuzinho-do-cerrado (*Anacardium humile* St.Hil.) por meio de marcadores RAPD. – Rev. Bras. Frutic. Jaboticabal, São Paulo **34**(1): 227–233.

CASTRO, A.A.J.F. 2007: Unidades de planejamento: uma proposta para o Estado do Piauí com base na dimensão diversidade de ecossistemas. – Publicações Avulsas em

Conservação de Ecossistemas, Universidade Federal do Piauí, CCN Biologia, Teresina, Piauí **18**: 1–28.

CASTRO, A.A.J.F. 2013: Cerrados do Nordeste do Brasil e Ecótonos Associados: Sítio 10 – Sítio Ecocem. – In: TABARELLI, M.; ROCHA, C.F.D.; ROMANOWSKI, H.P.; ROCHA, O. & LACERDA, L.D. (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, A.A.J.F.; CASTRO, A.S.F.; FARIAS, R.R.S.; SOUSA, S.R.; CASTRO, N.M.C.F.; SILVA, C.G.B.; MENDES, M.R.A.; BARROS, J.S. & LOPES, R.N. 2009: Diversidade de espécies e de ecossistemas da vegetação remanescente da Serra Vermelha, área de chapada, municípios de Curimatá, Redenção do Gurgéia, e Morro Cabe no Tempo, sudeste do Piauí. – Publicações Avulsas em Conservação de Ecossistemas, Universidade Federal do Piauí, CCN Biologia, Teresina, Piauí **23**: 1–72.

CHIPOJOLA, F.M.; MWASE, W.F.; KWAPATA, M.B.; BOKOSI, J.M.; NJOLOMA, J.P. & MALIRO, M.F. 2009: Morphological characterization of cashew (*Anacardium occidentale* L.) in four populations in Malawi. – African Journal of Biotechnology **8**(20): 5173–5181.

CLAUDE, J. 2008: Morphometrics with R. – Springer, New York.

COMISSÃO ILHA ATIVA. 2014: Cajuí: Boas práticas de manejo para o extrativismo sustentável. – Editora Sieart, Parnaíba, Piauí.

CORRÊA, M.P.F.; RUFINO, M.S.M.; VASCONCELOS, L.F.L.; RIBEIRO, V.Q.; SOARES, E.B. & ARAÚJO, E.C.E. 2002: Germinação e vigor de sementes de genótipos de cajuí (*Anacardium* spp.). – In: CONGRESSO BRASILEIRO DE FRUTICULTURA (eds.), Os novos desafios da fruticultura brasileira. – Anais do Congresso Brasileiro de Fruticultura 17, 2002, Belém, pp. 1–5.

CRAWLEY, M.J. 2013: The R Book. Second Edition. – Wiley, Chichester, U.K.

CRISÓSTOMO, J.R.; CAVALCANTI, J.J.V.; BARROS, L.M.; ALVES, R.E.; FREITAS, J.G. & OLIVEIRA, J.N. 2002: Melhoramento do Cajueiro-anão-precoce: Avaliação da qualidade do pedúnculo e a heterose dos seus híbridos. – Revista Brasileira de Fruticultura **24**(2): 477–480.

DASMOHAPATRA, R.; RATH, S.; PRADHAN, B.D. & ROUT, G.R. 2014: Molecular and agromorphological assessment of cashew (*Anacardium occidentale* L.) genotypes in India. – Journal of Applied Horticulture **16**(3): 215–221.

DECKER, J.S. 1936: Aspectos biológicos da flora brasileira. – Rotermund, São Leopoldo, Brazil.

DESAI, A.R. 2008: Molecular diversity and phenotyping of selected cashew genotypes of Goa, and physiological response of cv. GOA-1 to *in situ* moisture conservation. – PhD thesis, University of Agricultural Sciences, Dharwad, India.

DUCKE, A. 1922: Plantes nouvelles ou peu connues de la région amazonienne (II^{e} Partie). – Archivos do Jardim Botanico do Rio de Janeiro **3**: 3–269.

ENGLER, A. 1883: Anacardiaceae. – In: CANDOLLE, A. & CANDOLLE, C. (eds.), Monographiae Phanerogamarum, Vol. **4**. – G. Masson, Paris, pp. 171–500.

FERNANDES, A.G.; LOPES, A.S.; SILVA, E.V.; CONCEIÇÃO, G.M. & ARAÚJO, M.F.V. 1996: IV. Componentes Biológicos: Vegetação. – In: FUNDAÇÃO CEPRO (eds.), Macrozoneamento costeiro do Estado do Piauí. – Fundação Rio Parnaíba, Teresina, pp. 43–72.

FERREIRA, T.N.; FERREIRA, E.B.; FERREIRA, T.F.C.; ROCHA, A.J.G. & KARSTEN, J. 2015: Características físicas do cajuí (*Anacardium* sp.) do cerrado baiano. – Anais do I Congresso Brasileiro de Processamento mínimo e Pós-colheita de frutas, flores e hortaliças, Aracaju, Sergipe. 8 p. – Available at: http://octeventos.com/site/sites/pos-colheita/anais/PDF/00068_084913.pdf. Accessed 4 June 2018.

FILZMOSER, P. 2018: StatDA: Statistical Analysis for Environmental Data – R package, version 1.7. – Available at: https://CRAN.R-project.org/package=StatDA. Accessed 1 June 2018.

FILZMOSER, P. & GSCHWANDTNER, M. 2018: mvoutlier: Multivariate Outlier Detection based on Robust Methods. – R package, version 2.0.9. – Available at: https://CRAN.R-project.org/package=mvoutlier. Accessed 3 Apr. 2018.

FOULKES, A.S. 2009: Applied Statistical Genetics with R: For Population-Based Association Studies. – Springer, New York.

FOX, J.; WEISBERG, S. & PRICE, B. 2018: car: Companion to Applied Regression. – R package, version 3.0-0. – Available at: https://CRAN.R-project.org/package=car. Accessed 3 Apr. 2018.

FREITAS, B.M. & PAXTON, R.J. 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. – Journal of Applied Ecology **35**: 109–121.

FREITAS, B.M.; PAXTON, R.J. & HOLANDA-NETO, J.P. 2002: Identifying pollinators among an array of flower visitors, and the case of inadequate cashew pollination in NE Brazil. – In: KEVAN, P. & IMPERATRIZ FONSECA, V.L. (eds.), Pollinating Bees – The conservation link between agriculture and nature. – Ministry of Environment, Brasília, pp. 229–244.

FREITAS, B.M.; PACHECO FILHO, A.J.P.; ANDRADE, P.B.; LEMOS, C.Q.; ROCHA, E.E.M.; PEREIRA, N.O.; BEZERRA, A.D.M.; NOGUEIRA, D.S.; ALENCAR, R.L.; ROCHA, R.F. & MENDONÇA, K.S. 2014: Forest remnants enhance wild pollinator visits to cashew flowers and mitigate pollination deficit in NE Brazil. – Journal of Pollination Ecology **12**(4): 22–30.

FROTA JÚNIOR, J.I.; LIMA, E.N.; ARAÚJO, M.E.B.; LIMA, I.B. & LIMA, J.S. 2014: Divergência genética entre progênies de meio-irmãos de cajueiro anão precoce utilizando caracteres morfológicos. – Agrarian Academy, Centro Científico Conhecer, Goiânia **1**(2): 47–59.

GOMES, J.G.; OLIVEIRA, N.C.C.; LOPES, P.S.N. & BALLESTEIRO, M. 2011: 12630 – Componentes principais no estudo da diversidade de cajuí na sub-bacia do Rio dos Cochos. – Cadernos de Agroecologia **6**(2): 1–5. – Available at: http://aba-agroecologia.org.br/revistas/index.php/cad/article/view/12630. Accessed on 6 Dec. 2017.

GOMES, M.F.C. 2017: Diversidade e estrutura genética em populações de *Anacardium* spp. do Parque Nacional de Sete Cidades (PI) por meio de marcadores ISSR. – MSc dissertation, Programa de Pós-graduação em Genética e Melhoramento, Universidade Federal do Piauí, Teresina, Piauí, Brazil.

GOMES, S.O.; SOUZA, V.A.B.; COSTA, M.P.S.D.; SILVA, C.C.P.; VALE, E.M.; SOUSA, M. & SOUSA, J.P.B. 2013: Avaliação da qualidade física e química de cajuí (*Anacardium spp.*) na região Meio-Norte. – Revista Geintec **3**(3): 139–145.

GOOGLE EARTH. 2018: Google Earth Pro version 7.1.8.3036 (32-bit), build date 17 Jan 2017.

GUEDES, C.C.F.; GIANNINI, P.C.F.; SAWAKUCHI, A.O.; DEWITT, R. & AGUIAR, V.A.P. 2017: Weakening of northeast trade winds during the Heinrich stadial 1 event recorded by dune field stabilization in tropical Brazil. – Quaternary Research **88**(3): 369–381.

HARRELL Jr., F.E. 2018: Harrell Miscellaneous. – R package, version 4.1-1. – Available at: https://CRAN.R-project.org/package=Hmisc. Accessed 3 Apr. 2018.

HERVÉ, M. 2018: RVAideMemoire: Testing and Plotting Procedures for Biostatistics. – R package, version 0.9-69. – Available at: https://CRAN.R-project.org/package=RVAideMemoire. Accessed 3 Apr. 2018.

HETTERSCHEID, W.L.A. & BRANDENBURG, W.A. 1995: Culton versus taxon: conceptual issues in cultivated plant systematics. – Taxon **44**: 161–175.

HIJMANS, R.J. 2017: geosphere: Spherical Trigonometry. – R package, version 1.5-7. – Available at: https://CRAN.R-project.org/package=geosphere. Accessed 8 March 2018.

HILBERT, N.N.; GUEDES, C.C.F. & GIANNINI, P.C.F. 2016: Morphologic and sedimentologic patterns of active aeolian dune-fields on the east coast of Maranhão, northeast Brazil. – Earth Surface Processes and Landforms **41**(1): 87–97.

JENA, R.A.; SAMAL, K.S.; PAL, A.; DAS, B.K. & CHAND, P.K. 2016: Genetic diversity among some promising Indian local selections and hybrids of cashew nut based on morphometric and molecular markers. – International Journal of Fruit Science **16**(1): 69–93.

JOHNSON, D.V. 1972: The cashew of northeast Brazil: a geographical study of a tropical tree crop. – PhD thesis, University of California, Los Angeles. USA.

JOHNSON, D. 1973: The botany, origin, and spread of the cashew, *Anacardium occidentale* L. – Journal of Plantation Crops **1**: 1–7.

JOHNSON, D.V. 1974: O caju do Nordeste do Brasil: um estudo geográfico. –Translation by José Alexandre Robatto Orrico. ETENE, Fortaleza, 169 p. [cited in Pessoni 2007]

LEGENDRE, P. & LEGENDRE, L. 2012: Numerical Ecology. Third English Edition. – Elsevier, Amsterdam.

LIMA, V.P.M.S. 1988: A cultura do cajueiro no nordeste do Brasil. – ETENE-Banco do Nordeste, Fortaleza.

LOECHER, M. 2016: RgoogleMaps: Overlays on Static Maps. – R package, version 1.4.1. – Available at: https://CRAN.R-project.org/package=RgoogleMaps. Accessed 3 Apr. 2018.

LUCENA, V.M.X. 2006: Diversidade genética entre genótipos de cajueiro (*Anacardium occidentale* L.) e qualidade do fruto e pseudofruto. MSc thesis, Universidade Federal de Roraima, Boa Vista. Brazil.

LUZ, C.L.S. & PIRANI, J.R. 2017: Anacardiaceae. – In: Flora do Brasil 2020 (under construction). Jardim Botânico do Rio de Janeiro. – Available online at: http://reflora.jbrj.gov.br/reflora/floradobrasil/FB77853. Accessed on 21 Nov. 2017

MACHADO, O.X.B. 1949: Três novas espécies de *Anacardium* do Brasil Central. – Arquivos do Jardim Botânico do Rio de Janeiro **9**: 87–93.

MAIA, M.C.C.; OLIVEIRA, L.C.; ALMEIDA, A.S.; ARAÚJO, L.B.; VASCONCELOS, L.F.L. & LACERDA, M.N. 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.

MAIA, M.C.C.; ALMEIDA, A.S.; MACEDO, L.M.; RESENDE, M.D.V.; LACERDA, M.N.; VASCONCELOS, L.F.L.; OLIVEIRA, L.C. & SILVA, R.B. 2016: Heritability, repeatability, and genetic gains in a improvement population of cajuízeiro. – Revista Árvore, Viçosa-MG **40**(4): 715–722.

MASAWE, P.A.L.; ESUGU, J.F.O.; KASUGA, L.J.F.; MNENEY, E.E. & MUJUNI, D. 2013: Proceedings of the Second International Cashew Conference: Cashew, People and Environment. – CAB International, Wallingford, Oxford.

MATOS FILHO, C.H.A. 2017: Diversidade genética e seleção em populações de cajueiro comum. PhD thesis, Programa de Pós-Graduação em Agronomia-Agricultura Tropical, Universidade Federal do Piauí, Teresina, Brazil.

MITCHELL, J.D. & MORI, S.A. 1987: The cashew and its relatives (*Anacardium*: Anacardiaceae). – Memoirs of the New York Botanical Garden **42**: 1–76.

MITCHELL, J.D. 1992: Additions to *Anacardium* (Anacardiaceae). *Anacardium amapaënse*, a new species from French Guiana and eastern Amazonian Brazil. – Brittonia **44**: 331–338.

MÜLLER, K.R.; BARBOSA FILHO, V.M.; SILVA, G.F.; MENEZES, I.A.; ZEMOLIN, A.P.; FRANCO, J.L. & POSSER, T. 2013: Avaliação do potencial neuroprotetor do extrato de *Anacardium microcarpum* no modelo de Parkinsonismo induzido por Paraquat em *Drosophila melanogaster*. – BBR-Biochemistry and Biotechnology Reports, Número Especial **2**(3): 150–153.

MÜLLER, K.R.; MARTINS, I.K.; RODRIGUES, N.R.; CRUZ, L.C.; BARBOSA FILHO, V.M.; MACEDO, G.E.; SILVA, G.F.; KAMDEM, J.P.; MENEZES, I.R.A.; FRANCO, J.L. & POSSER, T. 2017: *Anacardium microcarpum* extract and fractions protect against Paraquat-induced toxicity in *Drosophila melanogaster*. – EXCLI Journal **16**: 302–312.

NAYAK, M.G.; MOHANA, G.S.; BHAT, P.S.; SAROJ, P.L.; SWAMY, K.R.M. & BHAT, M.G. 2014: Catalogue V – Minimum Descriptors of Cashew Germplasm Accessions. – ICAR-Directorate of Cashew Research, Puttur, Karnataka.

OKSANEN, J.; BLANCHET, F.G.; FRIENDLY, M.; KINDT, R.; LEGENDRE, P.; MCGLINN, D.; MINCHIN, P.R.; O'HARA, R.B.; SIMPSON, G.L.; SOLYMOS, P.; STEVENS, M.H.H.; SZOECS, E. & WAGNER, H. 2018: vegan: Community Ecology Package. – R package, version 2.4-6. – Available online at: https://CRAN.R-project.org/package=vegan. Accessed on 10 March 2018.

OLIVEIRA, M.E.A. 2004: Mapeamento, florística e estrutura da transição campo-floresta na vegetação (cerrado) do Parque Nacional de Sete Cidades, nordeste do Brasil. – PhD thesis, Instituto de Biologia da Universidade Estadual de Campinas, Campinas, São Paulo, Brazil.

PAIKRA, M.S. 2016: Studies on genetic divergence in cashew (*Anacardium occidentale* L.). – PhD thesis, Department of Horticulture, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India.

PAIVA, J.R.; CRISÓSTOMO, J.R. & BARROS, L.M. 2003: Recursos genéticos do cajueiro: coleta, conservação, caracterização e utilização. Documentos 65. – Embrapa Agroindústria Tropical, Fortaleza, 43 p.

PESSONI, L.A. 2007: Estratégias de análise de diversidade em germoplasma de cajueiro *(Anacardium spp.)*. – PhD thesis, Universidade Federal de Viçosa, Minas Gerais, Brazil.

PONTES, A.L. & RIBEIRO, R.M. 2006: Vocabulário da cultura e da industrialização do caju. – Embrapa Informação Tecnológica, Brasília, Brazil.

R CORE TEAM. 2018: R: A Language and Environment for Statistical Computing. – R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

RATTER, J.A.; BRIDGEWATER, S.; RIBEIRO, J.F.; FONSÊCA-FILHO, J.; RODRIGUES DA SILVA, M.; MILLIKEN, W.; PULLAN, M.; POTT, A.; OLIVEIRA-FILHO, A.; DURIGAN, G. & PENNINGTON, R.T. 2017 (ongoing): Analysis of the floristic composition of the Brazilian cerrado vegetation IV: Presentation of a Revised Data-Base of 367 Areas. – Available at: http://cerrado.rbge.org.uk/cerrado/download/download.php. Accessed 26 November 2017.

RIBEIRO, E.K.M.; RÊGO, M.M.C. & MACHADO, I.C.S. 2008: Cargas polínicas de abelhas polinizadoras de *Byrsonima chrysophylla* Kunth (Malpighiaceae): fidelidade e fontes alternativas de recursos florais. – Acta bot. bras. **22**(1): 165–171.

RIBEIRO, L.F. & TABARELLI, M. 2002: A structural gradient in cerrado vegetation of Brazil: changes in woody plant density, species richness, life history and plant composition. – Journal of Tropical Ecology **18**: 775–794.

RIPLEY, B. 2015: class: Functions for Classification. – R package, version 7.3-14. – Available at: https://CRAN.R-project.org/package=class. Accessed 3 Apr. 2018.

RIPLEY, B. 2018: MASS: Support Functions and Datasets for Venables and Ripley's MASS. – R package, version 7.3-49. Available at: https://CRAN.R-project.org/package=MASS. Accessed 3 Apr. 2018.

RIZZINI, C.T. 1969: Espécies novas de árvores do Planalto Central brasileiro. – Anais Acad. Brasil. Ci. **41**(2): 239–244.

RUFINO, M.S.M. 2004: Qualidade e potencial de utilização de cajuís (*Anacardium spp.*) oriundos da vegetação litorânea do Piauí. – MSc thesis, Programa Regional de Pós-Graduação em Desenvolvimento e Meio Ambiente, Universidade Federal do Piauí, Teresina, Brazil.

RUFINO, M.S.M.; VASCONCELOS, L.F.L.; CORRÊA, M.P.F.; RIBEIRO, V.Q.; SOARES, E.B. & SOUZA, V.A.B. 2002: Caracterização física e química do fruto e pseudofruto de genótipos de cajuí (*Anacardium spp.*). – EMBRAPA. Available online at: https://www.embrapa.br/web/mobile/publicacoes/-/publicacao/60017/caracterizacao-fisica-e-quimica-do-fruto-e-pseudofruto-de-genotipos-de-cajui-anacardium-spp. Accessed on 30 November 2017.

RUFINO, M.S.M.; CORRÊA, M.P.F.; ALVES, R.E.; BARROS, L.M. & LEITE, L.A.S. 2007: Suporte tecnológico para a exploração racional do cajuizeiro. – Documentos (Embrapa) **107**: 1–31.

RUFINO, M.S.M.; ALVES, R.E. & ARAGÃO, F.A.S. 2008a: Análise multivariada de genótipos de cajuizeiro em áreas nativas da região Meio-Norte do Brasil. – Proc. Interamer. Soc. Trop. Hort. **52**: 140–143.

RUFINO, M.S.M.; CORRÊA, M.P.F.; ALVES, R.E.A. & LEITE, L.A.S. 2008b: Utilização atual do cajuí nativo da vegetação litorânea do Piauí, Brasil. – Proceedings of the Interamerican Society for Tropical Horticulture **52**: 147–149.

SAINT-HILAIRE, A. 1831: Observations sur le genre *Anacardium* et les nouvelles espècies qu'on doit y faire entrer. – Ann.Sci.Nat.Sér. I, **23**: 268–274.

SAMAL, S.; ROUT, G.R. & LENKA, P.C. 2003: Analysis of genetic relationships between populations of cashew (*Anacardium occidentale* L.) by using morphological characterisation and RAPD markers. – Plant Soil Environ. **49**(4): 176–182.

SANTOS, R.C. & SANTOS JUNIOR, J.E. 2015: Divergência genética por análise multivariada de caracteres fenotípicos de *Anacardium humile* (St. Hilaire). – Rev. Ceres, Viçosa **62** (6): 553–560.

SANTOS-FILHO, F.S.S. 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, Brazil.

SANTOS-FILHO, F.S. & ZICKEL, C.S. 2013: Origem e estrutura da costa e a vegetação de restinga: o caso do litoral do Piauí. – In: SANTOS FILHO, F.S.; SOARES, A.F.C.L. & ALMEIDA JR., E.B. (eds.), Biodiversidade do Piauí: Pesquisas e Perspectivas, Vol. **2**. – Editora CRV, Curitiba, pp. 11–36.

SANTOS-FILHO, F.S.; ALMEIDA JR., E.B.; SOARES, C.J.R.S. & ZICKEL, C.S. 2010: Fisionomias das restingas do Delta do Parnaíba, Nordeste, Brasil. – Revista Brasileira de Geografia Física **3**: 218–227.

SARKAR, D. 2017: lattice: Trellis Graphics for R. – R package, version 0.20-35. – Available at: https://CRAN.R-project.org/package=lattice. Accessed 3 Apr. 2018.

SETHI, K. 2015: Studies on morphological and molecular diversity in cashew (*Anacardium occidentale* L.) hybrids. – PhD thesis, Orissa University of Agriculture and Technology, Bhubaneswar, India.

SILVA, C.B.; CASTRO, A.A.J.F.; FARIAS, R.R.S. & RAMOS NETO, M.B. 2013: Flora lenhosa em fragmentos de floresta estacional semidecidual no Parque Nacional de Sete Cidades (PN7C), Piauí, Brasil. – In: SANTOS FILHO, F.S.; SOARES, A.F.C.L. & ALMEIDA JR., E.B. (eds.), Biodiversidade do Piauí: Pesquisas e Perspectivas, Vol. **2**. – Editora CRV, Curitiba, pp. 101–119.

SILVA, C.B.; CASTRO, A.A.J.F.; FARIAS, R.R.S.; SOUSA, S.R. & LOPES, R.N. 2016: Fitossociologia da vegetação lenhosa em fragmentos de floresta estacional semidecidual no Parque Nacional de Sete Cidades (PN7C), Piauí, Brasil. – In: ALMEIDA, E.B. & SANTOS-FILHO, F.S. (eds.), Biodiversidade do Meio Norte do Brasil: Conhecimentos ecológicos e aplicações. – Editora CRV, Curitiba, pp. 67–89.

SPOONER, D.M.; HETTERSCHEID, W.L.A.; VAN DEN BERG, R.G. & BRANDENBURG, W.A. 2003: Plant Nomenclature and Taxonomy. – Horticultural Reviews, **28**:1–60. – John Wiley & Sons, Oxford.

STEKHOVEN, D.J. 2013: missForest: Nonparametric Missing Value Imputation using Random Forest. – Available at: https://CRAN.R-project.org/package=missForest. Accessed 4 Apr. 2018.

TAKAGI, K. 2011: Convex Hull around scatter plot in R. – Available at: https://chitchatr.wordpress.com/2011/12/30/convex-hull-around-scatter-plot-in-r/ Accessed on 3 April 2018.

THE INTERNATIONAL PLANT NAMES INDEX. 2018: – Available at: http://www.ipni.org Accessed 1 February 2018.

THERNEAU, T. & ATKINSON, B. 2018: rpart: Recursive Partitioning and Regression Trees. R package, version 4.1-13. – Available at: https://CRAN.R-project.org/package=rpart. Accessed 3 May 2018.

VARMUZA, K. & FILZMOSER, P. 2009: Introduction to multivariate statistical analysis in chemometrics. – CRC Press, Boca Raton.

VIDAL NETO, F.C.; BARROS, L.M.; CAVALCANTI, J.J.V. & MELO, D.S. 2013: Melhoramento genético e cultivares de cajueiro. – In: ARAÚJO, J.P.P. (ed.), Agronegócio caju: práticas e inovações. – Embrapa, Brasília, pp. 481–508.

VIEIRA, M.; MAYO, S.J. & ANDRADE, I.M. 2014: Geometric morphometrics of leaves of *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the coastal region of Piauí, Brazil. – Braz. J. Bot. **37**(3): 315–327.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Figure S1. Principal component analysis, data set 1. Ordination of individual means data set using scaled data. Best viewed in electronic version.

Figure S2. Barplots of the variable loadings of a linear discriminant analysis (LDA) of data set 1, computed from the scaled version of the individual means data set, using populations as the categorical variable. A. First discriminant axis. B. Second discriminant axis.

Figure S3. Cross-validation results from the linear discriminant analysis (LDA) of the scaled individual means data set (data set 1) using populations as the categorical variable. Each bar shows a population as constituted by the LDA cross-validation test, in which the bar name is that of the population whose individuals comprise its largest component. The length of each different coloured component of the bar is proportional to the number of individuals of the original populations (see inset legend) present in it. The more dominant one such component is, the more consistent and distinct the original population may be inferred to be as regards its morphological variation. Best viewed in electronic version.

Figure S4. Linear discriminant analysis of mean vectors of populations (data set 1). Ordination of population means shown in the space of the first two discriminant axes. Best viewed in electronic version.

Table S1. Notes on multivariate morphometric studies of *Anacardium occidentale* and *A. humile* published between 1999-2017, tabled in chronological order. Some studies were not obtainable online and were not seen, but they are reported here as cited by other authors. Only postgraduate and undergraduate studies that are available online are reported here.

Table S2. Results of univariate analysis tests. A. Raw data set. B. Individual means data sets (population multivariate outliers removed, individual replicates averaged). C. Individual means data set (Data set 1 only): Results for tests between the natural and domesticated population categories (factor dune – nondune). N = Shapiro-Wilk test of normal distribution; Bartlett = Bartlett parametric test of homogeneity of variances of populations; Fligner = Fligner-Killeen non-parametric test for equality of variances of populations; Anova = Parametric analysis of variance test for differences in means of populations; Kruskal = Kruskal-Wallis non-parametric test for differences in central values of populations; F = F parametric test to compare variances of two categories; t = Welch two sample parametric t-test for differences in means of two categories; Wilcoxon = Wilcoxon non-parametric test for differences in central values of two categories. All tests carried out in R version 3.5.0 (R Core Team 2018).

Table S3. Variables from our data set tabulated following Table IV of Mitchell & Mori (1987: 40), which compares *Anacardium occidentale* ecotypes. Dune populations are highlighted in beige and the cerrado region population in green.