Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Does size matter? Morphological and genetic similarities between cashew (Anacardium occidentale) and cajuí (A. humile)

Acalene Gonçalveis-Oliveira[1]
Federal University of Pernambuco
Thiago Nascimento[1]
Federal University of Pernambuco
Priscila Alves Barroso[2]
Federal University of Piauí
Cíntia Luíza Silva Luz[3]
State University of Campinas
Silvokleio Costa Silva[4]
Federal University of Piauí
Andrea Pedrosa-Harand[1]
andrea.harand@ufpe.br
Federal University of Pernambuco

Research Article

Keywords: Anacardiaceae, CMA/DAPI banding, Flow cytometry, Fluorescent in situ hybridization - FISH, morphological characterization, phylogeny

Posted Date: February 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8758033/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Cashew trees are recognized for their economic and ecological importance in the tropics, but their species delimitations and phylogenetic relationships remain unclear. Here, we analyzed the genetic diversity of their germplasms from the Cerrado and restinga biomes in Northeastern Brazil. Twenty-four accessions of *A. occidentale* L. and two of *A. humile* A.St.-Hil. were evaluated using morphological and genetic markers. The morphological variability extended beyond species boundaries, challenging their differentiation. Furthermore, molecular analyses of selected nuclear (*ITS*) and plastid (*matK*, *trnLF*, *ycf1* and *rps16*) DNA regions revealed no clear genetic separation between *A. occidentale* and *A. humile*, resulting in an unresolved clade. We also performed cytogenetic analyses and genome size measurements on twelve accessions of *A. occidentale* along with two of *A. humile* looking for possible undetected genomic differentiation. The results indicated a strong chromosomal stability, with $2n = 40$, $CMA^+/DAPI^-$ bands in the terminal region of the short arm of three chromosome pairs, one pair of 5S rDNA, and three pairs of 35S rDNA sites co-localized with $CMA^+/DAPI^-$ bands. Genome sizes were similar among the accessions, with an average value of 0.44 pg/1C (435 Mbp), suggesting no significant intra- or interspecific variation and no evidence of polyploidy. The lack of differentiation among the accessions may be attributed to incomplete lineage sorting, hybridization with introgression, or the possibility that they represent a single species, with domesticated accessions exhibiting larger fruits and pseudofruits.

Introduction

Sustainable management of plant resources is intrinsically linked to the preservation of their genetic diversity. One of the key steps toward this goal is identifying populations or genotypes that represent the existing variability within a species, using different methodologies. These approaches also enable the identification of accessions with potential for use in breeding programs (Borém et al. 2021; Salgotra and Chauhan 2023). In this context, a widely employed strategy to obtain relevant information to support the conservation of genetic diversity is the characterization of germplasm through the analysis of morphological traits, mainly due to the faster data acquisition compared to molecular techniques (Zuffo et al. 2016; Carneiro et al. 2019; França et al. 2020).

However, the phenotypic variability detected through morphological characterization may not accurately reflect underlying genetic diversity, as morphological traits are often influenced by environmental conditions and exhibit phenotypic plasticity (Govindaraj et al. 2015). In contrast, genetic diversity of different populations can be more directly assessed by molecular tools, such as nuclear markers - example.g., the *Internal Transcribed Spacer* (*ITS*) region - and/or plastid regions, such as *matK*, *rps16*, *rbcL*, and *ycf1* (Dong et al. 2015; Turchetto-Zolet et al. 2017). In addition to morphological and molecular approaches, cytogenetic analyses also provide valuable information for species characterization, providing information related to their complete karyotype (Costa et al. 2017).

The Anacardiaceae family comprises around 81 genera and approximately 800 species, with a primarily tropical and subtropical distribution. The genus *Anacardium* L. includes ten accepted species, with *Anacardium excelsum* (Bertero & Balb. ex Kunth) Skeels being the only species not found in Brazil (Mitchell et al. 2022; Powo 2024). Among these, *Anacardium occidentale* L. (cashew tree), a tree ranging from 1.5 to 15 meters in height, is one of the most economically important species in this family, along with *Mangifera indica* L. (mango) and *Pistacia vera* (pistachio). Its fruits (cashews nut) and pseudofruits (peduncles) are widely exploited by the agribusiness (Carneiro et al. 2019). Meanwhile, *Anacardium othonianum* Rizzini is currently considered a synonym of *A. occidentale* due to morphological similarities (Silva-Júnior et al. 2021). On the other hand, *Anacardium humile* A.St.-Hil. is a shrubby species ranging from 0.3 to 1.5 meters in height. It is endemic to Brazil and produces edible fruits and pseudofruits that are also exploited, which are harvested exclusively through extractivism due to their generally smaller sizes. The economic use of this fruit species is limited to local consumption, primarily by rural communities in the Northeast of Brazil, particularly in areas of the Cerrado biome where both species co-occur (Crespo and Souza 2014).

Among the species of *Anacardium*, only *A. occidentale* has undergone formal genetic improvement, aimed at developing clones with produce high-quality peduncles and nuts (Garruti et al. 2022). Although no breeding programs have been formally established for *A. humile*, some pre-breeding studies have already been conducted (Borges et al. 2018; Dos Santos et al. 2019; Pereira et al. 2019). Molecular phylogenetic studies within the genus are still limited to a few taxa (Rabah et al. 2017), hampering a understanding of the phylogenetic relationships among *Anacardium* species. On the other hand, in other genera of the *Anacardiaceae* family, such as *Rhus* L., *Pistacia*, *Schinus*, and *Spondias*, the use of molecular markers (*ITS*, *matK*, *rps16*, *rpl16*, *ndhF*, *trnL_trnF*, *psaA_ycf3*, *atpB_rbcL*, and *psbA_trnH*) has allowed the detection of polymorphisms and contributed to the understanding of phylogenetic relationships. These studies have produced well-supported clades, resolved monophyletic groups, and identified hybridization events (Yi et al. 2004; Xie et al. 2014; Nobre et al. 2018; Silva-Luz et al. 2019; Ariyarathne et al. 2020). Taxonomic classification within *Anacardium* remains controversial. For example, morphometric analyses by Vieira et al. (2014) were unable to clearly distinguish species such as *A. microcarpum* and *A. occidentale* due to substantial overlap in leaf traits. In turn, cytogenetic analyses are restricted to *A. occidentale*, with most reports indicating a chromosome number of $2n = 40$ (Gill and Singhal 1979; Gill et al. 1990; Pedrosa et al. 1999). However, other chromosome counts $2n = 30$ and 42 have also been reported (Machado 1944; Darlington and Janaki-Ammaal 1945; Aliyu and Awopetu 2007), suggesting possible intra-specific numerical variation. Genome size estimates are also available only for *A. occidentale*, averaging 0.85 pg/2C, with no evidence of polyploidy (Aliyu 2014).

The present study aimed to test the following hypotheses: (I) *Anacardium occidentale* and *A. humile* represent distinct evolutionary lineages within the genus; (II) morphological variability observed in *A. occidentale* and *A. humile* is correlated with their molecular differentiation; (III) karyotypic variation exists among accessions of *A. occidentale* and *A. humile*, and among other *Anacardium* species, which can be effectively used for germplasm characterization. To address these questions, we investigated the genetic diversity and phylogenetic relationships of *A. occidentale* and *A. humile* with other *Anacardium* species, using morphological, molecular, and cytogenetic tools to better understand the evolutionary history and diversification of the genus.

Materials and Methods

Plant Material and Taxonomic Delimitation

Cashew nuts from 37 accessions were collected from natural populations of *Anacardium occidentale* (cashew) and *A. humile* (cajuí) (Table S1), with most samples originating from six different locations in the state of Piauí, Brazil. These seeds were subsequently deposited in the Active Germplasm Bank of *Anacardium* Seeds at the Professora Cinobelina Elvas Campus of the Federal University of Piauí. All the accessions were taxonomically evaluated. The majority were classified as *A. occidentale*, although many did not conform to the typical size descriptions for fruits and pseudofruits. A second group of accessions was taxonomically identified as *A. humile*. However, vegetative and reproductive characteristics have been revealed as inefficient characters for taxa identification withing this group of plants, mostly due to considerable morphological variation, primarily displayed by the leaves. Furthermore, flower and fruit characteristics show considerable overlap between the two species, with the exception of longer peduncle found in some *A. occidentale* accessions, which may be result from the domestication process (Mitchell and Mori 1987). Despite this overlap, *A. occidentale* is usually easily distinguishable by its tree-like form, whereas *A. humile* is a subshrub characterized by a prominent underground trunk and ascending branches.

In addition, seeds of *A. occidentale* (BRS 226) and specimens of *A. occidentale* previously identified as *A. othonianum* (BGC 45, BGC 45.2, BGC 45.4 and BGC 46) were provided by Embrapa Agroindústria Tropical, located in Fortaleza, Ceará, Brazil. We also obtained seeds of *A. humile* from the central region of the Brazilian Cerrado. The germinated seedlings were cultivated in the Experimental Garden of the Laboratory of Cytogenetics and Plant Evolution, in Recife, Pernambuco, Brazil. Vouchers of the collected specimens were deposited in the UFP Herbarium at the Federal University of Pernambuco, Brazil (Table S1).

Morphoagronomic Characterization

For agronomic evaluations, 15 mature fruits (nuts) and pseudofruits (peduncles) were collected from each of the 26 accessions/mother trees (Table S1) and characterized agronomically. The *A. occidentale* accessions previously identified as *A. othonianum* provided by Embrapa, as well as *A. humile* from the central Brazilian Cerrado, were were excluded from this characterization, due to the limited number of seeds available. The characterization of the accessions, descriptive traits for fruits and peduncles (cashew apples or pseudofruits) established for cashew by the *International Board for Plant Genetic Resources* (IBPGR 1986) were used as the cashew apple shape (CAS); shape of cashew apple base (SCAB); ridges on cashew apple (RCA); cashew apple apex (CAA); grooves on apex cashew apple (GACA); cavity at apex cashew apple (CACA); skin of cashew apple (SCA); length cashew apple (LeCA); width cashew apple (WiCA); diameter cashew apple (DiCA); weight cashew apple (WeCA); relative position of suture and apex (RPSA); suture of nut (SN); shape of nut apex (SNA); shape of nut base (SNB); nut shape (NS); flanks of nut (FN); stylar scar on nut (SSN); nut length (NL); nut width (NW); nut thickness (NT); and weight of nut (WeN). Measurements (in mm) were taken using a digital caliper with a 150 mm range and ± 0.02 mm / 100 accuracy. Weights (in g) were obtained using an analytical scale balance (Model AY220 - Tecnal).

Fruit and peduncle color were measured with a portable spectrophotometer (model CbM-700D, Konica Minolta). Using the coordinates (*L*, *a*, and *b*) provided by the instrument, luminosity coordinates were measured on opposite sides of the nuts and peduncles: *L* [ranging from white (*+L*) to black (*-L*)], *a* [from red (*+a*) to green (*-a*)], and *b* [from yellow (*+b*) to blue (*-b*)]. Color measurement was performed in the visible spectrum (400 to 700 nm) for all three sections.

The total soluble solids content (TSS), expressed in °Brix, was determined from the juice extracted from each of the replicates (peduncle) per accession. After homogenization, two drops of juice from each replicate were placed on the prism of a refractometer (Model GT808 - ATC) to measure the °Brix value of each sample.

Cluster Analysis and Correlations

After performing all measurements, the quantitative morphological data were subjected to analysis of variance (ANOVA) to assess differences among accessions. The analysis was conducted using R software version 4.0.5 (R Core Team 2022) with ExpDes package (Ferreira et al. 2021). A completely randomized design was adopted, with 15 replicates (fruits and pseudofruits) per accession. To group the means into statistically homogeneous clusters, the Scott-Knott test was applied (Scott and Knott 1974).

To evaluate the multivariate effects of the accessions, a Multivariate Analysis of Variance (MANOVA) was performed. Upon confirming significant multivariate effects, Pillai's trace test was used to verify the significance level. Accessions were then grouped using the Unweighted Arithmetic Means Method (UPGMA), based on the generalized Mahalanobis distance matrix, derived from the quantitative traits. The consistency of the clusters was evaluated using the cophenetic correlation coefficient (Sokal and Rohlf 1962), and the Mojena method (1977) was employed to establish the cutoff point for defining cluster boundaries. To determine the relative contribution of each trait to overall genetic divergence, Singh' criterion (1981) was applied. The analyses were performed with the aid of the Biotools package (Da Silva 2021).

Additionally, Principal Component Analysis (PCA), was conducted using the covariance matrix of the original variables, obtaining from it the eigenvalues that constructed the eigenvectors. The studied accessions were plotted on a biplot for the first two principal components with the support of the factoextra package (Kassambara and Mundt 2020).

Cytogenetic Characterization

Young root tips from germinated seeds or potted plants grown in the Experimental Garden of the Laboratory of Plant Cytogenetics and Evolution were previously pre-treated with 2 mM 8-hydroxyquinoline for 7 hours at 18 °C and then fixed in methanol/acetic acid (3:1, v/v) for 2-24 h at room temperature and stored at -20 °C until further analysis. Each meristem was washed twice in distilled water and digested in enzymatic solution containing 2% cellulase (Onozuka®), 20% pectinase (Sigma®), and 2% pectolyase (Sigma®) for 3 hours at 37 °C. Subsequently, slides were prepared following a modified version of the air-drying technique described by De Carvalho and Saraiva (1993). In this modification, prepared slides were immersed in 60% acetic acid for 3 hours to reduce the excess cytoplasmatic content. After drying, the slides were stained with 4',6-diamidino-2-phenylindole (DAPI, 2 µg/mL) in glycerol (1:1, v/v) for selection of the best slides. They were then destained in ethanol: acetic acid (3:1, v/v) for 30 minutes at room temperature, transferred to absolute ethanol for at least 1 hour, air-dried, and aged for three days.

Double staining with the fluorochromes Chromomycin A_3 (CMA) and DAPI was performed as described by Vaio et al. (2018) with some modifications. After aged, the slides were stained with CMA (0.5 mg/mL) for 2 hours in a humid and dark chamber at room temperature, then counterstained with DAPI (1 µg/mL) in glycerol/McIlvaine buffer pH 7.0 (1:1, v/v) containing 2.5 mM $MgCl_2$ and stored for three days in the dark at room temperature. Images were captured using a Leica DM5500B epifluorescence microscope equipped with a Leica DFC345FX camera through the LAS AF software. At least 10 metaphases per accession were analyzed, and the best images were uniformly adjusted for brightness and contrast using Adobe Photoshop (v.21.0.2). The slides were subsequently destained, as described above, for later use in Fluorescence *in situ* hybridization (FISH).

FISH experiments followed the protocol Pedrosa et al. (2002). For the localization of 5S rDNA sites, a pool of pre-labelled oligonucleotide probes (PLOP, 5SrDNA_ang_1-4) conjugated with Cy3 was used (Waminal et al. 2018). A fragment of 25-5.8S-18S rDNA (35S rDNA, clone *pTa71*) from *Triticum aestivum* L. (Gerlach and Bedbrook 1979) was amplified by mini-prep and labelled with Alexa Fluor 488-dUTP (Invitrogen®) by nick translation. The hybridization mixture was composed of 50% (v/v) formamide, 10% (w/v) dextran sulfate, 2× SSC, 8 ng/µL for the plasmid probe and 25 ng/µL for the oligonucleotide probe. Chromosomes were denatured at 80°C for 10 minutes and incubated overnight in a pre-warmed humid chamber at 37 °C. Stringency washes (~ 76%) were carried out with two washes in 2× SSC, then two washes in 0.1× SSC, both at 42 °C and for 5 minutes each, after this another wash in 2× SSC was performed at room temperature for 10 min. All slides were counterstained with 1 µg/mL DAPI in mounting medium and analyzed as described above.

Flow Cytometry

The nuclear DNA content was estimated by flow cytometry for ten accessions of *A. occidentale* from different locations and one accession of *A. humile* (Table 1). Nuclear suspensions were prepared using the protocol of Aliyu (2012), with the following adaptations: leaf tissues from the internal standard *Solanum lycopersicum* L., var. Stupické polnírané, with a genome size of 1.96 pg/2C (Doležel et al. 1992), were simultaneously macerated with the sample in a Petri dish containing approximately 750 µL of *Woody Plant Buffer* (WPB) isolation buffer (Loureiro et al. 2007). The nuclear suspension was filtered through a 50 µm mesh, then 30 µL of propidium iodide (1 mg/mL) was added to stain the nuclei in suspension. Measurements were performed using a PARTEC Cyflow Space flow cytometer (Münster, Germany), each accession sample was measured three times on three different days. For each measurement, approximately 5,000 nuclei were quantified, and the results were interpreted by analyzing the graphs generated by FloMax software v. 2.3. Genome sizes (pg) were estimated for each accession using the formula [(mean sample fluorescence/mean standard fluorescence) × standard genome size] (Doležel 2005). The three accessions with lowest coefficient of variation (CV) were used to calculate the average 1C content of *A. occidentale*. Diferences in average genome size among the accessions of *A. occidentale* and *A. humile* was evaluated using Analysis of Variance (ANOVA), considering significance at $P < 0.05$.

Sampling and DNA Extraction

Leaf samples from the 26 individuals of *Anacardium* (Table S3), representing the agronomic variability of the species, were used for DNA extraction. Genomic DNA (gDNA) was extracted from 50-200 mg of fresh leaves; however, for some accessions, 50 mg of silica-dried leaves

were used when fresh material was unavailable. DNA extraction followed the method described by Ferreira and Grattapaglia (1998), and DNA quantity and purity were assessed using a NanoDrop 2000 spectrophotometer (Thermo Scientific).

PCR Amplification and Sequencing

From the 26 *Anacardium* individuals with extracted DNA, three plastid regions (*matK*, *trnL-F* and *rps16*) were successfully amplified for 17 accessions using universal primers previously described (Taberlet et al. 1991; Sang et al. 1997; Schäferhoff et al. 2010). These regions were available in GenBank for various *Anacardium* accessions and were used for a broad phylogenetic analyses of the family Anacardiaceae (Xie et al. 2014; Silva-Luz et al. 2019; Ariyarathne et al. 2020). For the nuclear locus (*ITS*), amplification was successful for only 17 accessions, using the 17SE and ITS4 primers described by White et al. (1990) and Sun et al. (1994). PCR reactions were performed in a total volume of 50 µL, containing: 20-100 ng of gDNA, 1× PCR buffer, 1× TBT [1g/L bovine serum albumin; 8.5 mM Tris-HCL (pH 8.0); 1% (v/v) Tween-20 and 750 mM Trehalose], 0.2 mM dNTPs, 3 mM $MgCl_2$, 0.1 µM of each primer, and 0.2 µL Taq polymerase (Thermofisher®). Amplifications were carried out using the PCR programs described in Table S2. PCR products were visualized on a 1% agarose gel. Successfully amplified products were purified using 75% isopropanol precipitation, quantified, and sent for sequencing on an ABI 3500 sequencer (Applied Biosystems®) at the Sequencing Platform of the Bioscience Centre at the Federal University of Pernambuco. Sequences were edited and aligned using the alignment tool in Geneious software version 7.1.4 (Kearse et al. 2012). All sequences processed in this study were deposited in GenBank (PV089646-PV089661, PV089703, PRJNA1224782).

Search for additional polymorphic regions in cashew plastomes

We also investigated the most polymorphic plastid regions from *A. occidentale* plastome sequences. For this, we mapped Illumina reads from different *A. occidentale* accessions available in GenBank (accession numbers SRX2990990) to its reference plastome (NC_035235) (Table S3) to assess the intra-specific polymorphism across several *loci* and identify the most informative for our phylogenetic analysis. This analysis revealed that the *ycf1* region was the most polymorphic. Consequently, out of the 26 accessions with extracted DNA, we successfully amplified *ycf1* in 24 accessions using primers described by Dong et al. (2015), following the amplification protocol outlined in Table S2. All PCR products were subsequently sequenced as described in the previous section.

Phylogenetic analyses

Sequences from five regions were obtained for individuals of *A. humile*, *A. occidentale*, along with sequences available in GenBank for *A. occidentale*, *A. excelsum* (Bertero & Balb. ex Kunth) Skeels, *A. parvifolium* Ducke, *A. spruceanum* Benth. ex Engl., *Fegimanra africana* (Oliv.) Pierre, *Mangifera indica* L., and *Spondias mombin* L. were included, with *S. mombin* indicated as the outgroup to root the tree. Phylogenetic relationships were inferred using Bayesian Inference (BI) in MrBayes v.3.2.6 (Ronquist et al. 2012). Analyses were first performed separately for each region, after which a concatenated alignment was generated for the plastid regions only. All analyses employed the General Time Reversible substitution model (GTR) with a gamma model of rate heterogeneity (Abadi et al. 2019). Four independent runs, each with four Markov Chain Monte Carlo (MCMC) chains were conducted, sampling every 1,000 generations for a total of 10,000,000 generations. Plastid, nuclear, and consensus trees, constructed using majority rule method and posterior probability (PP), were visualized and edited in FigTree v.1.4.2 (Rambaut et al. 2014).

Diversity, Differentiation, and Genetic Structure

Diversity, differentiation, and genetic structure were analyzed based on the obtained sequencing data. The nuclear (*ITS*) and concatenated plastidial (*matK*, *ycf1, rps16*, and *trnL_trnF*) sequences were edited and aligned using the alignment tool in Geneious software version 7.1.4 (Kearse et al. 2012). The resulting alignments were then used as input in the DNA Sequence Polymorphism (DnaSP v5.0) program (Librado and Rozas 2009) to determine the number of haplotypes (h) and haplotypic diversity (Hd). In addition, nucleotide diversity rates (π), population differentiation through pairwise F_{ST} (where values close to 0 indicate high genetic similarity and values close to 1 suggest greater dissimilarity), and molecular variance analysis (AMOVA) were calculated for both plastidial and nuclear sequence data using ARLEQUIN software v3.5.2 (Excoffier et al. 1992; Excoffier and Lischer 2010). The statistical significance of the AMOVA was tested using 10,000 permutations (Shepherd et al. 2016). Furthermore, two haplotype networks were constructed using the Median-Joining network method, implemented in NETWORK software (Bandelt et al. 1999).

Results

Morphoagronomic Characterization of Fruits and Peduncles of *A. occidentale and A. humile*

The morphoagronomic analysis of *A. occidentale* and *A. humile* accessions revealed qualitative and quantitative variation in fruit and peduncle traits. Among the evaluated accessions, only AN-708 exhibited a pear-shaped form (CAS), while others were cylindrical, rounded, or conical to obovate. Regarding the shape of cashew apple base (SCAB), only AN-701 and BRS 226 were classified as flattened, while the remaining

accessions presented angular, obliquely flattened, or rounded bases. They also varied in terms of ridges on cashew apple (RCA), grooves on apex cashew apple (GACA), and cavity at apex cashew apple (CACA). Except for AN-803, which presented an oblique cashew apple apex (CAA), all evaluated specimens had apices at the same level. Similarly, only AN-718 exhibited rough and opaque skin of cashew apple (SCA), while others displayed smooth and shiny skin (Table S4, Fig. 1).

Based on quantitative data, cluster analysis and divergence among *Anacardium occidentale* and *A. humile* accessions from different locations indicated the existence of genetic diversity (Table S5). The weight of the fresh cashew apple (WeCA) was the most discriminating trait. Group "a" consisted exclusively of commercial cashew accession BRS 226, with a WeCA of 104.02 g. Groups "b" and "c" included accessions from Parnaíba, with average WeCA values ranging from 29.07 g (AN-800) to 34.06 g (AN-802). Group "f," had WeCA values between 3.31 g (AN-716) to 6.92 g (AN-702), and comprised thirteen accessions from different locations and included the two *A. humile* accessions (AN-710 and AN-711). Evaluating of length cashew apple (LeCA), width (WiCA), and diameter (DiCA), it was clear that accession BRS 226, the only one belonging to group "a," had the highest values for these descriptors, while AN-716, also from *A. occidentale*, had the lowest. No geographic trend was detected for these size-related traits (Tables S1 and S5).

For Total Soluble Solids (TSS), accessions were grouped into five clusters (Table S5). Group "a" contained only BRS 226, which had the lowest °Brix value (11.1%). The highest TSS was recorded for AN-716, (20.0%), indicating a high sugar content, though this accession did not stand out in terms of weight or length, essential attributes for product acceptance and commercialization. Colorimetric analysis distinguished three colour patterns, ranging from yellow to red among cashew apples (Table S5). Luminosity ("*L*") and color parameters ("*a*" and "*b*") from spectrophotometer grouped the accessions into three distinct clusters (Table S5).

Qualitative analysis of cashew nut showed that the relative position of suture and apex (RPSA) was highly conserved, except for BRS 226 and AN-802, where the suture is in front of the nut apex (Table S6, Fig. 1). The shape of nut apex (SNA), ranged from pointed (only AN-706) to intermediate or rounded, while shape of nut base (SNB) varied between flat, obliquely flat (only AN-706), and rounded. Suture of nut (SN) were either angular or rounded, while nut shape (NS) ranged from oblong ellipsoid to kidney-shaped. The flanks of nut (FN) were bulging, flat, or rounded and the stylar scar on nut (SSN) was classified as either large or small (Table S6).

Quantitative nuts data also showed considerable variation (Table S7). The nut length (NL) divided accessions into eight groups, ranging from 1.43 cm (AN-707) to 3.63 cm (BRS 226). For nut width (NW) and nut thickness (NT), nine groups were identified: the smallest averages were from AN-715 (NW = 1.18 cm; NT = 0.8 cm), while the largest were BRS 226 (NW = 2.99 cm; NT = 2.22 cm). Weight of nut (WeN) grouped accessions into six categories, from 1.03 g (AN-718) to 10.93 g (BRS 226). Although accession BRS 226 had the highest nut weight, no consistent correlation was observed between nut length, width, and thickness among all studied accessions. Colorimetric analysis revealed predominantly gray nuts,sometimes with a greenish hue (Table S7).

Considering all variables together, two distinct groups and subgroups were identified (Fig. 2**A**). The cophenetic correlation coefficient (CCC) was r = 0.9580, indicating a strong fit between the dendrogram and the distance matrix. The first group consisted exclusively of *A. occidentale* BRS 226, while the second comprises by matrices of *A. occidentale* and *A. humile*, subdivided into three subgroups: (I) AN-700 and AN-704 (Bom Jesus); (II) accessions from Bom Jesus, Currais, Alvorada do Gurguéia, and Buriti dos Lopes; (III) accessions from Parnaíba. This structure reflects a clear geographic separation between Parnaíba and other populations.

The variables contributing most to cluster formation were the variables cashew apple length (LeCA), nut weight (WeN), nut thickness (NT), and nut length (NL) which together accounted for 54.3% of the observed divergence (Fig. 2**B**). Principal Component Analysis (PCA) explained 73.80% of the total variation (Fig. 2**C**) and its results were consistent with the dendrogram (Figs. 2**A** and 2**C**). The BRS 226 accession appeared as the most contrasting among the analyzed genotypes. The traits that contributed to this divergence were: cashew apple length (LeCA), cashew apple width (WiCA), cashew apple diameter (DiCA), cashew apple weight (WeCA), nut length (NL), nut width (NW), nut thickness (NT), and nut weight (WeN).

Chromosomal Numbers, CMA/DAPI Bands, rDNA Sites, and DNA Content

The cytogenetic analysis was conducted on 12 accessions of *A. occidentale* and two accessions of *A. humile*, revelling a stability in chromosome number, with a karyotype of $2n = 40$ for all accessions (Fig. 3, Fig. 4). This study report the first chromosome counts for *A. humile*. The CMA/DAPI staining analysis also indicated stability in the number of GC-rich heterochromatic bands (CMA^+) and the absence of AT-rich bands ($DAPI^+$). In all analyzed karyotypes, $CMA^+/DAPI^-$ bands were located specially in the terminal region of the short arm of three chromosome pairs, with one of these pairs being smaller and showing weaker staining intensity in the heterochromatic block (Fig. 3, Fig. 4). Fluorescence *in situ* hybridization analyses revealed one pair of subterminal 5S rDNA sites and three pairs of terminal 35S rDNA sites, one being smaller and all sites co-localizing with CMA^+ bands (Fig. 5).

The estimated genome sizes for *A. occidentale* and *A. humile* were relatively small, with a mean value of 0.88 pg/2C. Based on the three accessions with the lowest coefficient of variation (CV), the genome size was estimated at ~ 435 Mb/1C for both species (Table 1, Figure S1).

The analysis of variance (ANOVA) for *A. occidentale* and *A. humile* revealed a *p*-value of 0.304 ($p > 0.05$), suggesting no significant intraspecific variation in genome size among the evaluated accessions (Figure S2).

Table 1 Average genome size of ten accessions of *A. occidentale* and one accession of *A. humile*. The three *A. occidentale* accessions with the lowest CV (Coefficient of Variation) are indicated in bold

Species	Germplasm ID	2 *n*	Mean pg/2C	CV (%)	Mean Mbp/1C
A. occidentale	**AN-700**	40	**0.88**	**4.71**	**435**
A. occidentale	AN-701	40	0.88	5.79	431
A. occidentale	AN-704	40	0.93	7.55	456
A. occidentale	AN-705	40	0.92	6.35	450
A. occidentale	AN-706	40	0.89	5.50	439
A. occidentale	**AN-709**	40	**0.89**	**4.94**	**435**
A. occidentale	**AN-717**	40	**0.89**	**4.72**	**435**
A. occidentale	AN-719	40	0.88	5.04	430
A. occidentale	AN-801	40	0.91	6.87	446
A. occidentale	BGC-45	40	0.92	5.54	449
Mean			0.88	4.79	435
A. humile	AN-712	40	0.88	4.47	434

Phylogenetic Relationships in Anacardium

In this study, we generated 92 new sequences derived from five regions: 17 from *ITS*, 24 from *ycf1*, 17 from *matK*, *trnL_trnF*, and *rps16* each. Additionally, 24 sequences available in the GenBank database were incorporated into the data matrix (Table S2). All regions analyzed exhibited a high degree of conservation, with low genetic polymorphism, particularly in the plastid regions. Table 2 summarizes the characteristics of the plastid and nuclear regions used.

For some plastid regions (*matK*, *trnL-F*, and *rps16*) available in GenBank for several *Anacardium* accessions, concatenated phylogenetic analysis placed *Fegimanra africana* as the sister genus to *Anacardium*, which was recovered as a monophyletic group with high posterior probability (PP = 1; Figure S3). *Anacardium excelsum* was sister to all other *Anacardium* accessions (PP = 1). Additionally, *A. humile* and *A. occidentale* also formed a clade, but the relationships between these species were not well supported.

To further investigate intraspecific variation, we conducted an intraspecific plastome search to identify additional polymorphic regions among *A. occidentale* accessions. In this analysis identified the *ycf1* gene as the most polymorphic *locus*. The phylogenetic tree based on the *ycf1* marker, generated from 24 accessions, showed lower posterior probabilities than the tree obtained from the three concatenated plastid regions (Figure S4), indicating that ycf1 alone was insufficient to resolve phylogenetic relationships. Therefore, we concatenated the *ycf1* region with the three other plastid regions. To avoid missing data, eight individuals from the 24 *ycf1* accessions were excluded because they were not amplified for the three remaining plastid regions. The resulting concatenated plastid tree revealed five well-supported clades, consistent with the topology of the tree from of the three concatenated plastid regions (Figure S5). However, the relationships among *A. occidentale* and *A. humile* remained unresolved.

The ITS1-5.8S-ITS2 region also supported the monophyly of *Anacardium*, although with support (PP = 0.64; Figure S7). As in the plastid analysis, *A. excelsum* was recovered as sister to the remaining *Anacardium* accessions, with only three clades showed strong support (PP = 1). To improve resolution, we combined the plastid and nuclear datasets for Bayesian analysis (Fig. 6), which allowed to observe the monophyletic nature of the genus with a strong support (PP = 1), with *Fegimanra africana* as sister. *A. excelsum* appeared as the first species to diverge within the genus. Another well-supported clade separated *A. occidentale* and *A. humile* from *A. spruceanum* and *A. parvifolium*, each represented by a single accession. We also observed a well-supported clade containing a single *A. occidentale* accession distinct from all others, which were further divided into three well-supported subclades.

Accessions of *A. humile* were taxonomically identified based on morphological traits, particularly their shrubby growth habit. However, *A. occidentale* accessions exhibited an arboreal habit but were initially collected as *cajuís* (*A. humile*), because of their small fruits and pseudofruits, which closely resembled those of *A. humile*. In fact, most *A. occidentale* and *A. humile* accessions had fruits and pseudofruits with length and width measurements within the standards established for *A. humile* by Mitchell and Mori (1987). Our results indicate that many

A. occidentale accessions display intermediate morphology, with fruits and peduncles similar to *A. humile* and an arboreal habit characteristic of *A. occidentale*. To test whether these *A. occidentale* accessions might be hybrids between the two species, we removed all potential *A. occidentale* "hybrids" from both the plastid (Figure S6) and nuclear (Figure S8) datasets, as the presence of hybrids can alter tree topology and reduce support. Even after excluding the putative hybrids, plastid analysis still failed to clearly separate *A. humile* and *A. occidentale*, suggesting they may share the same plastid haplotypes. Conversely, the *ITS* tree supported a clade of *A. occidentale* (PP = 1) distinct from *A. humile*, although relationships between the two species remained unresolved (Figure S8).

Table 2 Polymorphic nuclear and plastid DNA regions used for phylogenetic analysis in *Anacardium* L.

Loci	Number of individuals	Alignment length (bp)	Informative sites (ingroup)	Variable sites (ingroup)	Conserved sites (ingroup)	Pairwise identity	GC Content
ITS	17	624	28	96	491	90.9%	65.1%
ycf1	24	872	23	66	797	95.7%	30.3%
rps16	17	819	1	8	800	96.1%	33.9%
trnL-F	17	787	2	20	769	96.3%	37.1%
matK	17	765	4	42	708	94.2%	35.1%

Diversity, Differentiation and Genetic Structure

To investigate the relationship between *Anacardium occidentale* and *A. humile* in greater detail, we analyzed genetic diversity, differentiation, and population structure using nuclear and plastid sequence data. For the nuclear dataset, the highest nucleotide diversity index (π) was observed in the *A. occidentale*-Parnaíba and *A. occidentale*-Alvorada populations, indicating considerably greater nucleotide diversity, while the *A. occidentale*-Currais population showed the lowest diversity (Table S8). In the plastid dataset, the group composed of cultivated *A. occidentale* accessions (BRS 226, BGC-45, and BGC-46) presented the highest nucleotide diversity index, whereas the *A. occidentale*-Bom Jesus population recorded the lowest, except in populations with very small samples sizes (π = 0.0) (Table S8).

Population differentiation analyses based on nuclear and plastidial data (Tables S9 and S10), showed no significant genetic differences between the *A. occidentale* and *A. humile* populations. Although some populations of *A. occidentale* and *A. humile* showed F_{ST} values close or equal to 1, these results were not statistically significant, as the sample sizes of each population was small. This factor reduces the significance of the F_{ST} index, making the results less representative of genetic differentiation between the analyzed populations (Tables S11 and S12). Additionally, negative F_{ST} values were recorded, indicating minimal or no genetic differentiation between groups.

The AMOVA for the nuclear data revealed that most variation occured within populations (96.67%), with only 9.96% attributed to differences between populations. Variation between the two species (*A. occidentale* and *A. humile*) was negative (-6.63), suggesting no detectable genetic differentiation (Table S13). For plastid data, most variation (86.24%) also occurred within populations, while variation between populations was higher than observed with nuclear data but represented only 13.82% of the total variation. As with the nuclear dataset, no significant genetic differentiation was observed between groups/species (Table S14).

Phylogeographic analyses of haplotype distributions and networks (Figures S9 and S10), showed patterns consistent with the diversity and genetic structure results. For the nuclear dataset, we identified 16 haplotypes (h = 16), with a different haplotype per individual, except for two individuals with the same haplotype (H3). All nuclear haplotypes belonged to *A. occidentale* except H13, which belonged to *A. humile*, resulting in a haplotypic diversity was Hd = 0.9926. For the plastidial dataset, we also identified 16 haplotypes, with haplotypes H10, H11, and H12 belonging to *A. humile*, and a haplotypic diversity of Hd = 1. The haplotype networks did not separate *A. occidentale* and *A. humile*, similarly to the phylogenetic results, which did not reveal species-specific clades. Nevertheless, the plastidial network showed greater similarities among geographically related accessions. This reinforces the hypothesis of possible hybridization, recent evolution, or the absence of genetic barriers between *A. occidentale* and *A. humile* populations.

Discussion

Morphoagronomic Diversity of the Pseudofruits and Fruits of A. occidentale and A. humile

The morphoagronomic characterization of the accessions revealed variation in most qualitative and quantitative traits of fruits and pseudofruits, demonstrating clear heterogeneity among the accessions. Morphological variation in the common cashew tree has been reported in previous studies, such as Asna et al. (2021), who described conical-ovoid pseudofruits in *A. occidentale* accessions, and França et al. (2020),

who identified a predominance of pyriform pseudofruits of *Anacardium* sp. When this phenotypic plasticity was evaluated across different species such as *A. occidentale*, *A. othonianum*, and *A. humile*, the most prevalent shapes described were the conical-ovoid, pyriform, and rounded (Rufino 2004; Castro et al. 2011), corroborating what was observed in most samples analyzed here. Variation in qualitative parameters of the nut (cashew), such as the reniform and oblong-ellipsoidal shapes, as reported by Sultana et al. (2022) for *A. occidentale*.

Our principal component analysis, Scott-Knott test, and UPGMA clustering confirmed the remarkable genetic diversity expressed in morphological dissimilarities. Variation for peduncle weight has previously been reported, with average peduncle weights in *Anacardium* species ranging from 0.89 to 54.18 g (Rufino 2004; Rocha et al. 2013; Santos and Santos-Júnior 2015; Pereira 2018), corroborating with our findings. Djolossè et al. (2019) reported an average peduncle weight of 74.12 g for *A. occidentale*, a value slightly lower than that observed here for the BRS 226 accession. For nut weight (WeN), we found clear variation among *A. humile*, from 0.48 to 4.16 g (Rufino 2004; Gomes et al. 2013; Santos and Santos-Júnior 2015; Pereira 2018; Borges et al. 2022). For cashew, Semporé et al. (2021) found nuts with weights ranging from 4.62 to 9.06 g, values compatible with those found in this study. Commercial standards for cashew nuts are typically 7 to 9 g (IBPGR 1986), meaning most nuts evaluated here would be rejected by the international market.

Previous studies reported variations in peduncle length among cashew tree clones at different maturation stages, with values between 57.93 mm and 86.20 mm (Gomes et al. 2006; Lopes et al. 2011). In our results, the apple length (LeCA) for the commercial accession BRS 226 matched values from the literature but was clearly higher than other *A. occidentale* and *A. humile* accessions. In natural *Anacardium* populations, peduncle lengths ranged from 8.00 mm to 43.60 mm (Rufino 2004; Rocha et al. 2013; Santos and Santos-Júnior 2015; Pereira 2018; Borges et al. 2022), consistent with our results. For nuts, the maximum length (NL) of *A. humile* was 16.1 mm here, though previous studies indicate values up to 20.5 mm (Lima et al. 1988; Santos and Santos-Júnior 2015; Pereira 2018). Mitchell and Mori (1987) reported *A. humile* nuts measuring 1.2–2.3 × 1–1.7 cm and peduncles of 1–3 × 1–2 cm, while for *A. occidentale* nuts measured 2–3.5 × 1–2 cm and peduncles 5–20 × 2–8 cm. Uur results show that *A. humile* peduncles are smaller and are in accordance with the measurements reported by Mitchell and Mori (1987). However, most wild accessions identified here as *A. occidentale* had fruit and pseudofruit length and width measurements within the standards established for *A. humile* by Mitchell and Mori (1987). These dimensions also fail to meet agronomic standards of the cashew processing industry, particularly for fruit and peduncle length and width, suggesting that almost all *A. occidentale* accessions studied here exhibit fruit and pseudofruit morphometry more typical of *A. humile*.

Peduncles were also evaluated for Total Soluble Solids (TSS - °Brix). The BRS 226 accession had the lowest value, with 11.1%, whereas other *A. occidentale* accessions ranged from 12.3 and 20.0%. The two *A. humile* accessions presented values of 17.4 and 19.6% (Table S5), in accordance with previous studies which reported values between 5.29% and 21.13% (Rufino 2004; Gomes et al. 2013; Pereira 2018). Most values in this study are mostly higher than those considered by the cashew juice industry, which establishes a TSS value between 10 and 12.22% for cashew (Paiva et al. 2000; Sancho et al. 2007; Oliveira et al. 2019), suggesting potential for use in beverages, ice creams, jams, and other food products, requiring less sugar addition in their preparation. Moreover, the sweet, fleshy, and juicy peduncle attracts a wide variety of dispersal agents, including primates, deer, birds, and especially frugivorous bats, which are the most efficient dispersers in propagating cashew and wild cashew trees, capable of transporting fruits over long distances (Mitchell and Mori 1987; Takehana et al. 2013).

Considering the similarity dendrogram for morphoagronomic traits among the accessions, two main groups were detected: the first associated with the BRS 226 accession and another comprising *A. occidentale* and *A. humile* accessions, with no separation of species. Pereira et al. (2019), when evaluating the genetic diversity of *A. humile*, identified six genetic groups among the 27 genotypes. Here, we observed that although some population structure was linked to geographic location, accessions from different localities also clustered together. This pattern may result from gene flow facilitated by pollinators such as bees, the main pollinators of *Anacardium* flowers, and by dispersal agents such as bats and birds, which can travel long distances. Furthermore, floral biology of *Anacardium* species promotes cross-pollination (Takehana et al. 2013; Hamrick 2012; Borges et al. 2018; Gomes et al. 2021).

Karyotypic Stability in Species of Anacardium

The present study revealed a remarkable karyotypic stability among *Anacardium* species. The chromosome number was consistently $2n = 40$ across all samples, with similar chromosomal morphology. This finding aligns with previous reports for *A. occidentale*, which also indicated $2n = 40$ (Gill and Singhal 1979; Gill et al. 1990; Pedrosa et al. 1999), whereas counts of $2n = 30$ and 42 (Machado 1944; Darlington and Janaki-Ammaal 1945; Aliyu and Awopetu 2007) could not be confirmed. Errors in chromosome counts may occur due to technical difficulties, the high chromosome number, or their small size, which often results in very similar morphology (Figueredo et al. 2016; Hoang et al. 2022). The previously reported $2n = 42$ may result from stretched secondary constrictions, which could have been mistaken for small chromosomes (Guerra et al. 1997; Melo et al. 2011). In such cases, CMA/DAPI banding provides a more precise karyotype characterization (Figueredo et al. 2016). Chromosome number stability is also observed in other woody genera, including *Schinus* (Pedrosa et al. 1999; Da Luz et al. 2015), *Pistacia* (Sola-Campoy et al. 2015; Zerey-Belaskri et al. 2018), *Spondias* (Almeida et al. 2007), *Mangifera* (Yonemori et al. 2010; Pierozzi and Rossetto 2011), *Eucalyptus* (Carvalho et al. 2017), and *Citrus* (Guerra et al. 2020) also exhibit chromosomal number stability, suggesting karyotypic conservation across species.

A highly conserved heterochromatin distribution was also observed in *Anacardium*. $CMA^+/DAPI^-$ bands were terminaly located in three chromosome pairs. In *Schinus*, only one chromosome pair exhibited terminal $CMA^+/DAPI^-$ bands (Las Peña et al. 2006), whereas in *Pistacia vera* L. multiple $DAPI^+$ bands occurred in the proximal and terminal chromosome regions (Sola-Campoy et al. 2015). In *Spondias*, the number and distribution of $CMA^+/DAPI^-$ bands varied among species (Almeida et al. 2007), suggesting a higher degree of structural rearrangements during the evolution. No variation was detected in the number or position of rDNA sites among the *Anacardium* accessions analyzed. A similar pattern was reported for *Mangifera indica*, with one 5S rDNA site and three 35S rDNA site pairs (Yonemori et al. 2010), consistent with the phylogenetic proximity between these two genera. Overall, the number of 5S sites appears constant within Anacardiaceae, while the number of 35S sites is more variable, though always terminally located (Almeida et al. 2007; Sola-Campoy et al. 2015).

The karyotypic stability found in *Anacardium* is also seen in other woody species, such as *Cenostigma* Tul., which shows stability in chromosome number, $CMA^+/DAPI^-$ bands, and in the number of 5S and 35S rDNA sites (Castro et al. 2023). In *Populus* L. and *Citrus*, highly conserved karyotypes have been reported, with no evidence of interchromosomal structural rearrangements, maintaining chromosomal synteny even after 14 million years of divergence in *Populus* and 9 million years in *Citrus* (He et al. 2020; Xin et al. 2020). These observations support the hypothesis that woody species tend to have more stable karyotypes compared to herbaceous ones, as they evolve more slowly, which may require longer periods for the generation and fixation of structural variations. We also observed nuclear DNA content stability, with genome sizes similar to those reported for *A. occidentale* accessions from Africa, with 0.85 pg/2C (419 Mb/1C; Aliyu 2014). Small genomes appear common in Anacardiaceae, with values reported for *Mangifera*, *Pistacia*, *Lannea* L., *Rhus*, and *Toxicodendron* L. ranging from 0.30 to 0.75 pg/1C, and polyploidy (tetraploidy) documented only in some *Mangifera* species (Arumuganathan and Earle 1991; Horjales et al. 2003; Ohri et al. 2004; Bai et al. 2012; Aliyu 2014; Zerey-Belaskri et al. 2018). To date, no polyploidy has been observed in *Anacardium*.

Phylogenetic Relationships and genetic differentiation in Anacardium

Our combined phylogenetic analyses confirmed a low level of polymorphism between the species of *A. occidentale* and *A. humile* evaluated. Both plastidial and nuclear data did not reveal a clear molecular distinction between *the two species*, despite the expectation that they represent distinct taxa. Molecular data have been successfully applied in phylogenetic reconstructions of other genera within Anacardiaceae family. For example, in *Rhus*, *Schinus*, and *Spondias*, phylogenetic studies using ribosomal DNA regions (*ITS* and *ETS*) and plastidial regions (*rps16* and *trnL-F*) have produced well-resolved phylogenies with well-supported clades, enabling clear species separation (Yi et al. 2004; Nobre et al. 2018; Silva-Luz et al. 2019). Ariyarathne et al. (2020) used *ITS* and *matK* regions to analyse endemic Anacardiaceae species from Sri Lanka and observed a close relationship between *A. occidentale* and the genus *Mangifera*, with good support, corroborated by phylogenomic analysis by Savadi et al. (2022). Nevertheless, we confirmed that *Fegimanra* is closer to *Anacardium* than *Mangifera*, as demonstrated by Xie et al. (2014).

We employed NGS data and assembled plastomes to identify the most informative loci for phylogenetic analysis. However, plastidial variation within the selected loci was limited. Rabah et al. (2017) sequenced the complete plastome of *A. occidentale* and the *trnK_matK*, *trnL-F*, and *ndhF* regions for one accession each of *A. nanum* A.St.-Hil., *A. humile*, *A. corymbosum* Barb.Rodr., and *A. excelsum*. Although sampling was limited, they recovered *A. nanum* and *A. humile* as sister species, with *A. occidentale* forming a distinct clade. Notably, *A. excelsum* diverged first and was the only species lacking a plastome insertion, suggesting that the modification likely occurred less than 20 million years ago (Xie et al. 2014), what is congruent with our findings.

Among our *Anacardium* accessions, we observed high morphological variability but low molecular polymorphism, particularly in plastidial regions, which proved less informative phylogenetically. Most molecular variation occurred among individuals within the same population, with no clear differentiation between species. Previous investigations on *Anacardium* genetic diversity and structure, including species such as *A. microcarpum*, *A. othonianum*, *A. occidentale*, *A. giganteum* W. Hancock ex Engl., and *A. excelsum*, using ISSR and SSR markers revealed low to high genetic diversity but low levels of genetic structure, suggesting substantial gene flow (Bocanegra-González and Guillemin 2018; Borges et al. 2018; Dos Santos et al. 2019; Gomes et al. 2021).

The absence of phylogenetic separation between *A. humile* and *A. occidentale* accessions could result from weak genetic differentiation and incomplete lineage sorting, suggesting they may represent incipient species that have not fully diverged. Generation time and growth habit can influence molecular evolution rates in plants. Herbaceous or annual plants, which are generally smaller with shorter generation times, tend to evolve faster, showing positive correlations between nucleotide substitution rates and herbaceous habit. Conversely, woody and tree species exhibit slower molecular evolution in both in the chloroplast and nuclear genome (Lanfear et al. 2013; Xin et al. 2020). Xie et al. (2014) estimated the divergence between *A. excelsum* and *A. occidentale* at approximately 20 million years ago. Considering this divergence of time and the fact that the evaluated accessions are long-lived trees and shrubs, the high conservation observed may be associated with these factors. In the genus *Cenostigma*, for example, genomic stability is observed, likely associated with the age of the genus, approximately 13.59 million years, the tree habit and the long-life cycle (Castro et al. 2023).

Genomic similarity may also be associated with the absence or low reproductive isolation between *A. humile* and *A. occidentale*, which is consistent with the high cytogenetic stability observed. This karyotypic conservation, the presence of morphological intermediates [accessions

with tree stature similar to cashew (*A. occidentale*) and fruits similar to those of *cajuí* (*A. humile*)], along with a supported clade of *A. occidentale* when intermediates were removed from the *ITS* tree, suggests possible hybridization. In a study involving natural populations referred to as *A. microcarpum* and *A. occidentale* from the coastal region of the state of Piauí, an attempt was made to distinguish the evaluated accessions of these two populations through leaf morphometry. The results of this study indicated that the populations of *A. microcarpum* have leaf morphometry very similar to those of *A. occidentale*; however, significant overlap in the data did not allow a clear distinction between these two taxa based on this characteristic. This raises the possibility of the existence of natural hybrids between *A. occidentale* and *A. microcarpum* in this region (Vieira et al. 2014).

Hybridization appears to be a common phenomenon within the genus *Anacardium* under natural conditions. Mitchell and Mori (1987) observed that three sympatric species from the Brazilian plateau (*A. occidentale*, *A. humile*, and *A. nanum*) bloom simultaneously where they coexist. Their flowers are morphologically indistinguishable and pollinated by the same insects, such as butterflies and bees, indicating few extrinsic barriers to cross-pollination. This may explain the occurrence of intermediate individuals between *A. occidentale* and *A. humile*, as well as between *A. humile* and *A. nanum*.

With remarkable morphological variation in habit and size of fruits and pseudofruits, it is also plausible to consider the influence of domestication on the overlap of these morphological traits. The process of plant domestication is characterized by genetic selection performed by humans to adapt wild plants to cultivation and human preferences, followed by breeding practices that can enhance productivity and resistance but reduce genetic diversity (N'Danikou and Tchokponhoue, 2020). *Anacardium occidentale* (cashew) is believed to have originated in the Cerrado of Central Brazil and subsequently colonized the sand dune restingas in Northeast Brazil. Central Brazil hosts a high diversity of *Anacardium* species, where the distribution of *A. occidentale* overlaps with the distribution areas of *A. humile*, *A. nanum*, and *A. corymbosum* (Mitchell and Mori 1987). Therefore, for now, we cannot exclude the possibility that the large fruit and cashew apple forms of *A. occidentale* (cashew) are domesticated forms of *cajuí* (*A. humile*), selected for larger fruits and pseudofruits. Under this scenario, this single species would have a wild range of habits but smaller fruits than its domesticated form. To test this hypothesis and better define species boundaries, future studies should use more informative molecular markers, such as SSRs or RADseq, ombined with a broader sampling of wild and domesticated accessions. This would help clarify the genetic structure of *Anacardium* species, their evolutionary relationships, and potential hybridization events.

Conclusion

Our results suggest significant morphological variability among *Anacardium* accessions from northeastern Brazil. However, this high level of phenotypic plasticity contrast with low levels of molecular variation observed in plastidial and nuclear regions, as well as the absence of cytogenetic differences. The loci used as markers here were not sufficiently informative to delimit species, suggesting either incomplete lineage sorting, the occurrence of hybridization, or differentiation within a single species due to domestication, processes that are not mutually exclusive.

Declarations

Acknowledgments The authors would like to thank Dr. Ana Cecília Ribeiro de Castro, from Embrapa Agroindústria Tropical, for providing seeds of *A. occidentale*, and technician Pedro Pereira Neves, from the Experimental Field of the UEP at Embrapa Meio-Norte, for the field support provided during the collections conducted in Parnaíba - PI. AG-O was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). This research was in part supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Finance Code 001).

Author's contributions SCS and AP-H conceived and designed the study. AG-O performed cytogenetic and molecular experiments and drafted the first version of the manuscript; TN performed plastome analyses; SCS collected the samples, performed mophoagronomic analyzed and co-supervised the work; PAB performed statistical analyses; CLSL performed taxonomical analyses; AP-H, provided resources and laboratory structure, and supervised the work. All authors discussed the data, read, and approved the final version of the manuscript.

Funding The study was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Data availability Raw data for Internal Transcribed Spacer (*ITS*) are available on NCBI under the accession numbers: PV089646, PV089647, PV089648, PV089649, PV089650, PV089651, PV089652, PV089653, PV089654, PV089655, PV089656, PV089657, PV089658, PV089659, PV089660, PV089661, PV089703. For the plastidial dataset, all raw sequences are available under the accession number: PRJNA1224782.

Competing interests The authors declare that they have no conflict of interest.

References

1. Abadi S, Azouri D, Pupko T, Itay M (2019) Model selection may not be a mandatory step for phylogeny reconstruction. Nat Commun 10:1-11. https://doi.org/10.1038/s41467-019-08822-w
2. Aliyu OM, Awopetu JA (2007) Chromosome studies in Cashew (*Anacardium occidentale* L.). Afr J Biotechnol 6:131–136. https://doi.org/10.4314/AJB.V6I2.56120
3. Aliyu OM (2012) Development of the flow cytometric protocol for ploidy analysis and determination of relative nuclear DNA content in Cashew (*Anacardium occidentale* Linn.). Am J Biochem Mol Biol 2:200–215. https://doi.org/10.3923/ajbmb.2012.200.215
4. Aliyu OM (2014) Analysis of absolute nuclear DNA content reveals a small genome and intra-specific variation in Cashew (*Anacardium occidentale* L.), Anacardiaceae. Silvae Genet 63:285–292. https://doi.org/10.1515/sg-2014-0036
5. Almeida CCS, Carvalho PCL, Guerra M (2007) Karyotype differentiation among *Spondias* species and the putative hybrid Umbu-cajá (Anacardiaceae). Bot J Linn Soc 155:541–547. https://doi.org/10.1111/j.1095-8339.2007.00721.x
6. Ariyarathne M, Yakandawala D, Barfus M, Heckenhauer J, Samuel R (2020) Molecular phylogeny and chromosomal evolution of endemic species of Sri Lankan Anacardiaceae. J Natl Sci Found Sri Lanka 48:289–303. https://doi.org/10.4038/jnsfsr. v48i3.9368
7. Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218. https://doi.org/10.1007/BF02672069
8. Asna AC, Jalaja SM, Smitha MS (2021) Phenotypic diversity and clustering of germplasm accessions of cashew for utilization and conservation. Electron J Plant Breed 12:1218–1226. https://doi.org/10.37992/2021.1204.167
9. Bai C, Alverson WS, Follansbee A, Waller DM (2012) New reports of nuclear DNA content for 407 U.S. plant species. Ann Bot 110:1623–1629. https://doi.org/10.1093/aob/mcs222
10. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol16:37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036
11. Bocanegra-González KT, Guillemin ML (2018) Guidelines for the restoration of the tropical timber tree *Anacardium excelsum*: first input from genetic data. Tree Genet Genomes 14:1–12. https://doi.org/10.1007/s11295-018-1271-z
12. Borém A, Miranda GV, Fritsche-Neto R (2021) Melhoramento de plantas. Oficina de Textos, São Paulo
13. Borges ANC, Lopes ACA, Britto FB, Vasconcelos LFL, Lima PSC (2018) Genetic diversity in a cajuí (*Anacardium* spp.) germplasm bank as determined by ISSR markers. Genet Mol Res 17:1–4. https://doi.org/10.4238/gmr18212
14. Borges WJO, Rambo MKD, Pires PS, Niculau ES, Silva FLN, Rambo MCD (2022) Evaluation of the physico-chemical potential and volatile profile of cashew (*Anacardium* spp.) in the Cerrado tocantinense. Res Soc Dev 11: 1–13. https://dx.doi.org/10.33448/rsd-v11i4.26966
15. Carvalho GMA, Carvalho CR, Soares FAF (2017) Flow cytometry and cytogenetic tools in *Eucalypts*: genome size variation x karyotype stability. Tree Genet. Genomes 13: 1-11. https://doi.org/10.1007/s11295-017-1186-0
16. Carneiro LA, Silva LS, Gomes MFC, Santos MF, Valente SES, Gomes RLF, Costa MF (2019) Morphological characterization and genetic divergence of a cashew population in Floriano, Piauí, Brazil. Genet Mol Res 18:1–18. http://dx.doi.org/10.4238/gmr18348
17. Castro ACR, Sobreira Júnior OV, Bordallo PN, Oliveira KGS, Bezerra CF (2011) Morphological Variability of Cashews from the Brazilian Savannah. Acta Hortic 2:863–869. https://doi.org/10.17660/ActaHortic.2011.918.114
18. Castro N, Mata-Sucre Y, Carvalho-Sobrinho J, Marques A, De Queiroz RT, Souza G (2023) Genomic stability in *Cenostigma* Tul., (Caesalpinioideae, Fabaceae): causes and consequences. Bot J LinnSoc 204:137–151. https://doi.org/10.1093/botlinnean/boad043
19. Costa L, Oliveira Á, Carvalho-Sobrinho J, Souza G (2017) Comparative cytomolecular analyses reveal karyotype variability related to biogeographic and species richness patterns in Bombacoideae (Malvaceae) Plant Syst Evol303:1131–1144. https://doi.org/10.1007/s00606-017-1427-6
20. Crespo MFV, Souza LI (2014) Cajuí: boas práticas e manejo sustentável. Sieart, Parnaíba
21. Da Luz L, Da Silva ACF, Laughinghouse HD (2015) Cytogenetic characterization of *Schinus terebinthifolius* Raddi (Anacardiaceae) accessions from Rio Grande do Sul state, Brazil. Caryologia 68:132–137. https://doi.org/10.1080/00087114.2015.1032573
22. Da Silva AR (2021) Biotools: Tools for Biometry and Applied Statistics in Agricultural Science. CRAN - Comprehensive R Archive Network. https://cran.r-project.org/package=biotools. Accessed 13 Apr 2025
23. Darlington CD, Janaki-Ammaal EK (1945) ChromosomeAtlas of Cultivated Plants. London
24. De Carvalho CR, Saraiva LS (1993) An air-drying technique for maize chromosomes without enzymatic maceration. Biotech Histochem 68:142–145. https://doi.org/10.3109/10520299309104684
25. Djolossè NK, Adoukonou-Sagbadja H, Gbèmavo CDSJ, Kodjo S, Badou A, Maliki R, Ahoyo-Adjovi NR (2019) Agro-morphological characterization of preselected cashew (*Anacardium occidentale* L.) mother trees in Benin farmer's plantations. J Agric Environ Int Dev113:17–34. https://10.12895/jaeid.20191.841
26. Doležel J, Sgorbati S, Lucretti S (1992) Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiol Plant85:625–631. https://doi.org/10.1111/j.1399-3054.1992.tb04764.x

27. Doležel J (2005) Plant DNA Flow Cytometry and estimation of nuclear genome size. Ann Bot 95:99–110. https://doi.org/10.1093/aob/mci005
28. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S (2015) Ycf1, the most promising plastid DNA barcode of land plants. Sci Rep 5:1–5. https://doi.org/10.1038/srep08348
29. Dos Santos JO, Mayo SJ, Bittencourt CB, De Andrade IM (2019) Genetic diversity in wild populations of the restinga ecotype of the cashew (*Anacardium occidentale*) in coastal Piauí, Brazil. Plant Syst Evol 305:913–924. https://doi.org/10.1007/s00606-019-01611-4
30. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479–491. https://doi.org/10.1093/genetics/131.2.479
31. Excoffier L, Lischer H (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10: 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x
32. Ferreira ME, Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análise genética. Embrapa Cenargen, Brasília
33. Ferreira EB, Cavalcanti PP, Nogueira DA (2021) ExpDes.pt: Pacote Experimental Designs (Português). CRAN - Comprehensive R Archive Network. https://doi.org/10.32614/CRAN.package.ExpDes. Accessed 13 Apr 2025
34. Figueredo A, Oliveira ÁWL, Carvalho-Sobrinho JG, Souza G (2016) Karyotypic stability in the paleopolyploid genus *Ceiba* Mill. (Bombacoideae, Malvaceae). Braz J Bot 39:1087–1093. https://doi.org/10.1007/s40415-016-0296-5
35. França KMA, Rocha LFCS, Sousa LFC, Melo RSS, Dantas ACA (2020) Caracterização morfológica de cajuí (*Anacardium* sp.) do Cerrado Sul Maranhense. Acta Tecnol 14:1–13. https://doi.org/10.35818/acta.v14i1.899
36. Garruti DS, Braga DC, Barbosa AED, Costa FNF, Da Silva NM, Vidal-Neto FC, Barros LM (2022) Atributos da qualidade de pedúnculos de cajueiro para consumo in natura. Embrapa Agroindústria Tropical, Fortaleza
37. Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885. https://doi.org/10.1093/nar/7.7.1869
38. Gill BS, Bir SS, Singhal VK (1979) In IOPB Chromosome number reports LXV. Taxon 28:630
39. Gill BS, Singhal VK, Bedi YS, Bir SS (1990) Cytological evolution in the woody taxa of *Pachmarhi* Hills. J Cytol Genet 25:308–320
40. Gomes JCM, Gomes NW, Silva LCA, Lima WA, DA Silva JM (2006) Caracterização pós-colheita de clones de cajueiro anão precoce no Oeste da Bahia. Bahia Agr 7:76–80
41. Gomes SO, Souza VABS, Costa MPSD, Silva CCP, Vale EM, Sousa M, Sousa JPB (2013) Avaliação da qualidade física e química de cajuí (*Anacardium* spp.) na região Meio-Norte. Geintec 3:139–145
42. Gomes MFDAC, Borges ANC, Batista GSS, Luz GDEA, Oliveira MEA, Lopes ACDEA, De Araújo ASF, Gomes RLF, Britto FB, Lima PSDAC, Valente SES (2021) Genetic diversity and structure in natural populations of Cajui from Brazilian Cerrado. Biosci J 37: 1–11. https://doi.org/10.14393/BJ-v37n0a2021-53974
43. Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genet Res Int431487:1–14. https://doi.org/10.1155/2015/431487
44. Guerra M, Pedrosa A, Silva AEB, Cornélio MTM, Santos KGB, Soares-Filho WS (1997) Chromosome number and secondary constriction variation in 51 accessions of a *Citrus* germoplasm bank. Braz J Genet 20:489–496. https://doi.org/10.1590/S0100-84551997000300021
45. Guerra S, Guerra M, Mendes S, Soares-Filho WS, Pedrosa-Harand A (2020) Karyotype variability of sour orange (*Citrus aurantium* L.) and the origin of its heteromorphic karyotype. Tree Genet Genomes 16:1–10. https://doi.org/10.1007/s11295-020-01471-x
46. Hamrick JL (2012) Gene movement in tropical tree populations – Tropical breeding systems: one and done? Heredity 109:330–331. https://doi.org/10.1038/hdy.2012.47
47. He L, Zhao H, He J, Yang Z, Guan B, Chen K, Hong Q, Wang J, Liu J, Jiang J (2020) Extraordinarily conserved chromosomal synteny of *Citrus* species revealed by chromosome-specific painting. Plant J 103:2225–2235. https://doi.org/10.1111/tpj.14894
48. Hoang PTN, Fuchs J, Schubert V, Tran TBN, Schubert I (2022) Chromosome numbers and genome sizes of all 36 Duckweed species (Lemnaceae). Plants 11:1–8. https://doi.org/10.3390/plants11202674
49. Horjales M, Redondo N, Blanco A, Rodríguez MA (2003) Cantidades de DNA nuclear en árboles y arbustos. Nova Acta Cient Compost Biol 13:23–33
50. IBPGR (1986) International Board for Plant Genetic Resources: Cashew Descriptors. IBPGR, Rome
51. Kassambara A, Mundt F (2020) factoextra: Extract and Visualize the Results of Multivariate Data Analyses. CRAN - Comprehensive R Archive Network. https://cran.r-project.org/package=factoextra. Accessed 13 Apr 2025
52. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649. https://doi.org/10.1093/bioinformatics/bts199

53. Lanfear R, Ho S, Davies TJ, Moles AT, Aarssen L, Swenson NG, Warman L, Zanne AE, Allen AP (2013) Taller plants have lower rates of molecular evolution. Nat Commun 4:1–7. https://doi.org/10.1038/ncomms2836
54. Las Peña ML, Bernardello GLM, Steibel P, Troiani H (2006) Cytogenetic studies in *Schinus* (Anacardiaceae) Missouri Botanical Garden. Arnaldoa 13:270–275
55. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187
56. Lima VPMS, Ramos AD, Franca FMC (1988) A cultura do cajueiro no Nordeste do Brasil. Banco do Nordeste do Brasil, Fortaleza
57. Lopes MMDEA, De Moura CFH, De Aragão FAZ, Cardoso TG, Filho JE (2011) Caracterização física de pedúnculos de clones de cajueiro anão precoce em diferentes estádios de maturação. Rev Ciênc Agron 42:914–920. https://doi.org/10.1590/S1806-66902011000400013
58. Loureiro J, Rodriguez E, Dolezel J, Santos C (2007) Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. Ann Bot 100:875–888. https://doi.org/10.1093/aob/mcm152
59. Machado O (1944) Estudos novos sobre uma planta velha – o cajueiro (*Anacardium occidentale* L.). Rodriguesia 8:19–48
60.
61. Melo CAF, Martins MIG, Oliveira MBM, Benko-Iseppon AM, Carvalho R (2011) Karyotype analysis for diploid and polyploid species of the *Solanum* L. Plant Syst Evol293:227–235. https://doi.org/10.1007/s00606-011-0434-2
62. Mitchell JD, Mori SA (1987) The cashew and its relatives (*Anacardium*: Anacardiaceae), Memoirs of the New York Botanical Garden-NYBG Press, New York
63. Mitchell JD, Pel SK, Bachelier JB, Warschefsky EJ, Joyce EM, Canadell LC, Silva-Luz CL, Coiffard C (2022) Neotropical Anacardiaceae (cashew family). Braz J Bot45:139–180. https://doi.org/10.1007/s40415-022-00793-5
64. Mojena R (1977) Hierarchical grouping methods and stopping rules: an evaluation. Comput J20:359–363. https://doi.org/10.1093/comjnl/20.4.359
65. N'Danikou S, Tchokponhoue DA (2019) Plant domestication for enhanced food security. In: Leal Filho W, Azul A, Brandli L, Ozuyar P, Wall T (eds) Zero Hunger. Encyclopedia of the UN Sustainable Development Goals, 1st edn. Springer, Cham, pp 1-12. https://doi.org/10.1007/978-3-319-69626-3_96-1
66. Nobre LLM, Santos JDOD, Leite R, Almeida C (2018) Phylogenomic and single nucleotide polymorphism analyses revealed the hybrid origin of *Spondias bahiensis* (family Anacardiaceae): de novo genome sequencing and comparative genomics. Genet Mol Biol41:878–883. https://doi.org/10.1590/1678-4685-GMB-2017-0256
67. Ohri D, Bhargava A, Chatterjee A (2004) Nuclear DNA amounts in 112 species of tropical hardwoods - new estimates. Plant Biol 6:555–561. https://doi.org/10.1055/s-2004-821235
68. Oliveira VF, Silva FG, Resende EC, Pereira PS, Silva FHL, Egea MB (2019) Physicochemical characterization of 'Cerrado' cashew (*Anacardium othonianum* Rizzini) fruits and pseudofruits. J Sci Food Agric 99:6199–6208. https://doi.org/10.1002/jsfa.9892
69. Paiva FFA, Garruti DS, Silva Neto RM (2000) Aproveitamento Industrial do caju. Embrapa-CNPAT, Fortaleza
70. Pedrosa PH, Gitaí J, Silva AEB, Felix LP, Guerra M (1999) Citogenética de angiospermas coletadas em Pernambuco. Acta Bot Bras13:49–60. https://doi.org/10.1590/S0102-33061999000100006
71. Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A (2002) Chromosomal map of the model legume *Lotus japonicus*. Genetics 161:1661–1672. https://doi.org/10.1093/genetics/161.4.1661
72. Pereira LD (2018) Caracterização e diversidade genética de frutos de cajuzinho-do-cerrado. Dissertation, Universidade Federal de Goiás
73. Pereira LP, Da Silva DFP, De Souza LKF, Pereira ETL, Da Assunção HF, Costa MM (2019) Genetic diversity of bushy cashew (*Anacardium humile* A. St.-Hil.) based on characteristics of fruits. Rev Bras Frutic 41:1–6. https://doi.org/10.1590/0100-29452019065
74. Pierozzi NI, Rossetto CJ (2011) Chromosome characterization of two varieties of *Mangifera indica* L. Rev Bras Frutic 33:546–551. https://doi.org/10.1590/S0100-29452011000500074
75. Plants of the World-POWO (2024) Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet. http://www.plantsoftheworldonline.org. Accessed 31 March 2024
76. R Core Team. R: (2022) A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.r-project.org/. Accessed 20 January 2023
77. Rabah SO, Lee C, Hajrah NH, Makki RM, Alharby HF, Alhebshi AM, Sabir JSM, Jansen RK, Ruhlman TA (2017) Plastome Sequencing of ten nonmodel crop species uncovers a large insertion of mitochondrial DNA in cashew. Plant Genome 10:1–14. https://doi.org/10.3835/plantgenome2017.03.0020
78. Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6. Available online at. 2014. http://beast.bio.ed.ac.uk/Tracer. Accessed 02 February 2024

79. Rocha MS, Figueiredo RW de, Araújo MA da M, Moreira-Araújo RS dos R (2013) Physicochemical Characterization and In Vitro Antioxidant Activity of Fruits from the Piauí Cerrado. Rev Bras Frutic 4:933–941. https://doi.org/10.1590/S0100-29452013000400003
80. Ronquist F, Teslenko M, Mark PVD, Ayres DL, Darling A, Höhna S, Larget B S, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029
81. Rufino MSM (2004) Qualidade e potencial de utilização de cajuís (*Anacardium* spp.) oriundos da vegetação litorânea do Piauí. Dissertation, Federal University of Piauí
82. Salgotra RK, Chauhan BS (2023) Genetic diversity, conservation, and utilization of plant genetic resources. Genes Acre14:1–20. https://doi.org/10.3390/genes14010174
83. Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am J Bot84:1120–1136. https://doi.org/10.2307/2446155
84. Sancho SDEO, Maia GA, De Figueiredo RW, Rodrigues S, de Sousa PHM (2007) Physicochemical changes in cashew apple (*Anacardium occidentale* L.) Juice processing. Food Sci Technol 27:878–882. https://doi.org/10.1590/S0101-20612007000400031
85. Santos RC, Santos-Júnior JE (2015) Genetic divergence for multivariate analysis of phenotypic characters of *Anacardium humile* (St. Hilaire). Rev Ceres 62:553–560. https://10.1590/0034-737X201562060007
86. Savadi S, Muralidhara BM, Godwin J, Adiga JD, Mohana GS, Eradasappa E, Shamsudheen M, Karun A (2022) De novo assembly and characterization of the draft genome of the cashew (*Anacardium occidentale* L.). Sci Rep12:1–13. https://doi.org/10.1038/s41598-022-22600-7
87. Semporé JN, Songré-Ouattara LT, Tarpaga WV, Bationo F, Dicko MH (2021) Morphological characterization and quality assessment of cashew (*Anacardium occidentale* L.) nuts from 53 accessions of Burkina Faso. J Agric Food Res6: 1–8. https://doi.org/10.1016/j.jafr.2021.100219
88. Silva-Júnior JF, Souza FVD, Pádua JG (2021) A arca de Noé das frutas nativas brasileiras. Embrapa, Brasília
89. Silva-Luz CLS, Pirani JR, Mitchell JD, Daly D, Capelli NV, Demarco D, Pell SK, Plunkett, GM (2019) Phylogeny of *Schinus* L. (Anacardiaceae) with a new infrageneric classification and insights into evolution of spinescence and floral traits. Mol Phylogenet Evol 133:302–351. https://doi.org/10.1016/j.ympev.2018.10.013
90. Singh D (1981) The relative importance of characters affecting genetic divergence. Indian J Genet Plant Breed 41:237–245
91. Sokal RR, Rohlf FJ (1962) The comparison of dendrograms by objeticve methods. Taxon 11:30–40. https://doi.org/10.2307/1217208
92. Sola-Campoy PJ, Robles F, Schwarzacher T, Ruiz Rejón C, De La Herrán R, Navajas-Pérez R (2015) The Molecular Cytogenetic characterization of pistachio (*Pistacia vera* L.) suggests the arrest of recombination in the largest heteropycnotic pair HC1. PLoS One 10:1–15. https://doi.org/10.1371/ journal. pone.0143861
93. Sultana Z, Al Mamum ASM, Islam AA, Mannan MD, Islam MD (2022) Morpho-molecular diversity of cashew nut (*Anacardium occidentale* L.) germplasm of Bangladesh. Plant Tissue Cult Biotechnol 32:167–180. https:// doi.org/10.3329/ptcb.v32i2.63551
94. Sun Y, Skinner DZ, Liang GH (1994) Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theor Appl Genet 89:26–32. https:// doi.org/10.1007/BF00226978
95. Schäferhoff B, Fleischmann A, Fischer E, Albach DC, Borsch T, Heubl G, Müller K.F (2010) Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. BMC Evol Biol10:1–22. https://doi.org/10.1186/1471-2148-10-352
96. Scott A, Knott M (1974) Cluster-analysis method for grouping means in analysis of variance. Biometrics 30:507–512. https://doi.org/10.2307/2529204
97. Shepherd LD, De Lange PJ, Cox S, McLenachan PA, Roskruge NR, Lockhart PJ (2016) Evidence of a strong domestication bottleneck in the recently cultivated New Zealand endemic root crop, *Arthropodium cirratum* (Asparagaceae). PLoS One 11:1–16. https://doi.org/10.1371/journal.pone.0152455
98. Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three noncoding regions of chloroplast DNA. Plant Mol Biol17:1105–1109. https://doi.org/10.1007/BF00037152
99. Takehana CLI, Ohashi ST, Jardim MAG, Dos Santos JUM (2013) Biologia floral e visitantes florais de *Anacardium giganteum* W. Hancock ex Engl. (Anacardiaceae) no município de Bragança, Pará. Rev Ciênc Agrár 56:202–211. https:// 10.4322/rca.2013.030
100. Turchetto-Zolet AC, Turchetto C, Zanella CM, Passaia G (2017) Marcadores moleculares na era genômica: metodologias e aplicações. Sociedade Brasileira de Genética, Ribeirão Preto
101. Vaio M, Nascimento J, Mendes S, Ibiapino A, Felix LP, Gardner A, Emshwiller E, Fiaschi P, Guerra M (2018) Multiple karyotype changes distinguish two closely related species of *Oxalis* (*O. psoraleoides* and *O. rhombeo-ovata*) and suggest an artificial grouping of section Polymorphae (Oxalidaceae). Bot J Linn Soc188:269–280. https://doi.org/10.1093/botlinnean/boy054

102. Vieira M, Mayo SJ, De Andrade IM (2014) Geometric morphometrics of leaves of *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the coastal region of Piauí. Braz J Bot 37:315–327. https://doi.org/10.1007/s40415-014-0072-3
103. Waminal NE, Pellerin RJ, Kim NS, Murukarthick J, Jee YP, Tae-Jin Y, Hyun HK (2018) Rapid and efficient fish using pre-labeled oligomer probes. Sci Rep 8:1–10. https://doi.org/10.1038/s41598-018-26667-z
104. White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Michael A (ed) PCR Protocols: a guide to methods and applications, 4th edn. Academic Press, San Diego, pp 482
105. Xie L, Yang ZY, Wen J, Li DZ, Yi TS (2014) Biogeographic history of *Pistacia* (Anacardiaceae), emphasizing the evolution of the Madrean-Tethyan and the eastern Asian-Tethyan disjunctions. Mol Phylogenet Evol 77:136–146. https://doi.org/10.1016/j.ympev.2014.04.006
106. Xin HY, Zhang T, Wu YF, Zhang WL, Zhang PD, Xi ML, Jiang JM (2020) An extraordinarily stable karyotype of the woody *Populus* species revealed by chromosome painting. Plant J 101:253–264. https://doi.org/ 10.1111/tpj.14536
107. Yi T, Miller AJ, Wen J (2004) Phylogenetic and biogeographic diversification of *Rhus* (Anacardiaceae) in the Northern Hemisphere. Mol Phylogenet Evol 33:861–879. https://doi.org/10.1016/j.ympev.2004.07.006
108. Yonemori K, Nishiyama K, Choi Y (2010) Physical mapping of 5S and 45S rDNAs by fluorescent in situ hybridization in mango (*Mangifera indica* L.). Acta Hortic 864:133–139. https://doi.org/10.17660/ActaHortic.2010.864.18
109. Zerey-Belaskri AE, Ribeiro T, Alcaraz ML, Zerey WE, Castro S, Loureiro J, Benhassaini H, Iñaki-Hormaza J (2018) Molecular characterization of *Pistacia atlantica* Desf. subsp. *atlantica* (Anacardiaceae) in Algeria: Genome size determination, chromosome count and genetic diversity analysis using SSR markers. Sci Hortic 227:278–287. https://doi.org/10.1016/j.scienta.2017.09.016
110. Zuffo AM, Busch A, Steiner F, Alves CZ (2019) Biometric characteristics of fruits, seeds and plants of *Hancornia speciosa* Gomes (Apocynaceae). Aust J Crop Sci 13:622–627. https://doi.org/10.21475/ajcs.19.13.04.p16

Figures

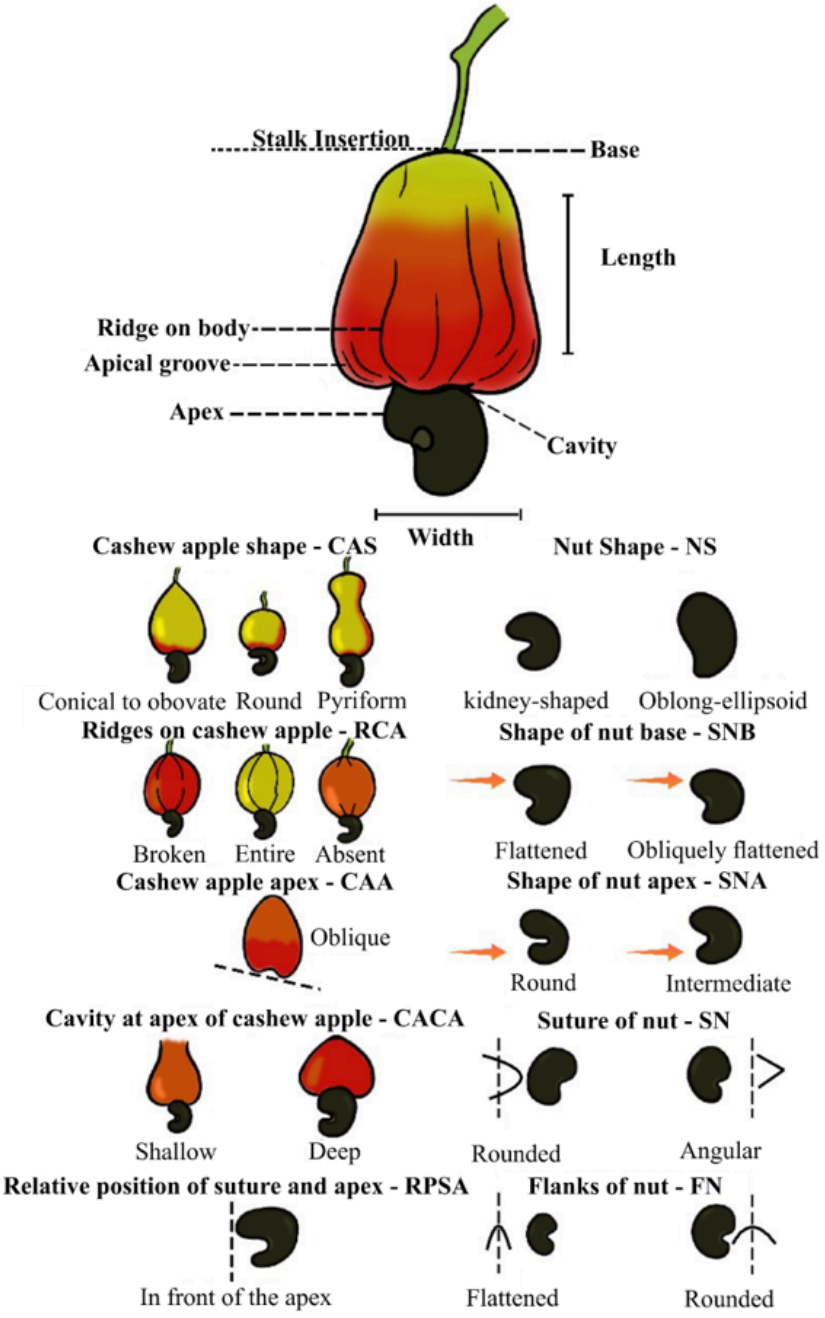

Figure 1

Qualitative traits related to the fruits (nuts) and peduncles (pseudofruits) of *A. occidentale* e *A. humile*used in the morphological analysis

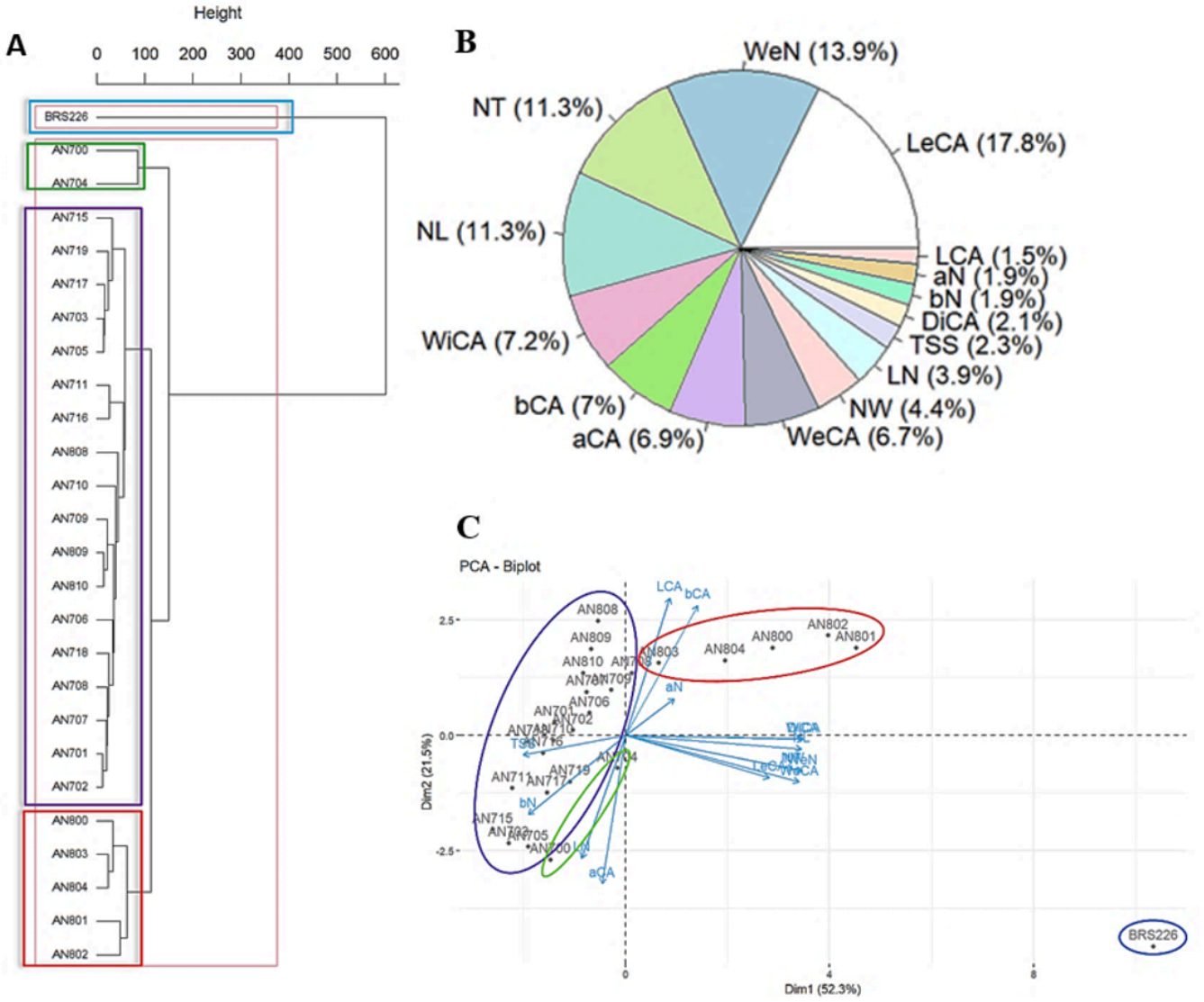

Figure 2

Similarity among accessions from natural populations of cashew, as well as the commercial accession BRS 226. **(A)**Dendrogram obtained using the UPGMA method; **(B)** Relative importance of the variables related to the quantitative data of the cashew apple and nut. Description of the evaluated traits: cashew apple length (LeCA); cashew apple width (WiCA); cashew apple diameter (DiCA); cashew apple weight (WeCA); total soluble solids (TSS); nut length (NL); nut width (NW); nut thickness (NT); nut weight (WeN); cashew apple color (LCA, aCA, and bCA); and nut color (LN, aN, and bN). **(C)** Principal Component Analysis (PCA-Biplot) of the morphoagronomic characteristics of the cashew apple and nut. In figures **(A)**and **(C),** except for BRS 226, the accessions were sampled from different locations: Alvorada do Gurguéia (AN-705, AN-706, AN-707, AN-708, and AN-709), Bom Jesus (AN-700, AN-701, AN-702, AN-703, AN-704, AN-710, and AN-711), Buriti dos Lopes (AN-808, AN-809, and AN-810), Currais (AN-715, AN-716, AN-717, AN-718, and AN-719), and Parnaíba (AN-800, AN-801, AN-802, AN-803, and AN-804). The first group is represented in blue, the second in pink, while subgroups I, II, and III are represented in green, purple, and red, respectively

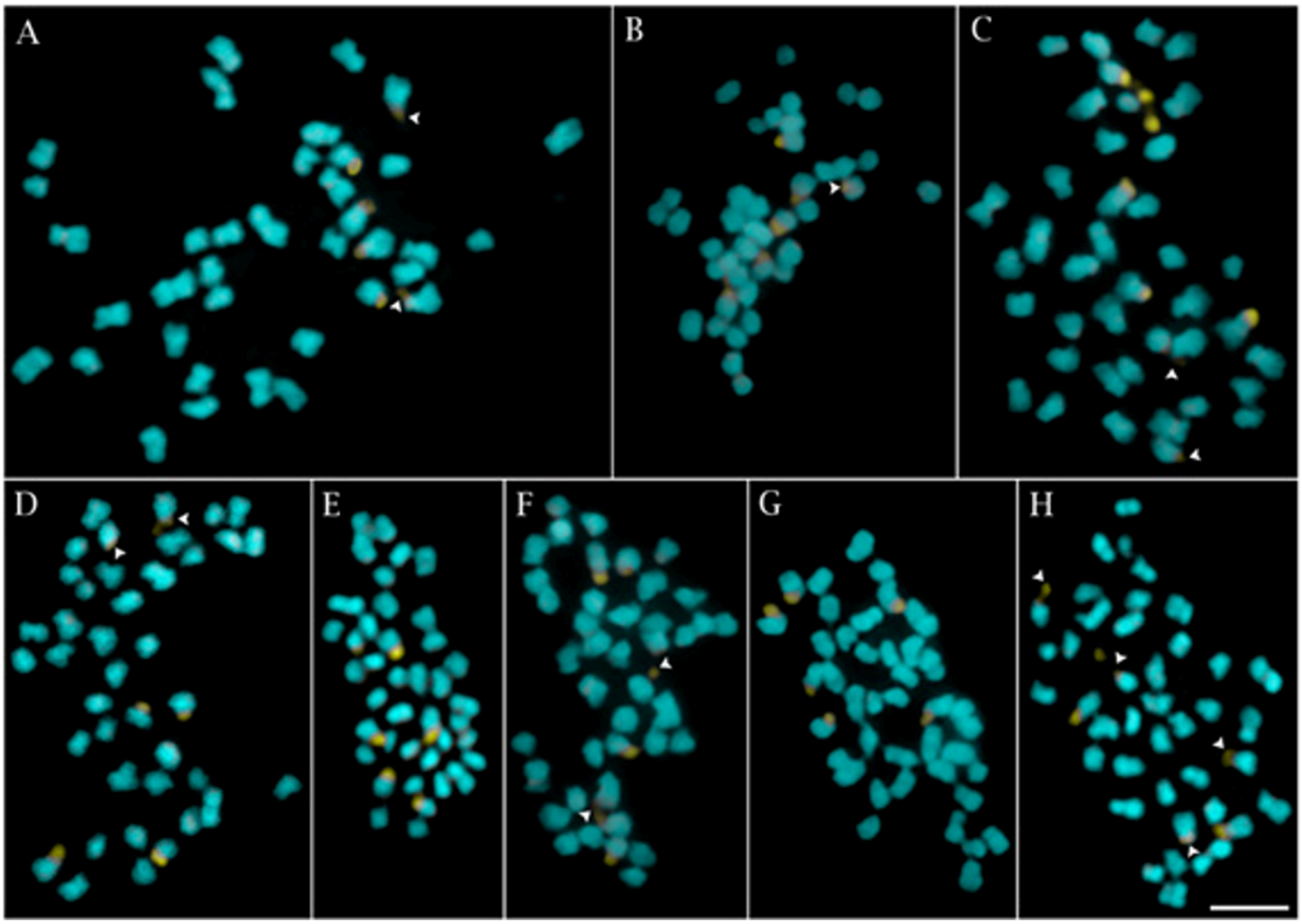

Figure 3

Mitotic metaphases of *A. occidentale* (A-D and F-H), *A. humile* (E), stained with CMA (yellow) and DAPI (blue). AN-706 (A), AN-715 (B), AN-704 (C), AN-701 (D), AN-712 (E), AN-803 (F), AN-708 (G), AN-714 (H). Arrowheads indicate weaker terminal $CMA^{+}/DAPI^{-}$ bands. Bar in (I) corresponds to 5 μm

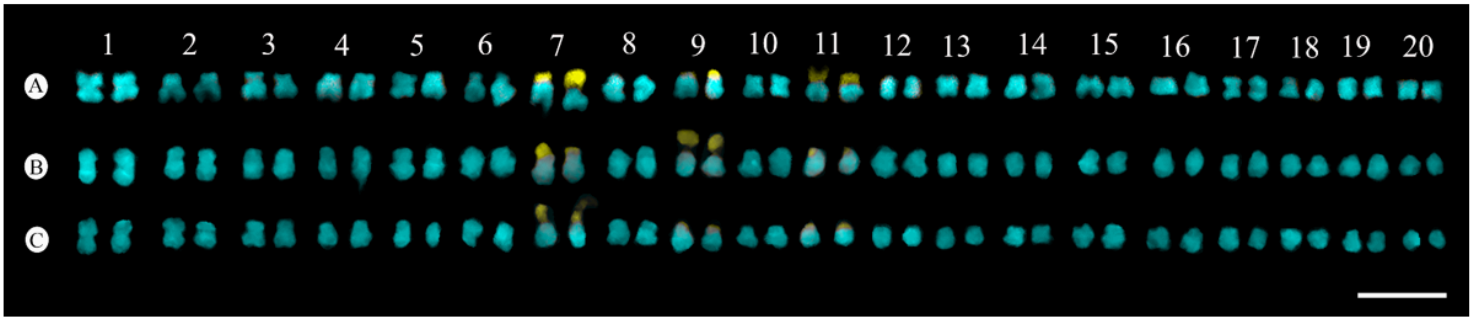

Figure 4

Karyograms of two *Anacardium* species, showing chromosome number, morphology, and distribution of CMA^{+} heterochromatic bands. (A) *A. occidentale* (BRS 226), (B) *A. occidentale* previously identified as *A. othonianum* (BGC-45), and (C) *A. humile* (AN-712). Bar corresponds to 5 μm

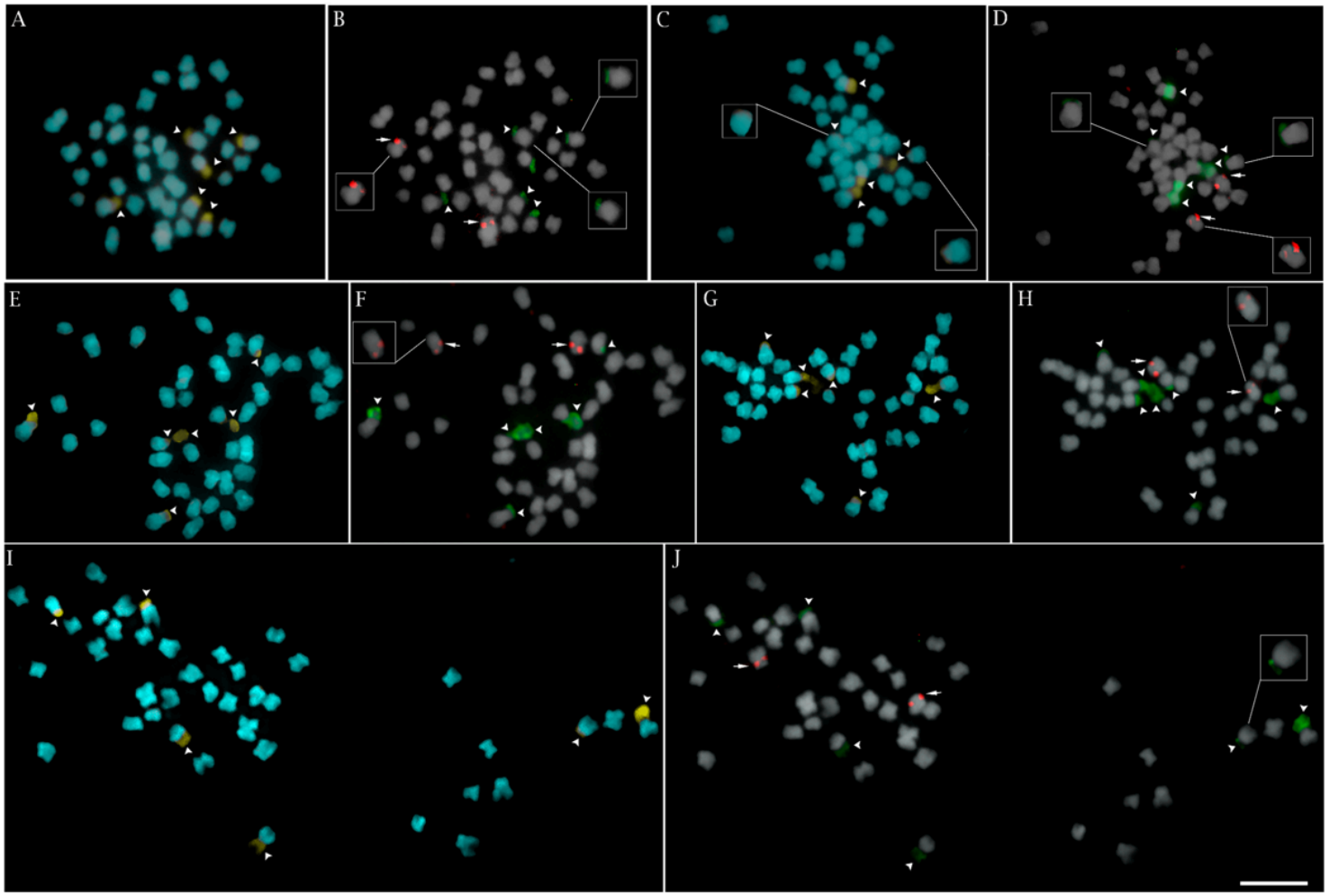

Figure 5

Distribution of $CMA^{+}/DAPI^{-}$ heterochromatin (A, C, E, G, I) and rDNA sites (B, D, F, H, J) in *Anacardium occidentale* (A-F, I-J,) and *A. humile* (G-H). 5S rDNA (red, arrows) and 35S rDNA (green, arrowheads). A - B) AN-707; C - D) AN-801; E - F) BGC-45 previously identified as *A. othonianum*; G - H) AN-712; and I - J) BRS 226. Scale bar in (J) corresponds to 5 µm

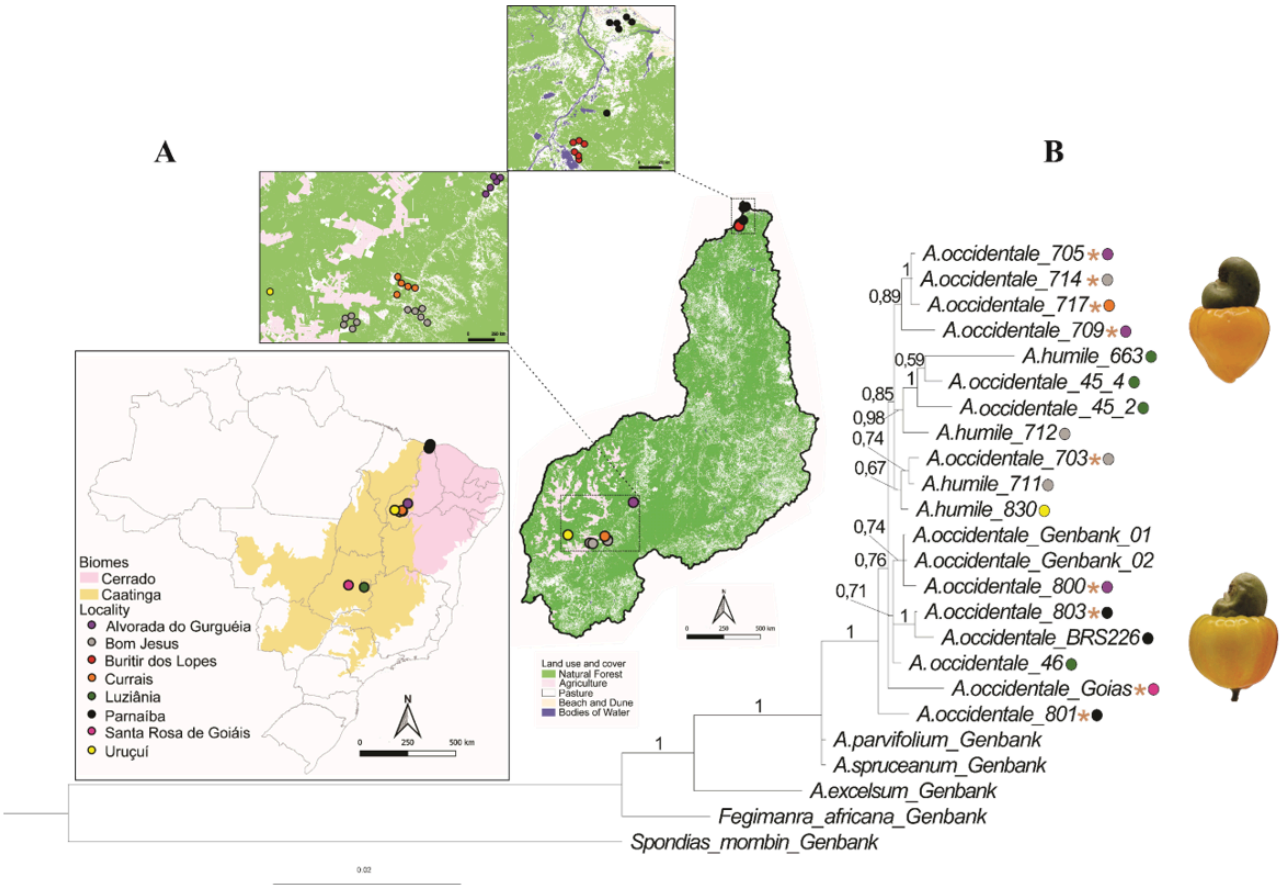

Figure 6

Geographical distribution and phylogenetic relationship of *Anacardium* accessions analyzed. **A**. Map of Brazil showing the locations and respective biomes where samples were collected, with the state of Piauí highlighted in green, indicating areas with more collection sites, along with an enlargement of these collection locations. **B**. Bayesian tree of *Anacardium* derived from the combined analysis of nuclear *ITS*and plastid regions (*matK*, *trnL_trnF*, *rps16*, and *ycf1*). Posterior probability (PP) values are shown for each node, and asterisks indicate accessions of *Anacardium occidentale* with *cajuí*-type fruits and apples, i.e., possible hybrids

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterialDoessizematterMorphologicalandgeneticsimilaritiesbetweencashewAnacardiumoccidentaleandcajuAA.humile.docx