UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA VEGETAL

ACALENE GONÇALVEIS DE OLIVEIRA

DIVERSIDADE GENÉTICA EM POPULAÇÕES NATURAIS DE *Anacardium occidentale* L. e *Anacardium humile* A.St.-Hil.

Recife - PE
2024

ACALENE GONÇALVEIS DE OLIVEIRA

DIVERSIDADE GENÉTICA EM POPULAÇÕES NATURAIS DE *Anacardium occidentale* L. e *Anacardium humile* A.St.-Hil.

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Vegetal da Universidade Federal de Pernambuco, como requisito para o título de mestre em Biologia Vegetal.

Área de concentração: Sistemática e Evolução.

Orientadora: Profª. Dra. Andrea Pedrosa-Harand
Coorientador: Prof. Dr. Silvokleio da Costa Silva

Recife - PE
2024

ACALENE GONÇALVEIS DE OLIVEIRA

DIVERSIDADE GENÉTICA EM POPULAÇÕES NATURAIS DE *Anacardium occidentale* L. e *Anacardium humile* A.St.-Hil.

Dissertação apresentada ao Programa de Pós-graduação em Biologia Vegetal da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre(a) em Biologia Vegetal.

Aprovada em 29/04/2024.

BANCA EXAMINADORA

Profª. Dra. Andrea Pedrosa Harand (Orientadora)
Universidade Federal de Pernambuco

Prof. Dr. Cícero Carlos de Souza Almeida (Examinador Externo)
Universidade Federal de Alagoas

Profª. Dra. Artemisa Nazaré Costa Borges Martins (Examinadora Externa)
Instituto Federal de Educação, Ciência e Tecnologia do Maranhão

Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Oliveira, Acalene Gonçalveis de.
Diversidade genética em populações naturais de Anacardium occidentale L. e Anacardium humile A.St.-Hil / Acalene Gonçalveis de Oliveira. - Recife, 2024.
101 f.: il.

Dissertação (Mestrado) - Universidade Federal de Pernambuco, Centro de Biociências, Programa de Pós-Graduação em Biologia Vegetal, 2024.
Orientação: Andrea Pedrosa-Harand.
Coorientação: Silvokleio da Costa Silva.
Inclui referências e anexo.

1. Anacardiaceae; 2. Análise filogenética; 3. Caracterização agronômica e citogenética. I. Pedrosa-Harand, Andrea. II. Silva, Silvokleio da Costa. III. Título.

UFPE-Biblioteca Central

Dedico este trabalho à Sra. Conceição, minha mãe,
meu exemplo de resiliência.

AGRADECIMENTOS

Gostaria de expressar minha gratidão a todos cuja colaboração foi fundamental para a concretização deste trabalho.

Agradeço, primeiramente, à Deus.

À Sra. Conceição, minha mãe, minha melhor amiga e exemplo de resiliência, obrigada pelo carinho e apoio que me foram concedidos nos momentos mais desafiadores da minha jornada acadêmica, por sempre acreditar em mim e me encorajar a estudar.

Ao Sr. Raimundo, meu pai, pelo cuidado, proteção, apoio e por estar sempre presente em minha vida. Muito obrigada.

Aos meus irmãos Cristiano, Ana Cristina e Maria, pela união, companheirismo e por sempre ouvirem meus desabafos.

Um carinhoso agradecimento ao meu avô Manoel, obrigada pelas sábias e lindas palavras de motivação.

A minha avó, Maria, obrigada pelo apoio.

Agradeço também as minhas amigas Juliana, Kaliane e Renata Fonseca, pela disponibilidade e pelas palavras de incentivo.

A Prof. Dra. Andrea, minha orientadora, agradeço por aceitar me orientar, pela empatia, pela enorme paciência, compreensão, principalmente nos momentos em que precisei de tempo, obrigada pelo profissionalismo e ensinamentos. Foi muito gratificante conhecê-la e trabalhar com você.

Ao Prof. Dr. Silvokleio, meu coorientador, agradeço pela amizade, enorme paciência, empatia, pelo carinho paterno, por ser sempre tão compreensivo comigo, pelas sábias palavras que me ajudaram em meu crescimento profissional e pessoal, pela dedicação ao nosso trabalho e pelas contribuições e ensinamentos para minha formação, obrigada por tudo.

Agradeço também a Thiago Henrique, pelo treinamento dado no início do trabalho no laboratório.

Agradeço a todos os alunos que passaram pelo Laboratório de Citogenética e Evolução Vegetal da UFPE, enquanto estive por lá: Ana, Júlia, Géssica, Natália, Gustavo Gomes,

Jéssica, Artur, Yennifer, Amália, Yhanndra, Lucas, Thiago Henrique, Claudio, Bruna, Gustavinho, Brena,

Renata Lima, Kaiane, Leylson e Augusto. Obrigada pelo acolhimento e pela companhia diária de vocês.

Agradeço especialmente a Ana, Lucas, Natália, Gustavo Gomes e Yennifer. Obrigada pela companhia constante, pelos passeios por Recife, pelas festinhas na casa de Lucas e pelas valiosas ajudas no laboratório.

Aos estudantes Marcos Vinícius, Felipe Marques, Ítalo José e João Pedro, pelo auxílio com as coletas de campo, essencial para o desenvolvimento deste trabalho.

Ao Dr. Thales Coutinho, também pela ajuda em coletas de materiais.

A profa. Dra. Priscila Alves Barroso, pela ajuda com as análises estatísticas.

Ao técnico Pedro Pereira Neves, do Campo Experimental da UEP da Embrapa Meio-Norte pelo apoio de campo dado nas coletas realizadas em Parnaíba – PI.

A curadora do herbário da UFPE, Marlene, juntamente com demais funcionários do herbário, a ajuda de vocês foi fundamental para a identificação de nosso material.

A taxonomista Cíntia Luíza da Silva Luz pela disponibilidade e identificação taxonômica de nosso material.

A Dra. Ana Cecília Ribeiro de Castro, pesquisadora da Embrapa Agroindústria Tropical, por fornecer sementes.

À CAPES e ao CNPQ pela concessão da bolsa de mestrado.

À Universidade Federal de Pernambuco (UFPE), pela oportunidade da realização do Curso de Mestrado.

Ao Programa de Pós-Graduação em Biologia Vegetal (PPGBV). Aos docentes e aos secretários deste programa de pós-graduação.

Enfim, agradeço a todos que participaram de forma direta e indireta para realização deste trabalho.

RESUMO

Anacardium occidentale L. e *A. humile* A. St. -Hil. são frutíferas pertencentes à família Anacardiaceae conhecidas, respectivamente, como caju e cajuí. Essas espécies apresentam ampla diversidade morfológica, com frutos e pseudofrutos apresentando formatos, tamanho e coloração diferentes. As relações filogenéticas entre *A. occidentale*, *A. humile* e as demais espécies do gênero ainda são obscuras, devido à ausência de uma filogenia, tornando difícil compreender a evolução do grupo. Além disso, as análises cariotípicas publicadas se limitam a contagens cromossômicas realizadas para *A. occidentale,* principal representante comercial do grupo. Assim, o presente trabalho teve o objetivo de analisar a diversidade genética de populações naturais de cajuís do Piauí, utilizando abordagens agronômicas, moleculares e citogenéticas, a fim de compreender sua diversidade e relações evolutivas. A abordagem agronômica contou com o emprego de descritores quali- e quantitativos para o pedúnculo (pseudofruto) e a castanha (fruto verdadeiro) em vinte e quatro acessos de *A. occidentale* e dois acessos de *A. humile*, revelando a existência de variabilidade genética significativa entre os germoplasmas estudados. O emprego de marcadores plastidiais e nucleares confirmou a natureza monofilética de *Anacardium*. Contudo, os marcadores utilizados não se mostraram eficientes para elucidar as relações filogenéticas entre as espécies, devido ao baixo polimorfismo. Doze acessos morfologicamente divergentes de *A. occidentale*, além dos dois acessos de *A. humile* revelaram estabilidade cariotípica, tanto no que diz respeito ao número cromossômico ($2n = 40$), quanto à distribuição de bandas $CMA^{+}/DAPI^{-}$, localizadas na região terminal do braço curto de três pares cromossômicos. Adicionalmente, os resultados obtidos por meio da FISH revelaram que todos os acessos apresentaram três pares de sítios de DNAr 35S colocalizados com as bandas $CMA^{+}/DAPI^{-}$ e um par de sítios de DNAr 5S localizados na região subterminal do braço curto de outro par cromossômico, sem colocalização com sítios $CMA^{+}/DAPI^{-}$. O tamanho do genoma de *A. occidentale* e *A. humile* foram estimados por citometria de fluxo, revelando-se relativamente pequenos, com um valor médio de 0,44 pg/1C (435 Mbp), 0,44 pg/1C (434 Mbp), respectivamente. Tais dados sugerem a inexistência de variação citogenética intra- ou interespecífica significativa entre os acessos, não havendo indícios de poliploidia entre as matrizes estudadas. Outros marcadores devem ser empregados para esclarecer os limites taxonômicos e relações entre *A. occidentale*, *A. humile* e espécies

irmãs, além de possível correlação entre a variação agronômica e uma estruturação genética da espécie.

Palavras-chave: Anacardiaceae; caracterização agronômica; análise filogenética; CMA/DAPI; DNAr 5S e 35S; FISH.

ABSTRACT

Anacardium occidentale L. and *A. humile* A. St.-Hil. are fruit-bearing species from the Anacardiaceae family, commonly known as cashew and cashew nut. These species exhibit a wide range morphological diversity, with fruits and pseudofruits displaying varying shapes, sizes, and colors. The phylogenetic relationships between *A. occidentale*, *A. humile*, and other species within the genus, remain unclear due to the absence of a phylogenetic analysis, making it difficult to understand the evolutionary history of the group. Furthermore, published karyotypic analyses are limited to chromosomal counts conducted on *A. occidentale*, the main commercial representative of the group. Therefore, the objective of this study was to analyze the genetic diversity of natural populations of cashew trees from Piauí, using agronomic, molecular, and cytogenetic approaches in order to understand their diversity and evolutionary relationships. The agronomic approach involved the use of qualitative and quantitative descriptors for the peduncle (pseudofruit) and the nut (true fruit) in twenty-four accessions of *A. occidentale* and two accessions of *A. humile*, revealing significant genetic variability among the studied germplasms. The use of plastidial and nuclear markers confirmed the monophyletic nature of *Anacardium*. However, the markers used were not efficient in elucidating the phylogenetic relationships between the species due to low polymorphism. Twelve morphologically divergent accessions of *A. occidentale*, along with the two accessions of *A. humile*, showed karyotypic stability, both in terms of chromosomal number ($2n = 40$) and in the distribution of $CMA^+/DAPI^-$ bands, located in the terminal region of the short arm of three chromosomal pairs. Additionally, results obtained through FISH revealed that all accessions had three pairs of 35S rDNA sites co-localized with the $CMA^+/DAPI^-$ bands and one pair of 5S rDNA sites located in the subterminal region of the short arm of another chromosomal pair, without co-localization with $CMA^+/DAPI^-$ sites. The genome sizes of *A. occidentale* and *A. humile* were estimated by flow cytometry, revealing relatively small sizes, with an average value of 0.44 pg/1C (435 Mbp) and 0.44 pg/1C (434 Mbp), respectively. These data suggest the absence of significant intra- or interspecific cytogenetic variation among the accessions, with no evidence of polyploidy

among the studied plants. Other markers should be employed to clarify the taxonomic boundaries and relationships between *A. occidentale*, *A. humile*, and sister species, as well as the possible correlation between agronomic variation and genetic structuring within the species.

Keywords: Anacardiaceae; agronomic characterization; phylogenetic analysis; CMA/DAPI; 5S and 35S rDNA; FISH.

SUMÁRIO

1. INTRODUÇÃO

O Cerrado brasileiro apresenta ampla extensão territorial, ocupando cerca de 24% do território nacional (COLLI et al., 2020). Esse bioma é o segundo maior domínio fitogeográfico do país, sendo considerado um *hotspot* de conservação da biodiversidade mundial por abrigar biota rica e endêmica (CEPF, 2023; IUCN, 2024). Todavia, vem sendo extremamente afetado pela devastação da vegetação original para o plantio de *commodities* (como soja, algodão, milho, sorgo e girassol) (INPE, 2023).

A flora desta região é composta de uma grande variedade de frutíferas nativas que têm despertado o interesse da indústria nacional, devido às suas características nutricionais e variedades de usos (REIS, SCHMIELE; 2019). Dessas podemos destacar o pequizeiro (*Caryocar brasiliense* Camb.), a mangabeira (*Hancornia speciosa* Gomes), o buritizeiro (*Mauritia flexuosa* L.f.) e o cajuizeiro (*Anacardium humile* A.St.-Hil.) (LAMBERS et al., 2020; SILVA-JÚNIOR et al., 2021). Apesar de sua relevância biológica inquestionável, o conhecimento sobre a diversidade genética da flora que compõe o Cerrado ainda é limitado.

Anacardium humile A. St. -Hil., é conhecido popularmente como cajuizeiro, cajuí ou cajuzinho-do-Cerrado, enquanto *A. occidentale* L., o cajueiro, é uma espécie cultivada presente em diferentes regiões do Brasil. Essas espécies pertencem à família Anacardiaceae, composta por ~81 gêneros e cerca de 800 espécies com distribuição tropical e subtropical (MITCHELL et al., 2022). Essas frutíferas desempenham um papel de grande relevância em termos nutricionais, ecológicos, econômicos e sociais, especialmente para as populações locais. Contudo, apesar de apresentar todos os potenciais supramencionados, o uso inapropriado dos ecossistemas para exploração predatória pode comprometer a variabilidade genética existente destas frutíferas (CRESPO; SOUZA, 2014).

Essas anacardiáceas apresentam uma alta variabilidade morfológica, sendo possível encontrar espécimes com pedúnculos com formatos que variam entre arredondados, piriformes ou achatados, com coloração variando de amarelo-claro a vermelho-intenso (CRESPO; SOUZA, 2014; CARNEIRO et al.,2019). A espécie *A. humile* ainda não é considerada uma planta cultivada, o que alerta para riscos na diminuição de sua diversidade genética e morfológica, especialmente quando considerada a ameaça ao seu bioma de ocorrência, o Cerrado (SILVA-JÚNIOR et al., 2021). Dessa

forma, pesquisas envolvendo a avaliação e caracterização do germoplasma de espécies do Cerrado, como *A. humile* e *A. occidentale*, são necessárias, a fim de avaliar a diversidade genética de seus representantes com vista a promover informações que subsidiem programas de conservação e melhoramento genético, bem como auxiliar no manejo racional e sustentável de populações naturais (AMABILE et al., 2018).

Nesse contexto, para que a variabilidade genética de uma população seja avaliada, é necessário caracterizar os acessos por meio de diversos métodos que contemplem características morfológicas, agronômicas, moleculares e bioquímicas de natureza qualitativa ou quantitativa (BORÉM et al., 2021). Assim, já foram realizadas pesquisas que avaliam a diversidade genética para algumas populações de caju (*A. occidentale*), explorando a diversidade genética por meio de avaliações morfológicas, tanto qualitativas quanto quantitativas, demonstrando a existência de uma notável diversidade genética entre os diversos acessos de caju (RUFINO et al., 2008; CHIPOJOLA et al., 2009; CARNEIRO et al., 2019).

Adicionalmente, diversos marcadores moleculares dominantes e codominantes foram utilizados para acessar a diversidade genética de diferentes espécies de *Anacardium*. Em um estudo referente a diversidade genética de duas populações de *Anacardium* spp. do Piauí e alguns acessos de *A. humile*, *A. microcarpum* Ducke, *A. othonianum* Rizzini, *A. occidentale* e *A. giganteum* W. Hancock ex Engl., com a utilização do marcador dominante ISSR (*Inter Simple Sequence Repeats*), foi observada uma alta diversidade genética entre os acessos avaliados, mas sem separação entre as populações ou espécies investigadas (BORGES et al., 2018). Populações naturais de *A. humile* localizadas no norte de Minas Gerais foram avaliadas via microssatélites, também denominado de SSR (*Simple Sequence Repeats*), revelando uma alta diversidade genética, mas com baixa estruturação e menor diversidade em populações localizadas em áreas mais degradadas e exploradas (COTA et al., 2017). Já analisando-se via marcador microssatélite acessos de cajueiro coletados em diferentes regiões da Índia, foi possível identificar uma variabilidade genética moderada entre os acessos (SAVADI et al., 2020).

Diante do exposto, os marcadores moleculares possibilitam acessar a variabilidade genética, sendo úteis em programas de melhoramento genético, além de fornecer o panorama para desenvolvimento de medidas de conservação (BORÉM et al., 2021). Atualmente, os dados de sequenciamento de regiões dos genomas de várias espécies vegetais têm possibilitado o desenvolvimento de estudos filogenéticos e filogeográficos

capazes de evidenciar a diversificação dos organismos no tempo evolutivo e compreender a sua história evolutiva (KADLEC et al., 2017).

Em plantas, as regiões mais comuns em estudos filogenéticos e filogeográficos são as sequências plastidiais, como por exemplo, *rbcL*, *ndhF*, *rps16*, *matK* e regiões que não codificam genes (íntrons e espaçadores intergênicos). A principal vantagem da utilização de regiões do cloroplasto em estudos filogenéticos decorre do pequeno tamanho desse genoma, herança geralmente uniparental e uma ordem de genes e estrutura do DNA bem conservada entre as angiospermas (DANIELL et al., 2016). Além destas, são utilizadas sequências do genoma nuclear, como o espaçador interno transcrito (*ITS* - *Internal Transcribed Spacer*), região altamente variável, que apresenta herança biparental, com uma alta taxa de evolução, tornando-se vantajosa para as análises filogenéticas (TURCHETTO-ZOLET et al., 2017).

As primeiras relações de parentesco para as espécies do gênero *Anacardium* foram estabelecidas por Mitchell e Mori (1987), empregando caracteres morfológicos. Por meio desse estudo, foi possível separar clados distintos como *A. humile* e *A. occidentale*. Apenas cinco espécies do gênero *Anacardium* foram incluídos numa análise filogenética molecular mais recente, com um acesso cada, empregando três marcadores plastidiais (5' *trnK_UUU-matK*, *trnL-F*, e 3' *ndhF* (RABAH et al., 2017), embora análises mais amplas tenham sido realizadas em outros gêneros de Anacardiaceae, a exemplo de *Rhus* L., *Pistacia* L., *Schinus* L. e *Spondias* L. (YI et al., 2004; XIE et al., 2014; SILVA-LUZ, 2017; NOBRE et al., 2018; ARIYARATHNE et al., 2020). Esses estudos utilizaram regiões do DNA ribossomal nuclear (*ITS* e *ETS*- *External Transcribed Spacer*) e regiões do cloroplasto (*matK, rps16, rpl16, ndhF*, *trnL-F, psaA-ycf3, atpB-rbcL* e *psbA-trnH*) gerando filogenias bem resolvidas, com clados bem suportados, sendo possível determinar as relações entre espécies pertencentes a cada gênero e separá-los em grupos geneticamente distintos, assim como identificar a origem de híbridos, revelando que estas regiões testadas foram eficazes e informativas.

Além das avaliações agronômicas e moleculares, as caracterizações citogenéticas também fornecem informações úteis para programas de pré-melhoramento e melhoramento genético, auxiliando na determinação de números cromossômicos, nível de ploidia, tamanho do genoma, construções de mapas físicos e identificação de híbridos (BRAMMER; JUNIOR, 2022). Para o gênero *Anacardium,* o conhecimento citogenético tem se limitado basicamente a contagens cromossômicas, utilizando a técnica de coloração convencional, e informações sobre o tamanho do genoma, restringindo-se

apenas à espécie comercial (*A. occidentale*). O número cromossômico reportado para o cajueiro é 2*n* = 40 (GILL; SINGHAL, 1979; GILL et al., 1990; PEDROSA et al., 1999), embora haja relatos de 2*n* = 30 e 42 (MACHADO, 1944; DARLINGTON; JANAKI-AMMAAL, 1945; ALIYU; AWOPETU, 2007), com tamanho médio do genoma de 0.85 pg/2C, sem indício de poliploidia (ALIYU, 2014).

Assim, tendo em vista o que foi apresentado, o presente estudo tem como objetivo testar as seguintes hipóteses: (I) a variabilidade morfológica apresentada por *A. occidentale* e *A. humile* está correlacionada com a distribuição geográfica e estruturação genética dos acessos; (II) existe variação cariotípica entre os acessos de *A. occidentale* e *A. humile*; (III) *A. occidentale* e *A. humile* são espécies distintas do gênero *Anacardium*. Para isso, investigamos a diversidade genética e estruturação de *A. occidentale* e *A. humile* e espécies relacionadas, por meio de análises agronômicas, moleculares e citogenéticas para entender a evolução e diversificação do grupo.

2. REVISÃO DE LITERATURA

2.1. Caracterização agronômica em estudos de diversidade genética

A diversidade genética se refere a toda variação biológica herdada durante o processo evolutivo, em grande parte ocasionada pelo processo de recombinação do material genético (DNA) ou por mutações nas sequências nucleotídicas (SALGOTRA; CHAUHAN, 2023), resultando em variações nos genótipos. Tais variações genotípicas, contidas no genoma de cada espécie, podem ser repassadas de uma geração para outra, podendo se expressar ou não. A manifestação dessa variação genotípica constitui o fenótipo, que pode ser evidenciado, por exemplo, por meio de características morfológicas. Assim, a variação fenotípica constitui o reflexo de alguma variação genotípica, embora também seja influenciada pelo ambiente. Logo, a maioria das pesquisas realizadas com diversidade morfológica também se refere à diversidade genética, pressupondo que os caracteres expressos podem refletir a variação do genótipo (ELLEGREN; GALTIE, 2016).

Os estudos sobre a diversidade genética têm como objetivo conhecer e quantificar o nível de variabilidade existente e sua distribuição, quer seja entre linhagens, cultivares e populações (CRUZ et al., 2011). Geralmente, a quantificação desta diversidade genética é feita por meio de avaliações moleculares, utilizando marcadores moleculares, tais como SSR (*Simple Sequence Repeats*), AFLP (*Amplified Fragment Length Polymorphism*) ou RAPD (*Random Amplified Polymorphic DNA*), que não sofrem a influência ambiental. No entanto, a diversidade genética de diferentes espécies também é avaliada por meio de características morfológicas, em especial as características agronômicas de interesse comercial (FU, 2015).

Nesse sentido, a caracterização de germoplasmas realizada por meio de descritores agronômicos visa caracterizar espécimes com base em uma ampla gama de variáveis, tais como, no caso de fruteiras, peso dos frutos, espessura da polpa, número de sementes por fruto, cor dos frutos, entre outras. Através da aplicação de métodos estatísticos, torna-se possível estimar a diversidade genética e agronômica presente em uma população (GOVINDARAJ et al., 2015). Logo, é possível determinar a distância genética entre genótipos, utilizando métodos biométricos que se baseiam em estatísticas multivariadas. Diversos métodos estão disponíveis para investigar a diversidade genética e agronômica, como as análises multivariadas que envolvem medidas de dissimilaridade, incluindo a distância de Mahalanobis, a técnicas de agrupamento como o método de

Scott-Knott e UPGMA (*Unweighted Pair-Group Method using Arithmetic Avarages*), bem como análises de dispersão gráfica que incluem análise de componentes principais e análise de variáveis canônicas (CRUZ et al., 2011).

No entanto, é importante ressaltar que, em alguns casos, a variabilidade fenotípica detectada por meio de caracterizações agronômicas e morfológicas realizadas por análises estatísticas pode não refletir diretamente a variabilidade genética, uma vez que os caracteres morfológicos podem apresentar plasticidade fenotípica e influência ambiental. Portanto, a avaliação da diversidade genética em diferentes populações pode ser complementada e correlacionada com ferramentas moleculares ou genômicas (GOVINDARAJ et al., 2015).

2.2. Utilização das regiões plastidiais e nuclear em análises filogenéticas

A filogenética é a ciência que busca reconstruir a história evolutiva da vida na Terra por meio da construção relações genéticas entre os diferentes organismos. Inicialmente as filogenias foram construídas utilizando apenas informações morfológicas. Entretanto, a partir da introdução dos avanços no sequenciamento de DNA, diversos métodos filogenéticos foram e ainda são extensamente empregados para inferir as relações evolutivas em diversos níveis taxonômicos (KAPLI et al., 2020). Com isso, uma variedade de regiões genômicas têm sido utilizadas para esse propósito, oferecendo uma alternativa confiável e independente das características morfológicas. Isso possibilita lidar de forma precisa com os desafios taxonômicos decorrentes das diferenças na interpretação das variações morfológicas, visto que as análises filogenéticas utilizando regiões do DNA e marcadores moleculares que, por sua vez, não sofrem interferência ambiental (WU et al., 2020).

Nesse contexto, o sequenciamento de genes ou regiões específicas do DNA se tornou uma prática comum em estudos filogenéticos. Inicialmente, os estudos nesse campo se concentraram na análise de sequências como *rbcL, ndhF, atpB, matK* e regiões não codificantes do genoma do cloroplasto, incluindo íntrons e espaçadores intergênicos, para investigar diversas questões, inclusive filogeográficas (TURCHETTO-ZOLET et al., 2017). As regiões não codificantes em particular, são altamente valiosas em estudos de sistemática filogenética em níveis infragenéricos devido à sua maior variabilidade em comparação com as regiões gênicas. No entanto, os avanços das tecnologias de sequenciamento permitem agora a utilização de dados em escala genômica, possibilitando

a análise de milhares de *loci* de genomas específicos. Isso viabiliza a construção de árvores filogenéticas baseadas, por exemplo, em sequenciamento de plastomas, área conhecida como filogenômica (YOUNG; GILLUNG, 2019).

Apesar dos avanços da filogenômica, as análises utilizando *loci* universais ainda são comuns e possibilitam estabelecer relações filogenéticas entre diferentes grupos taxonômicos e investigar processos evolutivos relacionados (KADLEC et al., 2017; YU et al., 2018). Tanto as regiões nucleares (*ITS - Internal Transcribed Spacer*) quanto as plastidiais (*matK, rps16, rbcL*, entre outros) são amplamente utilizadas devido à facilidade de amplificação por *primers* universais, permitindo sua aplicação em diferentes grupos taxonômicos (GUO et al., 2023).

A utilização de regiões nucleares e plastidiais em estudos que inferem as relações filogenéticas entre espécies e gêneros é vantajosa devido a uma série de características distintas. Por exemplo, o pequeno tamanho do genoma e a estrutura do DNA conservada são atributos notáveis do genoma plastidial (DANIELL et al., 2016). Por outro lado, a alta taxa de evolução é uma característica relevante do genoma nuclear (TURCHETTO-ZOLET et al., 2017). Portanto, essas regiões já foram aplicadas em diversos estudos, como os realizados em *Epidendrum* (Laeliinae; Orchidaceae), *Myrcia* e *Eugenia* (Myrtaceae), que empregaram uma variedade de regiões, tais como *ITS, ETS, atpl_atpH, matk, rpl32_trnL, rps16-trnK, trnL_trnF, trnQ_rps16, psbA-trnH, ndhF* e *rpl16.* Esses estudos resultaram em filogenias bem resolvidas, com clados bem suportados, embora alguns clados tenham apresentado suporte mais baixo (AMORIM et al., 2019; GIARETTA et al., 2019; PESSOA et al., 2021).

2.3. Citogenética no estudo evolutivo de plantas

2.3.1. Características cariotípicas

Os cromossomos metafásicos representam um dos principais alvos de estudos citogenéticos. Cada cromossomo metafásico é constituído por duas cromátides-irmãs, sendo cada uma delas formada por uma única fita de dupla de DNA associada a uma variedade de proteínas histônicas e não histônicas. Essas fitas de DNA passam por estágios de condensação até atingirem o nível máximo de compactação, caracterizando assim os cromossomos metafásicos (GUERRA, 1988; PAULSON et al., 2021; KUBALOVÁ et al., 2023).

Estes cromossomos possuem regiões de grande interesse para estudos citogenéticos, que incluem a constrição primária, conhecida como centrômero, a constrição secundária, geralmente associada a regão organizadora do nucléolo, as regiões satélites e os telômeros. A morfologia desses cromossomos é definida pela posição do centrômero, dividindo-o em braços cromossômicos. Dependendo da razão entre os braços, eles podem ser classificados em quatro tipos distintos: metacêntrico, submetacêntrico, acrocêntrico e telocêntrico (GUERRA, 1986; STIMPSON; SULLIVAN, 2013).

Nesse sentido, grande parte das pesquisas em citogenética se concentra na análise detalhada do cariótipo, especialmente quanto à morfologia, número e tamanho dos cromossomos, simetria cariotípica, distribuição da heterocromatina e eucromatina, quantidade e localização do DNA ribossomal (DNAr), bem como o tamanho do genoma (LEVIN, 2002; HE et al., 2020). Tais abordagens são essenciais para a compreensão da história evolutiva e da diversificação de distintos grupos taxonômicos (GUERRA, 2012).

2.3.2. Heterocromatina e coloração com fluorocromos

A heterocromatina é um constituinte do genoma eucarioto que se encontra frequentemente mais compactada em comparação com a eucromatina, desempenhando uma variedade de funções como, por exemplo, o silenciamento de genes que se encontram muito próximo de uma região enriquecida por esse tipo de cromatina (ALLSHIRE; MADHANI, 2018). Essa heterocromatina pode ser subdividida em duas categorias principais: a heterocromatina constitutiva, compactada e composta por sequências de DNA repetitivo geralmente em tandem, formando blocos visíveis nos cromossomos quando corados com corantes específicos; e a heterocromatina facultativa, que exibe ciclos alternados em que a cromatina pode estar transcricionalmente ativa ou inativa (MORRISON; THAKUR, 2021). A heterocromatina constitutiva está distribuída preferencialmente nas regiões proximais, terminais e entorno da constrição secundária dos cromossomos. É o local onde se situa grande parte do DNA satélite, mas que não é composto exclusivamente por esse tipo de sequência (PENAGOS-PUIG; FURLAN-MAGARIL, 2020).

Em plantas, a heterocromatina constitutiva é investigada principalmente por meio do bandeamento C e da dupla coloração com os fluorocromos cromomicina A3 (CMA) e 4'6-diamidino-2-fenilindol (DAPI), que evidenciam as regiões ricas, respectivamente, em

GC (guanina e citosina) e AT (adenina e timina), corando diferencialmente os cromossomos (BARROS-SILVA; GUERRA, 2023). A utilização desses e de outros fluorocromos têm desempenhado um papel importantíssimo nas análises cariotípicas de vários representantes da biodiversidade mundial.

Análises citogenéticas utilizando estes fluorocromos permitem caracterizar de forma mais detalhada os cariótipos. Por exemplo, em representantes da subtribo Laeliinae (Orchidaceae), as espécies analisadas no estudo apresentaram um padrão de bandas CMA^+ observadas nas regiões pericentroméricas e bandas $DAPI^+$ presentes nas regiões terminais dos cromossomos (QUERINO et al., 2020). Em algumas espécies de Bignoniaceae foram encontradas bandas de CMA^+ localizadas nas regiões proximais e terminais dos cromossomos (CORDEIRO et al., 2020). A diferenciação cariotípica confiável de espécies cítricas foi observada pela primeira vez após dupla coloração com os fluorocromos CMA e DAPI. Após a dupla coloração, os cromossomos de *Citrus aurantium* L. (laranja azeda) exibiram muitas bandas CMA^+ brilhantes, localizadas principalmente na região terminal dos cromossomos, permitindo identificar fórmulas cariotípicas distintas entre os acessos destas cultivares (GUERRA et al., 2020).

2.3.3. Hibridização *in situ* fluorescente

A hibridização *in situ* fluorescente (FISH) é uma das técnicas mais importantes dentro da citogenética. Sua aplicação em espécies vegetais tem possibilitado a localização de várias sequências de DNA e a identificação dos pares cromossômicos (BADAEVA et al., 2017). Nesta técnica, sequências de DNA ou RNA marcadas com fluoróforos são utilizadas como sondas e hibridizadas *in situ*, revelando o sítio exato onde a sequência relacionada ocorre no cromossomo (GUERRA, 2004; VESELINYOVÁ et al., 2021).

Diversos tipos de sondas podem ser utilizados como marcadores cromossômicos, voltando-se, especialmente, na detecção de sequências repetitivas organizadas em tandem, como o DNA satélite e o DNAr, além de sondas genômicas e cromossômicas (GUERRA, 2004). Embora as sondas de DNAr 5S e 35S ainda mantenham sua relevância, a FISH tem experimentado um avanço notável com o desenvolvimento e aplicação de sondas baseadas em oligonucleotídeos sintéticos (*oligos*) (JIANG, 2019).

A FISH, quando associada com outras técnicas citogenéticas, como a coloração cromossômica diferencial com fluorocromos CMA e DAPI, têm contribuído para a caracterização dos cariótipos quanto à constituição e distribuição da heterocromatina,

trazendo luz aos prováveis eventos envolvidos na evolução cariotípica de espécies (GUERRA, 2000; ORTIZ, 2017). A eficácia dessas técnicas tem sido comprovada em grupos como *Caesalpinia*, em que diferenças no padrão de bandas CMA^+, com eventos de amplificação e eliminação foram observadas entre as espécies, assim como a presença de sítios de DNAr 5S e 35S, com todos os sítios 35S colocalizados com os blocos de CMA^+ (VAN-LUME et al., 2017).

2.3.4. O DNAr 5S e 35S

Os ribossomos são organelas celulares consideradas fábricas proteicas, uma vez que desempenham um papel fundamental no processo de tradução dos RNAs mensageiros e, portanto, na síntese de proteínas (NEVES et al., 2005). No contexto das plantas, o DNA ribossomal (DNAr) 35S é uma unidade repetitiva constituída pelos genes ribossômicos 18S, 5.8S e 25S. Esses genes estão estruturalmente separados por dois espaçadores internos, denominados ITS1 e ITS2, além de uma sequência intergênica mais longa, o IGS (*Intergenic spacers*), sendo transcritos em um único RNA (ROA; GUERRA, 2015; STEPANENKO et al., 2022).

Por outro lado, o DNAr 5S consiste em uma unidade de repetição mais simples que constitui um locus independente. Tanto o DNAr 35S quanto o DNAr 5S são constituídos por sequências de nucleotídeos repetidas em tandem em uma ou várias regiões do genoma, formando sítios nos cromossomos. É importante destacar que essas sequências são evolutivamente conservadas e similares entre os eucariotos, no caso do DNAr 35S, ou entre a maioria das angiospermas, no caso do DNAr 5S (GUERRA, 2004).

A distribuição dos sítios de DNAr ocorre preferencialmente nos braços curtos e nas regiões subterminais e terminais dos cromossomos. Estes sítios podem variar tanto em posição ao longo do cromossomo quanto em tamanho e quantidade, permitindo que sejam determinadas variações intragenéricas e até intraespecíficas (ROA; GUERRA, 2012). Em espécies próximas do gênero *Macroptilium* (Benth) (Fabaceae), os sítios de DNAr podem variar extensamente entre acessos nativos e domesticados (DE BARROS et al., 2023). Estudos realizados com cultivares de *Citrus* L. (Rutaceae) identificaram sítios de DNAr 35S localizados nas regiões proximais e terminais de seus cromossomos, coincidindo com algumas bandas de CMA^+ (GUERRA et al., 2020; MONTENEGRO et al., 2023).

Além da variação no número de sítios entre diferentes espécies, é igualmente comum observar a estabilidade no número de sítios de DNAr, embora possam ocorrer pequenas diferenças em sua localização. *Theobroma cacao* L. (cacau) e *T. grandiflorum* Schum (cupuaçu) (Malvaceae) apresentam cariótipos com um único sítio de DNAr 5S localizado na região proximal de um par cromossômico e um único sítio de DNAr 35S localizado na região terminal de um outro par cromossômico (DANTAS; GUERRA, 2010). Em espécies diploides de *Annona* L. (Annonaceae), observa-se também uma estabilidade de sítios de DNAr 5S e35S, localizados nas regiões dos braços curtos dos cromossomos (SANTOS et al., 2023).

Um estudo envolvendo seis espécies do gênero *Spondias* (Anacardiaceae) revelou a presença de apenas um sítio de DNAr 5S e 35S por conjunto cromossômico haploide. Nos cromossomos, os sítios de DNAr 35S estão sempre localizados na região terminal do braço curto, enquanto o sítio de DNAr 5S foi detectado na região intersticial em um outro cromossomo, com exceção de apenas uma espécie, *S. tuberosa* Arruda, que apresenta o sítio 5S na região terminal do braço curto (ALMEIDA et al., 2007). Essa uniformidade na localização dos sítios de DNAr sugere conservação estrutural dentro do gênero *Spondias*.

2.4. O gênero *Anacardium* L.

O gênero *Anacardium* L. pertence a tribo Anacardieae, da família Anacardiaceae, a qual é reconhecida por sua significativa relevância tanto econômica quanto ecológica, e inclui várias espécies arbóreas e frutíferas como o caju (*A. occidentale* L.), a manga (*Mangifera indica* L.), o pistache (*Pistacia vera* L.), a pimenta rosa (*Schinus terebinthifolia* Raddi), o umbu (*Spondias tuberosa* Arruda), o cajá (*Spondias mombin* L.) e a seriguela (*Spondias purpúrea* L). A família compreende ~81 gêneros e cerca de 800 espécies tropicais e subtropicais, estendendo-se até as regiões temperadas (MITCHELL et al., 2022).

O gênero *Anacardium* é composto por árvores, arbustos e subarbustos nativos dos neotrópicos. Apresentam flores reunidas em uma inflorescência do tipo panícula, composto com forma piramidal, com flores hermafroditas, com 7 - 10 estames, sendo um ou dois geralmente maiores (MITCHELL; MORI,1987; MITCHELL et al., 2022). Os agentes polinizadores das flores de *Anacardium* são as abelhas, principalmente a espécie *Apis mellifera* L. (TAKEHANA et al., 2013). Na maioria das espécies, a floração costuma

acontecer durante a estação seca, enquanto o ápice de frutificação ocorre entre os meses de agosto a dezembro. Com a fertilização das flores, desenvolve-se o fruto (castanha) e o pseudofruto ou pedúnculo (hipocarpo) carnoso, comestível e suculento. Esse pedúnculo é consumido por animais que fazem a dispersão das sementes, sendo os morcegos frugívoros e o ser humano seus principais agentes dispersores (MITCHELL; MORI,1987).

Dentre os representantes do gênero *Anacardium*, o cajueiro (*A. occidentale*) apresenta porte arbóreo, sendo a espécie de maior importância econômica (VIEIRA et al., 2006). Já o cajuizeiro *(A. humile*), é um subarbusto, tradicionalmente conhecido como cajuzinho, cajuzinho-do-cerrado ou cajuí, que se destaca como uma espécie endêmica da flora do Cerrado brasileiro (SILVA et al., 2004; CRESPO; SOUZA, 2014). Além de *A. humile*, *A. giganteum* e *A. spruceanum* também são conhecidas popularmente como cajuí, sendo as duas reportadas para a região Norte do Brasil (FLORA DO BRASIL, 2020).

O termo caju é oriundo da palavra indígena "*acaiu*", que na língua tupi significa "noz que se produz", da qual deriva a palavra cajuí (caju + tupi "*i"*, pequeno) (VIEIRA et al., 2006). O termo cajuí é bastante genérico e se refere ao pequeno tamanho da castanha (fruto verdadeiro) e pedúnculo (pseudofruto), que medem aproximadamente 1-2 × 1-2 cm e 1-3 × 1-2 cm, respectivamente (Figura 1; MITCHELL; MORI,1987).

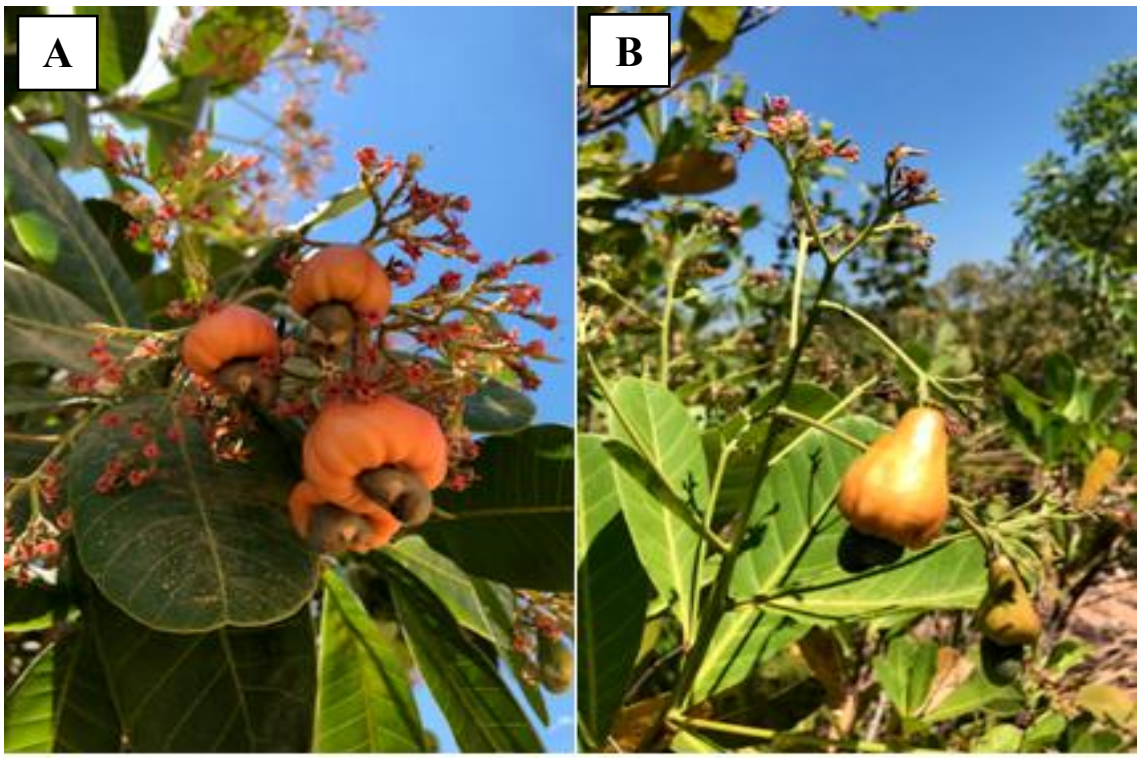


Figura 1 - Aspectos morfológicos de caju e cajuí. **(A)** Caju (*A. occidentale*), com

inflorescência, fruto e pseudofruto do acesso AN-714; **(B)** Cajuí (*A. humile*), fruto e pseudofruto do acesso AN-712. Fotos de Silvokleio da Costa Silva.

2.4.1. Origem e distribuição geográfica

As diversas evidências encontradas na literatura apontam que *Anacardium* é originário da região que inclui o Norte da América do Sul e parte da América Central (PAIVA et al., 2003). O gênero está distribuído naturalmente desde Honduras na América Central, até o Sul do Paraná e leste do Paraguai. Na Venezuela, Colômbia e Equador ocorrem apenas a oeste dos Andes (KEW SCIENCE, 2017; FLORA DO BRASIL, 2020). O Brasil é considerado o maior e mais importante centro de diversidade de *Anacardium*, com dois principais centros, um localizado na Amazônia e o outro na região do Cerrado (MITCHELL; MORI, 1987), sendo o Nordeste brasileiro o maior centro de diversidade da única espécie cultivada comercialmente, o cajueiro (PAIVA et al., 2003).

As espécies não cultivadas, com importância econômica restrita ao consumo local nas áreas produtoras, a exemplo do cajuizeiro, também são distribuídas na região Nordeste do Brasil, especialmente nas áreas de Cerrado, como no Piauí (RUFINO et al., 2007). Essas frutíferas apresentam uma boa adaptação ao ambiente e crescem tanto em solos ricos ou pobres em nutrientes, ocorrendo em áreas de ecótonos Cerrado-Caatinga e de restingas e carrasco (restingas densas), ecossistemas que cobrem as dunas (Figura 2; MITCHELL; MORI, 1987).

Figura 2 - Habitats naturais do cajueiro (*A. occidentale*). (**A**) Acesso AN-800; (**B**) AN-801, localizados no município de Parnaíba, PI. Fotos de Silvokleio da Costa Silva.

2.4.2. *Anacardium* e sua importância estratégica no Brasil e no mundo

As espécies pertencentes ao gênero *Anacardium* apresentam importância ecológica, social e econômica em diversos países (SCHWEIGGERT et al., 2016). Dentre os representantes do gênero *Anacardium*, o cajueiro (*A. occidentale*) é a espécie de maior importância econômica, tendo seus frutos (castanhas) e pseudofrutos (pedúnculos) explorados pela agroindústria (VIEIRA et al., 2006).

Nesse sentido, a Costa do Marfim se destaca como o principal produtor mundial de castanha de caju, com 848,7 mil toneladas em 2020. Já o Brasil se classificou na 11ª colocação, com uma produção de 139,9 mil toneladas no mesmo ano (BRAINER, 2022). No Brasil, o principal estado produtor de castanha de caju é o Ceará, seguido pelos estados do Piauí e Rio Grande do Norte. Logo, o Nordeste brasileiro é responsável por quase toda a produção nacional de castanhas de caju e por toda a exportação nacional (BRAINER, 2022; CONAB, 2023).

Apesar de não ocupar as primeiras posições quanto a produção de castanhas, o Brasil se destaca como sendo o principal e maior produtor mundial de pedúnculos (BRAINER, 2022). O país também é considerado o maior consumidor de derivados do pedúnculo tais como sucos, cajuínas e doces (OLIVEIRA, 2018). Cabe destacar também que os frutos e pseudofrutos das espécies de *Anacardium*, inclusive de *A. occidentale* e *A. humile*, são ricos em vitaminas (C e A) e minerais como cálcio, ferro e fósforo, contribuindo para um aumento na procura desses materiais para consumo *in natura* (SILVA et al., 2001).

O cajuizeiro *(A. humile*) também apresenta frutos comestíveis que são explorados apenas de maneira extrativista (SILVA et al., 2004; CRESPO; SOUZA, 2014), uma vez que a castanha do cajuí não é aceita pela grande indústria de beneficiamento da castanha do caju devido ao seu pequeno tamanho. Mesmo assim, após ser torrada, sua amêndoa pode ser consumida sozinha ou ser adicionada como ingrediente na composição de doces, tortas e bolos, produtos estes consumidos pelas famílias extrativistas de regiões de ocorrência desta frutífera (RUFINO et al., 2007).

O seu pedúnculo é bastante apreciado por causa de seu sabor mais adocicado quando comparado ao caju, podendo ser consumido *in natura* ou empregado na fabricação de doces, suco, picolé, cajuína, proteína vegetal (carne) e bebidas alcoólicas (VIEIRA et al., 2006; LIMA, 2008; MONTEIRO; JUNQUEIRA, 2018; CARNEIRO et al., 2019). A aceitação dos produtos obtidos e a comercialização do fruto e pseudofruto

em feiras livres faz com que o cajuizeiro seja foco de exploração (RUFINO et al., 2007; GOMES, et al., 2011; CRESPO; SOUZA, 2014).

2.4.3. Classificação botânica e relações filogenéticas

O gênero *Anacardium* é constituído por um pequeno número de espécies, todas originárias das regiões das Américas Central e do Sul e um maior número de representantes do gênero situados na Amazônia e no Cerrado. De acordo com a taxonomia clássica, o gênero é composto por 21 espécies (JOHNSON,1973). Todavia, Mitchell e Mori (1987), por meio da taxonomia numérica, reduziram o número de espécies para 10. Entretanto, a taxonomia do gênero ainda é objeto de debate, com proposta de 15 (CUNHA, 2002) e 20 espécies (POWO, 2023).

Considerando as 10 espécies aceitas para *Anacardium* por Mitchell e Mori, (1987) e Flora do Brasil (2020), nove ocorrem no Brasil: i) *A. amapaense* J. D. Mitch.; ii) *A. corymbosum* Barb. Rodr. (cajuzinho); iii) *A. giganteum* W. Hancock ex Engl. (cajuí, caju-açu); iv) *A. humile* A. St.-Hil. (cajuí, cajuzinho, caju-do-cerrado, caju-mirim, caju-anão, caju-rasteiro, cajuzinho-do-campo, caju-do-campo, cajuhy, cajuzinho-do-cerrado); v) *A. microsepalum* Loes. (cajuí-da-várzea, cajurana); vi) *A. nanum* A. St.-Hil. (cajuzinho); vii) *A. occidentale* L. (caju, caju-anão ou acajaiba); viii) *A. parvifolium* Ducke; e ix) *A. spruceanum* Beth.ex Engl. (cajuí, caju-assu, cajueiro-do-mato). Apenas *A. excelsum* (Bertero & Balb. ex Kunth) Skeels não ocorre no Brasil, estando presente na Venezuela e de Honduras ao Equador. Além de *A. humile*, *A. giganteum* e *A. spruceanum* também são conhecidas popularmente como cajuí, sendo *A. humile* de ampla distribuição no Centro-Oeste, Sudeste e Nordeste, e as outras duas reportadas para a região Norte do Brasil apenas. Assim, o termo popular "cajuí" não é válido para distinguir as espécies, visto que se refere a várias espécies que possuem frutos e pseudofrutos pequenos (VIEIRA et al., 2014).

As primeiras relações de parentesco para as espécies do gênero *Anacardium* foram estabelecidas empregando uma análise cladística por meio de caracteres morfológicos como habitat, base foliar, textura foliar, inflorescência e outros caracteres botânicos, evidenciando que *A. humile* e *A. occidentale* ocupam clados distintos. Além disso, as espécies *A. othonianum* e *A. microcarpum* foram identificadas como sinônimas de *A. occidentale*, uma vez que são espécies próximas e muito parecidas morfologicamente (MITCHELL; MORI, 1987). Mais recentemente, foi realizado o sequenciamento

completo do plastoma de *A. occidentale* e de regiões marcadoras *trnK-matK*, *trnL_trnF* e *ndhF* de *A. nanum, A. humile, A. corymbosum* e *A. excelsum*, para investigar a organização do genoma plastidial e a origem da inserção de um segmento do genoma mitocondrial no plastoma. Embora a amostragem seja limitada, *A. nanum* e *A. humile* se mostraram como espécies irmãs, com *A. occidentale* irmã deste clado e *A. excelsum* a primeira a divergir e única que não compartilhou esta inserção, que provavelmente ocorreu há menos de 20 milhões de anos atrás (RABAH et al., 2017).

Entretanto, ainda não existe na literatura um estudo filogenético mais amplo para o gênero *Anacardium*, apesar de já terem sido realizadas análises filogenéticas para outros gêneros pertencentes à família Anacardiaceae, como *Rhus* L., *Pistacia* L., *Schinus* L. e *Spondias* L. (YI et al., 2004; XIE et al., 2014; SILVA-LUZ, 2017; NOBRE et al., 2018; ARIYARATHNE et al., 2020). Estes estudos utilizaram regiões nucleares e plastidiais e conseguiram gerar filogenias bem resolvidas, sendo possível determinar as relações entre espécies pertencentes a cada gênero, revelando que estas regiões testadas foram informativas para os gêneros mencionados.

2.4.4. Estudos agronômicos e citogenéticos no gênero *Anacardium* L.

Diversos estudos objetivaram comparar a diversidade genética de espécies do gênero *Anacardium* por meio de descritores morfológicos e agronômicos. Com foco na caracterização da diversidade de ecótipos de caju cultivados na Costa do Marfim, uma notável variabilidade morfológica e agronômica entre esses ecótipos foi observada, confirmando a hipótese de alta diversidade genética, com a formação de agrupamentos de procedências distintas e distantes umas das outras (KOUAKOU et al., 2018). Variações fenotípicas significativas também foram registradas entre os acessos cultivados de caju localizados em Benin, na África ocidental (DJOLOSSÈ et al., 2019).

Outros estudos foram conduzidos no Brasil para avaliar a diversidade genética de populações naturais e cultivadas de caju no Piauí, utilizando descritores morfológicos e agronômicos. Foram incluídos caracteres relativos às folhas, inflorescências, frutos e pseudofrutos, sendo possível observar diferenciação fenotípica e a presença de diversidade genética para as populações avaliadas (CARNEIRO et al., 2019; COSTA et al., 2020; MATOS-FILHO et al., 2023). Em um estudo que avaliou a diversidade genética com características físico-químicas de frutos e pseudofrutos de cajuzinho-do-cerrado (*A. humile*) da coleção biológica *ex situ* localizada no estado de Goiás, foi

confirmada a existência de variabilidade genética dentro das populações analisadas (PEREIRA et al., 2019). Da mesma forma, constatou-se variabilidade genética tanto para variáveis morfométricas como para multicategóricas ao avaliar a diversidade fenotípica e morfológica de populações naturais de *A. humile* localizadas em área do Cerrado em Minas Gerais (SANTOS; SANTOS-JÚNIOR, 2015). Também foi registrada variação fenotípica e divergência genética entre matrizes de *A. othonianum*, avaliadas por meio de características físico-químicas dos frutos e pseudofrutos (OLIVEIRA, 2018).

No que diz respeito aos conhecimentos citogenéticos relacionados ao gênero *Anacardium*, até o momento foram publicadas apenas contagens cromossômicas para o caju (*A. occidentale)*, principal representante comercial do grupo. O número cromossômico amplamente documentado para o cajueiro é de $2n = 40$ (GILL; SINGHAL, 1979; GILL et al., 1990; PEDROSA et al., 1999), embora tenha havido também relatos de $2n = 30$ e 42 (MACHADO, 1944; DARLINGTON; JANAKI-AMMAAL, 1945; ALIYU; AWOPETU, 2007). Até o momento, não há dados sobre números cromossômicos, padrões de distribuição de bandas heterocromáticas, número e distribuição dos sítios de DNAr 5S e 35S para outras espécies do gênero *Anacardium* em publicações científicas indexadas.

As informações sobre o tamanho do genoma no gênero se restringem à espécie *A. occidentale*, cujo tamanho médio do genoma é relatado como 0.85 pg/2C, sem indícios de poliploidia (ALIYU, 2014). Portanto, é crucial conduzir análises intra- e interespecíficas no gênero, a fim de investigar possíveis variações significativas em níveis agronômico, molecular e citogenético.

3. RESULTADOS

Artigo a ser submetido à revista Tree Genetics & Genomes (A2)

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

Acalene Gonçalveis-Oliveira[1], Thiago Nascimento[1], Priscila Alves Barroso[2], Cíntia Luíza da Silva Luz[3], Silvokleio da Costa Silva[4], Andrea Pedrosa-Harand[1]*

[1]Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Biosciences Centre, Federal University of Pernambuco. Recife, PE, 50.670-901, Brazil.

[2]Laboratory of Breeding and Data Analysis, Professor Cinobelina Elvas Campus, Federal University of Piauí. Bom Jesus, PI, 64.900-000, Brazil.

[3]Laboratory of Plant Taxonomy, Department of Plant Biology, Institute of Biology, State University of Campinas, Campinas, Brazil.

[4]Laboratory of Genetics and Germplasm Conservation, Professora Cinobelina Elvas Campus, Federal University of Piauí. Bom Jesus, PI, 64.900-000, Brazil.

*Corresponding author. Email: andrea.harand@ufpe.br

ORCIDs:

Acalene Gonçalveis-Oliveira	https://orcid.org/0000-0001-6720-3134
Thiago Nascimento	https://orcid.org/0000-0002-7742-0260
Priscila Alves Barroso	https://orcid.org/0000-0003-1749-1961
Cíntia Luíza da Silva Luz	https://orcid.org/0000-0002-7144-5109
Silvokleio da Costa Silva	https://orcid.org/0000-0002-9975-1212
Andrea Pedrosa-Harand	http://orcid.org/0000-0001-5213-4770

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 analysed the genetic diversity of their germplasms in Cerrado and restinga biome areas of 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. Moreover, a lack of molecular differentiation was observed across selected nuclear (*ITS*) and plastid (*matK*, *trnLF*, *ycf1* and *rps16*) DNA regions 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, however our analysis revealed a strong chromosomal stability, with $2n = 40$, $CMA^+/DAPI^-$ bands in the terminal

region of the short arm of three chromosome pairs and the presence of one pair of 5S rDNA and three pairs of 35S rDNA sites, the latter co-localized with CMA^{+}/$DAPI^{-}$ bands. Genome size were similar, with an average value of 0.44 pg/1C (435 Mbp), suggesting no significant intra- or interspecific variation among the accessions and the absence of polyploidy. The lack of differentiation among the accessions may result from incomplete lineage sorting, hybridization with introgression, or the existence of a single species with domesticated accessions expressing larger fruits and pseudofruits.

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

Introduction

Sustainable management of plant resources is intrinsically linked to the preservation of their genetic diversity. One of the most important tasks to achieve this goal is the identification of populations or genotypes that represent the existing variability within a species, using different methods. These methods 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 for 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 analyses (Zuffo et al. 2016; Carneiro et al. 2019; França et al. 2020).

However, the phenotypic variability detected through morphological characterization may not reflect genetic diversity, as morphological traits may exhibit phenotypic plasticity and be influenced by the environment (Govindaraj et al. 2015). Thus, genetic diversity of different populations can be more directly assessed by molecular tools, such as nuclear markers - for example, 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 studies, cytogenetic analyses are also useful for species characterization, providing information related to their complete karyotype (Costa et al. 2017).

In this context, the Anacardiaceae family comprises around 81 genera and approximately 800 species with tropical and subtropical distribution. The genus *Anacardium* L. includes ten accepted species, with *Anacardium excelsum* (Bertero & Balb. ex Kunth) Skeels being the only one that does not occur in Brazil (Mitchell et al. 2022; Powo 2024). *Anacardium occidentale* L. (cashew tree), a tree that can range 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 tree) and *Pistacia vera* (pistachio), with its fruits (cashews) and pseudofruits (peduncles) exploited by the agribusiness industry (Carneiro et al. 2019). Meanwhile, *Anacardium othonianum* Rizzini is classified as 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 shrub ranging from 0.3 to 1.5 meters in height, standing out as an endemic species with edible fruits and pseudofruits that are also exploited, but exclusively through extractivism due to their generally smaller sizes. Its economic importance is limited to local consumption, primarily by rural communities in the Northeast of Brazil, in areas of the Cerrado biome, were both species cooccur (Crespo and Souza 2014).

Among the *Anacardium* species, only *A. occidentale* has been targeted for genetic improvement, aiming to obtain clones that produce high-quality peduncles and cashews (Garruti et al. 2022). For *A. humile*, there is still no formal breeding program established, although pre-breeding studies have already been conducted (Borges et al. 2018; Dos Santos et al. 2019; Pereira et al. 2019). Molecular phylogenetic studies in the *Anacardium* genus are limited to a few taxa (Rabah et al. 2017), which hinders a clear understanding of the phylogenetic relationships among its 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, generating well-supported clades, separating the analysed species into monophyletic groups, and identifying hybridization events (Yi et al. 2004; Xie et al. 2014; Nobre et al. 2018; Silva-Luz et al. 2019; Ariyarathne et al. 2020). The taxonomic classification of *Anacardium* is still controversial, and morphometric analyses, such as the one performed by Vieira et al. (2014), were unable to clearly distinguish species such as *A. microcarpum* and *A. occidentale* due to the large overlap of leaf data. Cytogenetic analyses are restricted to *A. occidentale*, with a reported chromosome number of $2n = 40$ (Gill and Singhal 1979; Gill et al. 1990; Pedrosa et al. 1999), although there are also reports of $2n = 30$ and 42 (Machado 1944; Darlington and Janaki-Ammaal 1945; Aliyu and Awopetu 2007), suggesting intra-specific numerical variation. Genome size has also been estimated for *A. occidentale* only, with an average value of 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* are distinct lineages within the genus; (II) the morphological variability present in *A. occidentale* and *A. humile* is correlated with their molecular differentiation; (III) there is karyotypic variation among accessions of *A. occidentale* and *A. humile*, and among *Anacardium* species, which can be successfully used for germplasm characterization. We also investigated the genetic diversity of *A. occidentale* and *A. humile* and their phylogenetic relationships with other *Anacardium* species, using morphological, molecular, and cytogenetic tools to understand the evolution and diversification of the genus.

Materials and Methods

Plant Material and Taxonomic Delimitation

Cashew nuts from 37 accessions were collected from natural populations of cashew and *cajuí* (Table S1), with the majority of the seeds collected from six different locations in Piauí state, then deposited into the Active Germplasm Bank of *Anacardium* Seeds at the Professora Cinobelina Elvas Campus of the Federal University of Piauí, Brazil. All the accessions were taxonomically evaluated, and most of them classified as *A. occidentale*, although they do not fit the size descriptions for fruit and pseudofruit. Another group of accessions was taxonomically identified as *A. humile*. Most 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. Additionally, there is significant overlap of flower and fruit characteristics, except for the longer peduncle observed in some *A. occidentale* accessions, which may result from the domestication process (Mitchell and Mori 1987). However, *A. occidentale* is often easily distinguishable by its tree-like form, while *A. humile* is a subshrub with a substantial 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 maintained in the Experimental Garden of the Laboratory of Cytogenetics and Plant Evolution, in Recife, Pernambuco. 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 not considered for agronomic characterization, since only a few seeds were supplied. For 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). For this, a digital caliper with a range of 150 mm and an accuracy of ± 0.02 mm / 100 mm was used for measurements (in mm), while weighing (g) of samples was performed using an analytical scale balance (Model AY220 - Tecnal).

The color of the fruits and peduncles was measured using 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 region, in the range of 400 to 700 nm, for all three sections.

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

Cluster Analysis and Correlations

After performing all measurements, the quantitative morphological data were used as input for analysis of variance (ANOVA) to check for differences between accessions, using R software version 4.0.5 (R Core Team 2022) with ExpDes package (Ferreira et al. 2021). A completely randomized design with 15 replicates (fruits and pseudofruits) was considered. The Scott-Knott test was performed to separate accessions means into homogeneous groups (Scott and Knott 1974).

Multivariate analysis of variance (MANOVA) was used to test the effects of the mean vectors associated with the accessions. Once this effect was verified, the Pillai's trace method test was used to check its significance. The accessions were clustered using the unweighted arithmetic means method - UPGMA. The clustering was based on the generalized Mahalanobis distance matrix, obtained from the quantitative characteristics evaluated in each of the accessions. The consistency of the clusters was assessed by the cophenetic correlation coefficient (Sokal and Rohlf 1962). The cluster analysis was based on distance matrices obtained using the cutoff point established by the Mojena method (1977). Singh (1981) criterion was used to quantify the relative contribution of the characteristics to genetic divergence. The analyses were performed with the aid of the Biotools package (Da Silva 2021).

The data obtained were also subjected to Principal Component Analysis (PCA), processed with 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 after stored at -20 °C for later 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 with a modification of the air-drying technique described by De Carvalho and Saraiva (1993), in which after preparation, the 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, then destained in ethanol: acetic acid (3:1, v/v) for 30 minutes at room temperature, and 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, with a Leica DFC345FX camera through the LAS AF software. At least 10 metaphases per accession were analysed, and the best images were uniformly adjusted for brightness and contrast using Adobe Photoshop (v.21.0.2). The slides were destained once more, as described above, for later use in Fluorescence *in situ* hybridization (FISH).

The FISH experiments were performed as described by Pedrosa et al. (2002). For the localization of 5S rDNA sites, a pool of pre-labelled oligonucleotide probes (PLOP, 5SrDNA_ang_1-4) with Cy3 was used (Waminal et al. 2018), and 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 then 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 analysed 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. Approximately 5,000 nuclei were quantified for each measurement, and the results were interpreted by analysing the graphs generated by FloMax software v. 2.3. Genome sizes in picograms (pg) were estimated for each accession according to the formula [(mean sample fluorescence/mean standard fluorescence) × standard genome size] (Doležel 2005). The three accessions with measurements showing the lowest coefficient of variation (CV) were used to calculate the average 1C content of *A. occidentale.* The variation 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 due to the unavailability of fresh leaves. DNA extraction followed the method of Ferreira and Grattapaglia (1998) and was subsequently quantified using a NanoDrop 2000 spectrophotometer (Thermo Scientific).

PCR Amplification and Sequencing

From the 26 individuals with extracted DNA, we successfully amplified three plastid regions (*matK*, *trnL-F* and *rps16*) 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*), we could amplify 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®). Amplification of the regions was carried out using the PCR programs described in Table S2. All 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 here 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 available in GenBank from different *A. occidentale* accessions (accession numbers SRX2990990) to its reference plastome (NC_035235) (Table S3) to assess the intra-specific polymorphism of some *loci* and determine which regions would be most informative for our phylogenetic analysis. As a result, the *ycf1* region was found to be the most polymorphic. Therefore, out of the 26 accessions with extracted DNA, we successfully amplified 24 accessions for the *ycf1* region using primers described by Dong et al. (2015), following the amplification cycle described in Table S2. Subsequently, all amplified PCR products were sequenced as described in the previous section.

Phylogenetic analyses

Sequences obtained from five regions of individuals belonging to *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) with MrBayes v.3.2.6 (Ronquist et al. 2012). All analyses were performed for each region separately, then a concatenated alignment was performed only for the plastid regions. These analyses employed a General Time Reversible substitution model (GTR) with a gamma model of rate heterogeneity (Abadi et al. 2019). Four independent runs with four Markov Chain Monte Carlo (MCMC) chains were conducted, sampling every 1,000 generations for 10,000,000 generations. Plastid, nuclear, and consensus trees using majority rule 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 analysed based on the sequencing data obtained. 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 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). Additionally, nucleotide diversity rates (π), population differentiation through pairwise F_{ST} (where values

close to 0 indicate high genetic similarity, while values close to 1 suggest greater dissimilarity), and molecular variance analysis (AMOVA) were calculated for the 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 through 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 characteristics. Only the AN-708 accession 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 others varied between 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). On the other hand, 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 had rough and opaque skin of cashew apple (SCA), while others had smooth and shiny skin (Table S4, Fig. 1).

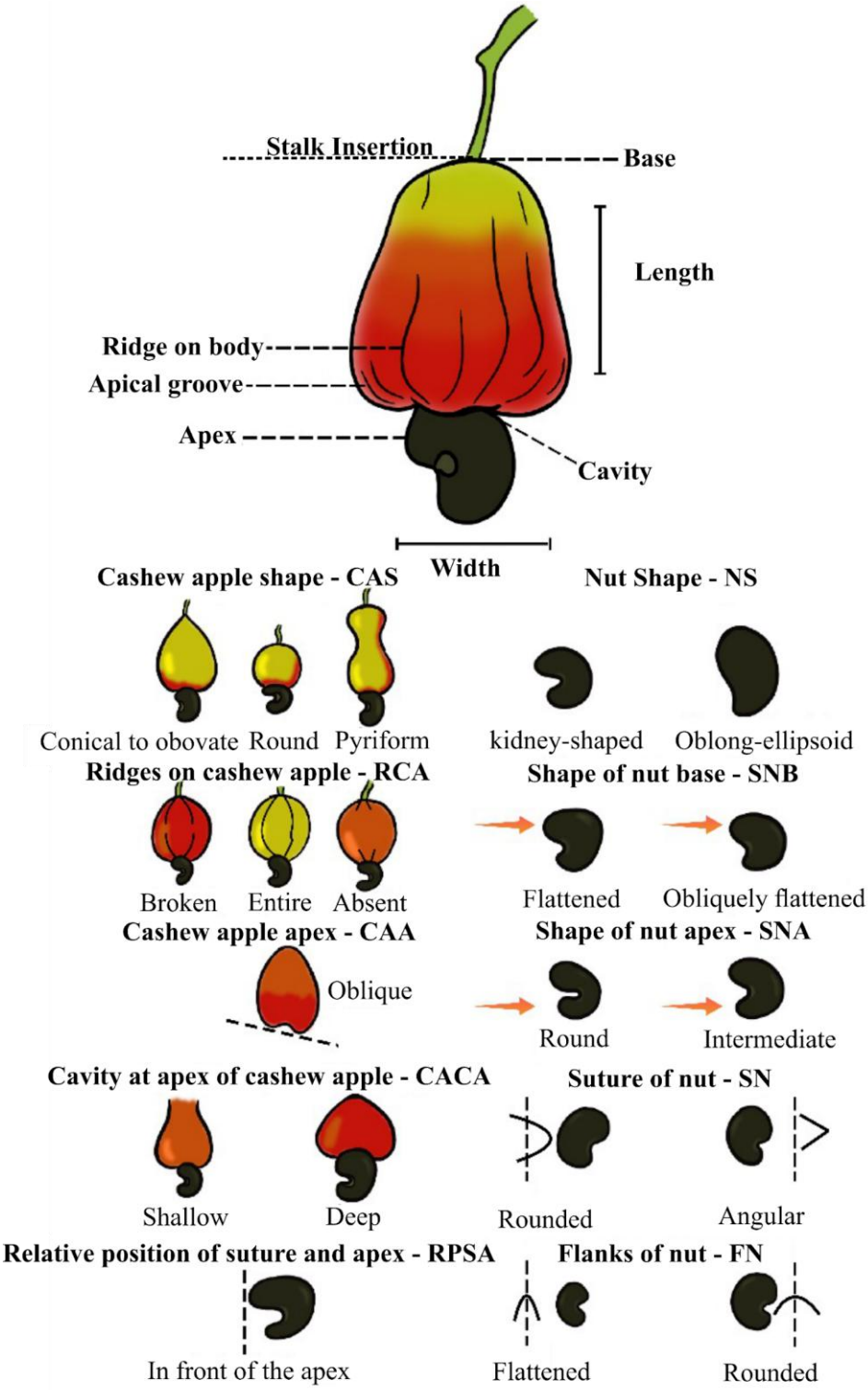


Fig. 1 Qualitative traits related to the fruits (nuts) and peduncles (pseudofruits) of *A. occidentale* e *A. humile* used here as the basis for the morphological analysis.

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 trait that most indicated distinct groups. Group "a" exclusively consisted of commercial cashew accession BRS 226, weighing 104.02 g, while groups "b" and

"c" showed average WeCA values ranging from 29.07 g (AN-800) to 34.06 g (AN-802), originating from Parnaíba municipality. Group "f," with WeCA ranging from 3.31 g (AN-716) to 6.92 g (AN-702), consisted of thirteen accessions from different locations and included the two *A. humile* accessions (AN-710 and AN-711). Similarly, evaluating the results obtained for length (LeCA), width (WiCA), and diameter cashew apple (DiCA), it was clear that accession BRS 226, the only one belonging to group "a," had the highest values for both descriptors. On the other hand, accession AN-716, also from *A. occidentale*, was the lowest among those evaluated. However, no trend was observed between such descriptors and their geographical distribution (Tables S1 and S5).

The Total Soluble Solids (TSS) of cashew apples distinguished the individuals into five different clusters (Table S5). Group "a" consisted of only BRS 226, recording the lowest °Brix content (11.1%) among the analysed matrices. Meanwhile, the accession AN-716 exhibited the highest soluble solids content (°Brix) with a value of 20.0%, indicating a high sugar content. However, this accession does not stand out in terms of weight and length, essential attributes for product acceptance and commercialization. From the cluster analysis performed with colorimetric data, significant differences were detected, distinguishing three colour patterns that vary from yellow to red among the evaluated cashew matrices (Table S5). The luminosity parameter "*L*," as well as the color parameters "*a*" and "*b*," provided by the portable spectrophotometer to analyse the color of the fruits and peduncles, grouped the evaluated accessions into three clusters (Table S5).

When evaluating qualitative data from the cashew nut, it was observed that the relative position of suture and apex (RPSA) is highly conserved among most analysed accessions, except for BRS 226 and AN-802, where the suture is in front of the nut apex (Table S6, Fig. 1). Regarding the shape of nut apex (SNA), it ranged from pointed (only AN-706) to intermediate or rounded. Similarly, the shape of nut base (SNB) varied between flat, obliquely flat (only AN-706), and rounded (Table S6). For the suture of nut (SN), angular or rounded accessions were observed, while the nut shape (NS) varied from oblong ellipsoid to kidney-shaped and the flanks of nut (FN) descriptor varied between bulging, flat, or rounded. Finally, the stylar scar on nut (SSN) resulted in the division of accessions into either large or small stylar scars (Table S6).

Quantitative data regarding the nuts also presented a wide variation (Table S7). The nut length (NL) divided the individuals into eight groups, with values ranging from 1.43 g (AN-707) to 3.63 g (BRS 226). However, when evaluating nut width (NW) and nut thickness (NT), nine groups were formed, with the lowest averages obtained for accession AN-715 (NW = 1.18 cm; NT = 0.8 cm), while the highest values were observed for accession BRS 226 (NW = 2.99 cm; NT = 2.22 cm). Regarding weight of nut (WeN), the six distinct groups ranged from 1.03 g (AN-718) to 10.93 g (BRS 226). Although accession BRS 226 of *A. occidentale* stood out for the higher nut weight, it was not possible to establish direct relationships with nut length, width, and thickness among all studied accessions. Colorimetric analysis revealed gray nuts, with a greenish hue in some cases (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 good representation of the distance matrix. The first group consisted exclusively of *A. occidentale* BRS 226. The second group was formed by matrices of *A. occidentale* and *A. humile*, subdivided into three subgroups. The first subgroup is

formed by AN-700 and AN-704 from Bom Jesus municipality. The second subgroup consists of accessions from Bom Jesus, Currais, Alvorada do Gurguéia, and Buriti dos Lopes. The third subgroup consisted exclusively of accessions from Parnaíba. Thus, geographical structuring between Parnaíba versus remaining populations could be observed.

The characteristics that most contributed to the formation of the clusters are shown in (Fig. 2**B**). The variables cashew apple length (LeCA), nut weight (WeN), nut thickness (NT), and nut length (NL) accounted for 54.3% of the divergence observed among the analysed accessions (Fig. 2**B**). Principal Component Analysis (PCA) separated the accessions along both axes, explaining 73.80% of the variation (Fig. 2**C**). There was congruence between the data presented by the Principal Component Analysis (PCA-Biplot) and the dendrogram generated by the UPGMA analysis (Figs. 2**A** and 2**C**). The BRS 226 accession appeared as the most contrasting among the analysed 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).

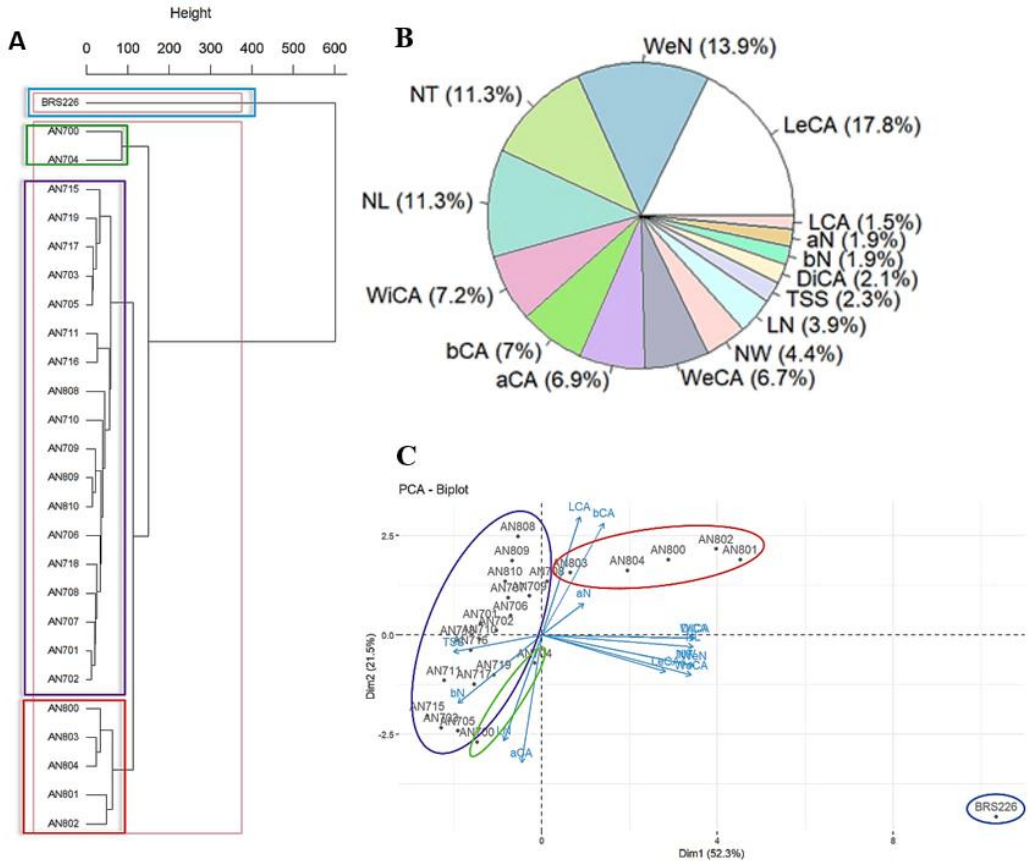


Fig. 2 Similarity among accessions from natural populations of cashew, as well as the commercial accession BRS 226. **(A)** Dendrogram obtained using the UPGMA method; **(B)** Relative importance of the variables related to the quantitative data of the cashew apple and nut. Description of the evaluated 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.

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

The cytogenetic analysis was conducted with 12 accessions of *A. occidentale*, as well as the two accessions of *A. humile* and revealed stability in chromosome number, with a karyotype of $2n = 40$ for all accessions (Fig. 3, Fig. 4). Here we report the first chromosome counts for *A. humile*. The CMA/DAPI staining analysis also indicated stability in terms of the number of GC-rich heterochromatic bands (CMA^+) and absence of AT-rich bands ($DAPI^+$). We observed $CMA^+/DAPI^-$ bands located specially in the terminal region of the short arm of three chromosome pairs in all analysed karyotypes, with one of these pairs being smaller and showing weaker staining intensity in the heterochromatic block (Fig. 3, Fig. 4). Our 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* accessions are relatively small, with a mean value of 0.88 pg/2C. Considering the three accessions with the lowest coefficient of variation (CV), the genome size of *A. occidentale and A. humile* was 1C = ~ 435 Mb (Table 1, Figure S1). The analysis of variance (ANOVA) conducted on the accessions of *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 accessions of the two species evaluated (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

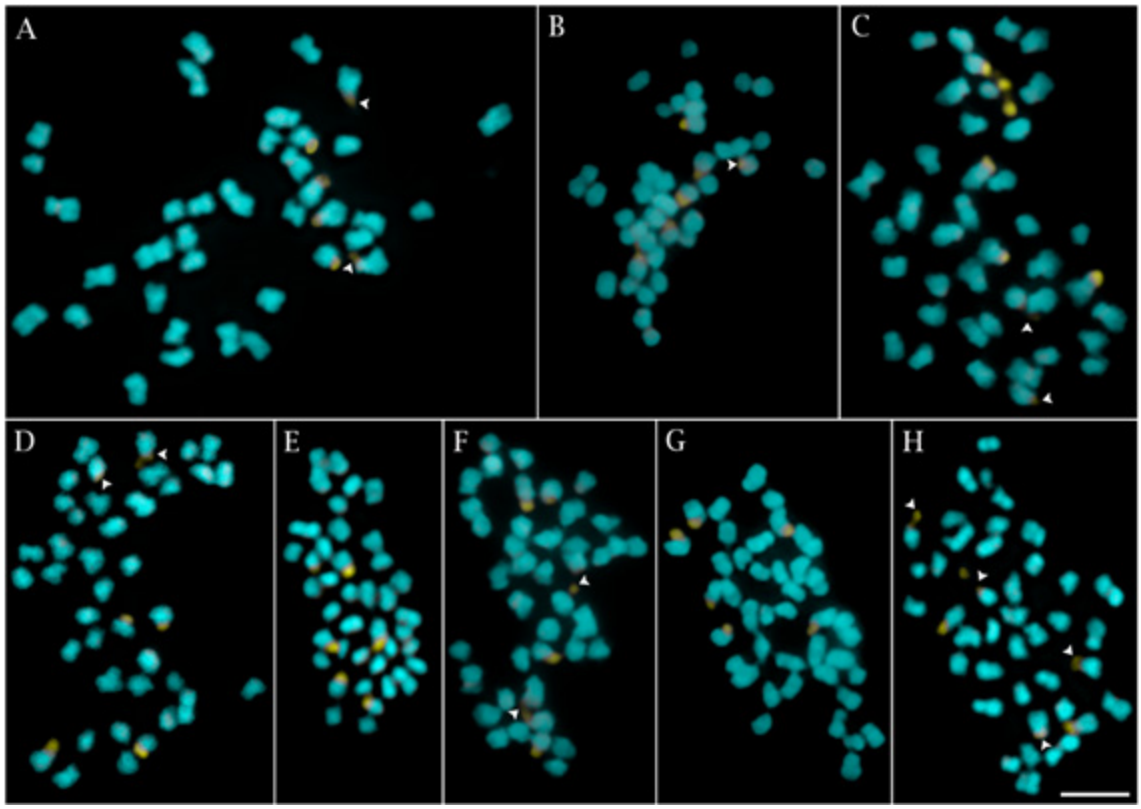


Fig. 3 Mitotic metaphases of *A. occidentale* (A-D and F-H), *A. humile* (E), stained with CMA (yellow) and DAPI (blue). AN-706 (A), AN-715 (B), AN-704 (C), AN-701 (D), AN-712 (E), AN-803 (F), AN-708 (G), AN-714 (H). Arrowheads indicate weaker terminal $CMA^{+}/DAPI^{-}$ bands. Bar in (I) corresponds to 5 µm

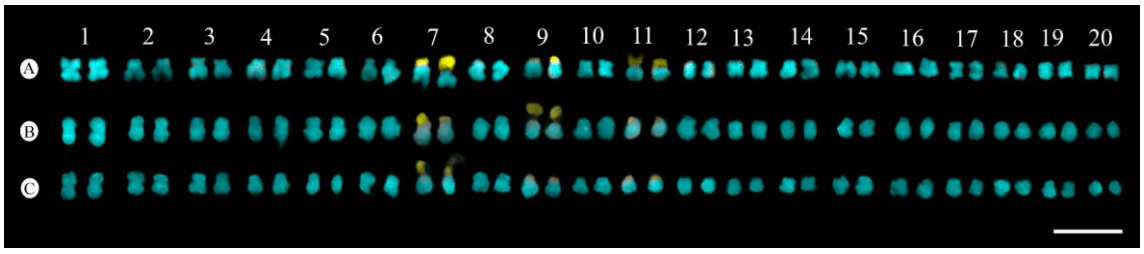


Fig. 4 Karyograms of two *Anacardium* species, showing chromosome number, morphology, and distribution of CMA^{+} heterochromatic bands. (A) *A. occidentale* (BRS 226), (B) *A. occidentale* previously identified as *A. othonianum* (BGC-45), and (C) *A. humile* (AN-712). Bar corresponds to 5 µm

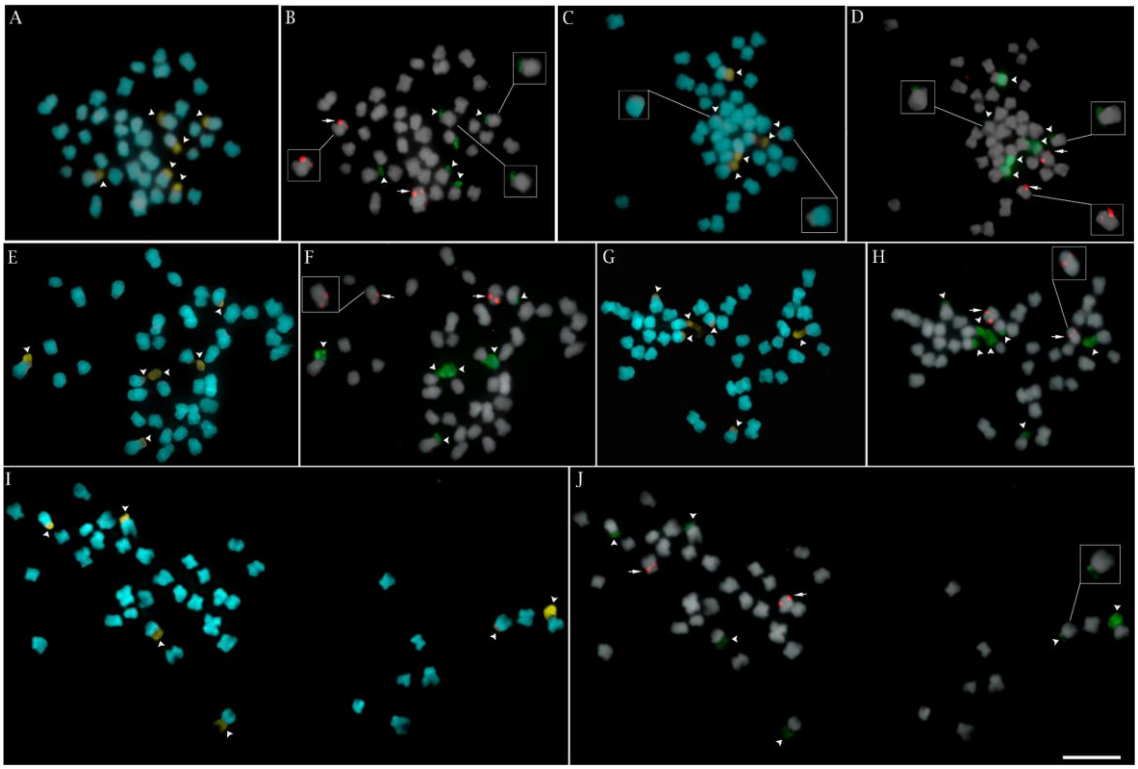


Fig. 5 Distribution of $CMA^{+}/DAPI^{-}$ heterochromatin (A, C, E, G, I) and rDNA sites (B, D, F, H, J) in *Anacardium occidentale* (A-F, I-J,) and *A. humile* (G-H). 5S rDNA (red, arrows) and 35S rDNA (green, arrowheads). A - B) AN-707; C - D) AN-801; E - F) BGC-45 previously identified as *A. othonianum*;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

Phylogenetic Relationships in Anacardium

During this study we generated a total of 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 on GenBank database were also added to the data matrix (Table S2). All regions analysed exhibited a high degree of conservation, with low genetic polymorphism, especially for the plastid regions. Table 2 summarizes the characteristics of the plastid and nuclear regions used in this study.

For some of the plastid regions (*matK*, *trnL-F*, and *rps16*) available in GenBank for several *Anacardium* accessions, the concatenated phylogenetic analysis showed *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 among these species were not well supported.

In this context, we chose to conduct an intraspecific plastome search to identify additional polymorphic regions among *A. occidentale* accessions (Figure S4). As a result of this comparative analysis, we identified that the *ycf1* gene was the most polymorphic *locus*. The tree based on the *ycf1* marker amplified for 24 accessions showed lower posterior probabilities than the concatenated tree of the three plastid regions (Figure S5), not being enough for predicting phylogenetic relationships alone. Therefore, we concatenated the *ycf1* region with the other three plastid regions, and to avoid missing data, eight

individuals from the 24 *ycf1* accessions were removed because they were not amplified for the three remaining plastid regions. The resulting concatenated plastid tree showed five well-supported clades, similar to those of the three concatenated regions (Figure S6). However, relationships among *A. occidentale* and *A. humile* remained unresolved.

The data from the ITS1-5.8S-ITS2 region also supported the monophyly of the genus *Anacardium*, albeit without strong support (PP = 0.64; Figure S8). Again, *A. excelsum* was sister to the remaining *Anacardium* accessions, but only three clades showed strong support (PP = 1). Thus, we combined the plastid and nuclear data for Bayesian analysis (Fig. 6), which allowed to observe the monophyletic nature of the genus with a strong support (PP = 1). *Fegimanra africana* remained as sister to *Anacardium*, and *A. excelsum* appeared as the first species to diverge within the genus. Another well-supported clade separates the species *A. occidentale* and *A. humile* from *A. spruceanum* and *A. parvifolium*, with one accession each. Additionally, we observed a well-supported clade that separates a single accession of *A. occidentale* from all other accessions, with three other well-supported subclades.

The *A. humile* accessions were taxonomically identified based on morphological characteristics, mainly by their shrubby habit. However, *A. occidentale* accessions exhibited an arboreal habit but were initially collected as *cajuís* (*A. humile*), as they had small fruits and pseudofruits very similar to those of *A. humile*. Indeed, most *A. occidentale* and *A. humile* accessions had fruits and pseudofruits with length and width measurements within the standards established by Mitchell and Mori (1987) for *A. humile*. Thus, our data indicates that many *A. occidentale* accessions have an intermediate morphology, with fruits and peduncles similar to *A. humile* and an arboreal habit similar to *A. occidentale*. Therefore, to test whether these *A. occidentale* accessions could be hybrids between both species, we removed all possible *A. occidentale* "hybrids" from the plastid (Figure S7) and nuclear (Figure S9) trees, as the presence of hybrids can alter tree topology and reduce support. Even without the supposed hybrids, it was not possible to clearly separate *A. humile* and *A. occidentale* in the plastid analysis, suggesting they may share the same plastid haplotypes. On the other hand, the *ITS* tree supported a clade of *A. occidentale* (PP = 1) separate from *A. humile*, but it did not resolve the relationships among these species (Figure S9).

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%

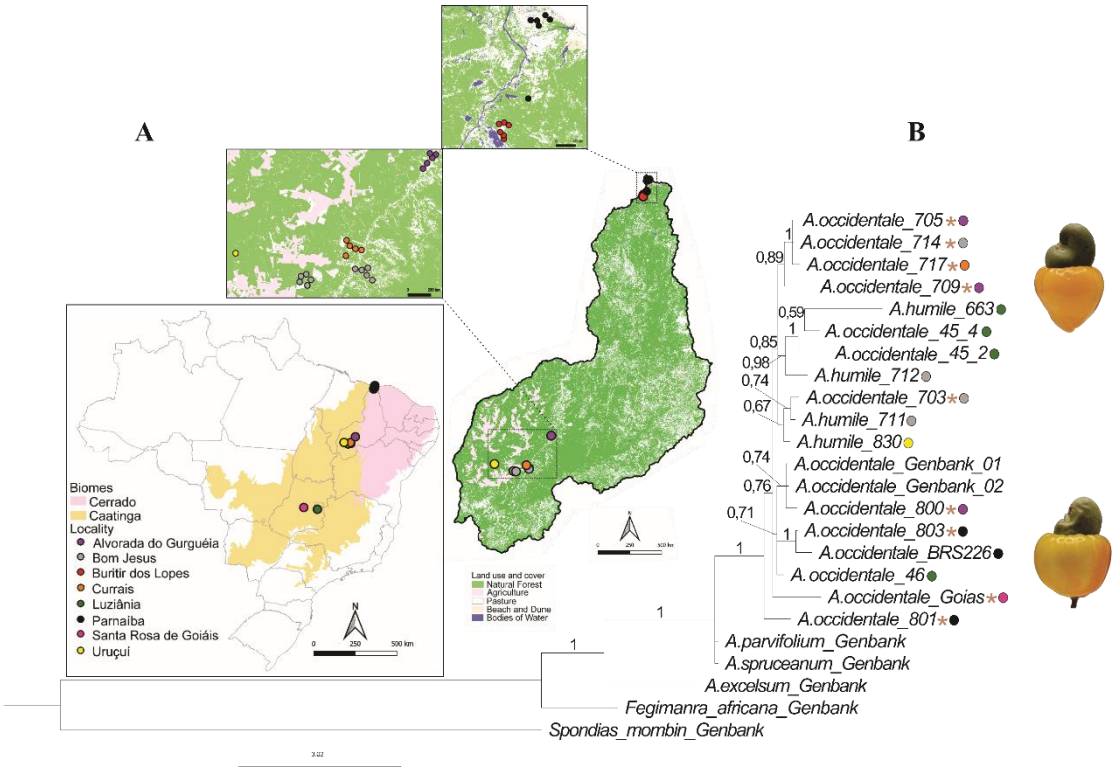


Fig. 6 Geographical distribution and phylogenetic relationship of *Anacardium* accessions analysed. **A**. Map of Brazil showing the locations and respective biomes where samples were collected, with the state of Piauí highlighted in green, indicating areas with more collection sites, along with an enlargement 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

Diversity, Differentiation and Genetic Structure

In other to investigate the relationship between *Anacardium occidentale* and *A. humile* in more detail, genetic diversity, differentiation, and population structure were analysed using nuclear and plastid sequence data. The highest nucleotide diversity index (π) for the nuclear data was observed in the *A. occidentale*-Parnaíba and *A. occidentale*-Alvorada populations, indicating considerably higher nucleotide diversity, while the *A. occidentale*-Currais population showed the lowest nucleotide diversity index (Table S8). Regarding the plastid data, the group composed of cultivated *A. occidentale* accessions (BRS 226, BGC-45, and BGC-46) presented the highest nucleotide diversity index, while the lowest index was recorded for the *A. occidentale*-Bom Jesus population (Table S8), except for populations with a reduced number of samples (π = 0.0).

Regarding the population differentiation from the nuclear and plastidial data (Tables S9 and S10), F_{ST} values indicated 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 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 analysed populations (Tables S11 and S12). Additionally, negative values were recorded, suggesting that the genetic difference between the groups is very small or non-existent.

The AMOVA result for the nuclear data revealed that most of the variation is observed within populations (96.67%), with only a small fraction (9.96%) attributed to differences between populations. The variation between groups, corresponding to the two species, *A. occidentale* and *A. humile*, was also negative (-6.63), suggesting that there is no genetic difference between the species (Table S13). The AMOVA results for plastidial data indicate that most of the genetic variation (86.24%) occurs within populations, and the variation between populations was higher than observed with nuclear data but represented only 13.82% of the total variation. Again, no significant genetic differentiation was observed between groups/species (Table S14).

While analysing individuals from a phylogeographic perspective through haplotype distributions and networks (Figures S10 and S11), a certain concordance is observed between the diversity and genetic structure analyses. The total number of haplotypes for the nuclear dataset was h = 16, with a different haplotype per individual, except for 2 individuals with the same haplotype (H3). Except for haplotype H13, which belongs to *A. humile*, all the other haplotypes belonged to *A. occidentale*. The nuclear haplotypic diversity was Hd = 0.9926. For the plastidial dataset, the total number of haplotypes was also h = 16, with haplotypes H10, H11, and H12 belonging to *A. humile*, and the plastidial haplotypic diversity was Hd = 1. Haplotype networks did not differentiate *A. occidentale* and *A. humile*, similarly to the phylogenetic analysis that did not indicate the formation of species-specific clades. Nevertheless, the plastidial network indicates higher 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 identified variation for most of the qualitative and quantitative traits of fruits and pseudofruits, thus demonstrating heterogeneity among the accessions. Morphological variation in the common cashew tree had already been demonstrated by previous studies, such as those conducted by Asna et al. (2021), who described conical-ovoid pseudofruits for *A. occidentale* accessions, and França et al. (2020), who identified a predominance of pyriform shapes in pseudofruits of *Anacardium* sp. When this phenotypic plasticity was analysed among different species such as *A. occidentale*, *A. othonianum*, and *A. humile*, the most prevalent traits described were the conical-ovoid, pyriform, and rounded shapes (Rufino 2004; Castro et al. 2011; Ferreira et al. 2015), corroborating what was observed in most samples investigated in this study. We also observed 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 methods confirmed the remarkable genetic diversity expressed in morphological dissimilarities. Variation for peduncle weight had already 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 the

data presented here; although, Djolossè et al. (2019) reported an average weight of 74.12 g for *A. occidentale*, a value slightly lower than that found here for the BRS 226 accession. As for nut weight (WeN), there is clear variation among the accessions of *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), while for cashew, Semporé et al. (2021) found nuts with weights ranging from 4.62 to 9.06 g, compatible with those found in this study. It is internationally established that the commercial weight of cashew nuts ranges from 7 to 9 g (IBPGR 1986). Thus, most of the nuts evaluated here would be rejected by the international market.

Previous studies observed variations in peduncle length in cashew tree clones at different maturation stages, with values between 57.93 and 86.20 mm (Gomes et al. 2006; Lopes et al. 2011). In our results, the apple length (LeCA) obtained for the commercial accession (BRS 226) was similar to the values previously published. However, accession BRS 226 shows clearly higher values when compared to other *A. occidentale* accessions and to the *A. humile* accessions. In natural populations of *Anacardium* species, peduncle lengths ranged from 8.00 to 43.60 mm (Rufino 2004; Rocha et al. 2013; Santos and Santos-Júnior 2015; Pereira 2018; Borges et al. 2022). These values are similar to those found in our results. Regarding the nut, the maximum length (NL) of *A. humile* was 16.1 mm, which indicates that the length of *A. humile* nuts can vary up to 20.5 mm (Lima et al. 1988; Santos and Santos-Júnior 2015; Pereira 2018). Mitchell and Mori (1987) recorded values for *A. humile* nuts of 1.2–2.3 × 1–1.7 cm and peduncles of 1–3 × 1–2 cm, while for *A. occidentale*, the values are 2–3.5 × 1–2 cm for the nut and 5–20 × 2–8 cm for the peduncle. Based on these measurements, our results show that the peduncles of *A. humile* are smaller and are in accordance with the measurements established by Mitchell and Mori (1987). However, most of the wild accessions identified here as *A. occidentale* showed fruit and pseudofruit length and width measurements within the standards established by Mitchell and Mori (1987) for *A. humile*. These values also do not meet the agronomic standards established by the cashew processing industry, especially regarding fruit and peduncle length and width. Thus, almost all *A. occidentale* accessions exhibit fruit and pseudofruit morphometry that characterizes them as similar to *A. humile*, and not to the morphological description of *A. occidentale*.

The peduncles were also evaluated for Total Soluble Solids (TSS - °Brix). The BRS 226 accession presented the lowest value, with 11.1%, while the other *A. occidentale* accessions showed values between 12.3 and 20.0%. The two *A. humile* accessions presented values of 17.4 and 19.6% (Table S5), in accordance to previous studies which reported values between 5.29% and 21.13%, evidencing considerable variation among different matrices (Rufino et al. 2002; Rufino 2004; Gomes et al. 2013; Pereira 2018). The values found 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 that the peduncles of these matrices could be used by the food industry in the production of beverages, ice creams, jams, and other 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 the other with the *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 evaluated. In our results, we observed that, despite some degree of population structure due to geographic location, there was also clustering among accessions from different localities. This may be influenced by the occurrence of possible gene flow, which can be 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; in addition to floral biology that promotes cross-pollination in *Anacardium* species (Takehana et al. 2013; Hamrick 2012; Borges et al. 2018; Gomes et al. 2021).

Karyotypic Stability in Species of Anacardium

The present study demonstrated a remarkable karyotypic stability among *Anacardium* species. The observed chromosome number was consistently $2n = 40$, 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. Chromosome counting errors may arise due to technical difficulties, a high chromosome number, or their small size, which often results in very similar morphology (Figueredo et al. 2016; Hoang et al. 2022). The 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 allows more precise karyotype characterization (Figueredo et al. 2016). Other woody genera such as *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 (Lás Penas and Bernardello 2006), whereas in *Pistacia vera* L. multiple $DAPI^+$ bands were observed in the proximal and terminal chromosome regions (Sola-Campoy et al. 2015). In *Spondias*, $CMA^+/DAPI^-$ band number and distribution differed among species (Almeida et al. 2007), suggesting a higher level of structural rearrangements during the evolution of these species. No variation was detected in the number and position of rDNA sites among the *Anacardium* accessions analysed. In *Mangifera indica*, a similar pattern was found: one 5S rDNA site and three 35S rDNA site pairs (Yonemori et al. 2010), consistent with the phylogenetic proximity between these two genera. In this context, the number of 5S sites appears to be 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).

Karyotypic stability observed in *Anacardium* has also been reported in other woody species, such as in *Cenostigma* Tul., which also 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 were observed, with no 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 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 cashew (*A. occidentale*) accessions from Africa, with 0.85 pg/2C (419 Mb/1C; Aliyu 2014). Small genomes appear to be common in Anacardiaceae, as reported for *Mangifera*, *Pistacia*, *Lannea* L., *Rhus*, and *Toxicodendron* L., with DNA content ranging from 0.30 to 0.75 pg/1C, and records of polyploidy (tetraploidy) only in some *Mangifera* representatives (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, polyploidy has not 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. Thus, the plastidial and nuclear data did not reveal a clear distinction between *A. occidentale* and *A. humile*, as they were shown to be very similar from a molecular standpoint. However, *A. humile* and *A. occidentale* were expected to be two distinct taxa. Several studies have successfully used molecular data in phylogenetic reconstruction for different genera within the 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*) allowed well-resolved phylogenies with well-supported clades, enabling the separation of species (Yi et al. 2004; Nobre et al. 2018; Silva-Luz et al. 2019). Ariyarathne et al. (2020) used *ITS* and *matK* regions for phylogenetic analyses with endemic Anacardiaceae species from Sri Lanka. In this analysis, a close relationship between *A. occidentale* and the genus *Mangifera* was observed, with good support, corroborated by a 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 select the most informative loci for phylogenetic analysis, but the plastidial variation of the selected loci was limited. Rabah et al. (2017) focused on the complete plastome sequencing of *A. occidentale* and the sequencing of *trnK_matK*, *trnL-F*, and *ndhF* from one accession each of *A. nanum* A.St.-Hil., *A. humile*, *A. corymbosum* Barb.Rodr., and *A. excelsum*. Despite the limited sampling, they recovered *A. nanum* and *A. humile* as sister species, while *A. occidentale* formed a distinct clade. Notably, *A. excelsum* was the first to diverge and the only one that did not share a plastome insertion, suggesting that this modification likely occurred less than 20 million years ago (Xie et al. 2014), what is congruent with our analysis.

Among our *Anacardium* accessions, we observed high morphological variability but low molecular polymorphism in the evaluated sequences, especially in plastidial regions, making them less phylogenetically informative. Most of the molecular variation was observed among individuals of the same population, with no differentiation between species. Previous investigations on the genetic diversity and structure of natural populations of *Anacardium* species, including *A. microcarpum*, *A. othonianum*, *A. occidentale*, *A. giganteum* W. Hancock ex Engl., and *A. excelsum*, using ISSR and SSR markers, revealed

a low to high genetic diversity but low levels of genetic structure, suggesting high levels of gene flow (Bocanegra-González and Guillemin 2018; Borges et al. 2018; Dos Santos et al. 2019; Gomes et al. 2021).

The lack of phylogenetic separation between the accessions of *A. humile* and *A. occidentale* could be caused by weak genetic differentiation and incomplete lineage sorting. It is possible that they represent two incipient species that have not yet fully differentiated. Generation time and habit can influence the rate of molecular evolution in plants. Herbaceous or annual plants tend to evolve more rapidly. Indeed, positive correlations have been established between nucleotide substitution rates and herbaceous habits, as these species are often small and have shorter generation times. On the other hand, woody and tree species tend to have slower molecular evolution, both in the chloroplast and nuclear genome (Lanfear et al. 2013; Xin et al. 2020). Xie et al. (2014) reported that the divergence between *A. excelsum* and *A. occidentale* occurred 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 in the evaluated materials. 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, may suggest 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).

In fact, hybridization appears to be a common phenomenon within the genus *Anacardium* under natural conditions. As observed by Mitchell and Mori (1987), the three sympatric species of *Anacardium* from the Brazilian plateau (*A. occidentale*, *A. humile*, and *A. nanum*) bloom simultaneously where they coexist. Additionally, their flowers are virtually identical morphologically and are pollinated by the same insects, such as butterflies and bees. The authors suggest that there are few extrinsic barriers to crossing between these species, which may explain the 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 characters. The process of plant domestication is characterized by genetic selection performed by humans to adapt wild plants to cultivation and human preferences. This is followed by breeding practices that, although increasing the productivity and resistance of plants, also tend to reduce the existing genetic diversity, as observed by 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. The

Central Brazil region is characterized by a large 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. Considering this hypothesis, this single species would have a wild range of habits but smaller fruits than its domesticated form. The use of other molecular markers, such as SSRs or RADseq, could be more informative in a larger sampling of wild and domesticated accessions to elucidate the genetic structure of *Anacardium* species, their species boundaries, and possible hybridization.

Conclusion

Our results suggest significant morphological variability among *Anacardium* accessions from northeastern Brazil. However, this high level of phenotypical plasticity contrast with low levels of molecular variation in plastidial and nuclear regions, as well as the lack of cytogenetic variation. 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, which are not mutually exclusive.

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 analysed 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.

Conflict of interest

The authors declare that they have no conflict of interest.

Data Archiving Statement

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.

References

Abadi S, Azouri D, Pupko T, Itay M (2019) Model selection may not be a mandatory step for phylogeny reconstruction. Nat Commun10:1-11. https://doi.org/10.1038/s41467-019-08822-w

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

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

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

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

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

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

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

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

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Becerra JX (2003) Evolution of Mexican *Bursera* (Burseraceae) inferred from ITS, ETS, and 5S nuclear ribosomal DNA sequences. Mol Phylogenet Evol 26:300–309. https://doi.org/10.1016/s1055-7903(02)00256-7

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

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

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

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

Borém A, Miranda GV, Fritsche-Neto R (2021) Melhoramento de plantas. Oficina de Textos, São Paulo

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

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 Linn Soc 204:137–151. https://doi.org/10.1093/botlinnean/boad043

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

Carvalho GMA, Carvalho CR, Soares FAF (2017) Flow cytometry and cytogenetic tools in *Eucalypts*: genome size variation × karyotype stability. Tree Genet Genomes 13:1–11. https://doi.org/10.1007/s11295-017-1186-0

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 Evol 303:1131–1144. https://doi.org/10.1007/s00606-017-1427-6

Crespo MFV, Souza LI (2014) Cajuí: boas práticas e manejo sustentável. Sieart, Parnaíba

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

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

Darlington CD, Janaki-Ammaal EK (1945) Chromosome Atlas of Cultivated Plants. London

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

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

Doležel J, Sgorbati S, Lucretti S (1992) Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiol Plant 85:625–631. https://doi.org/10.1111/j.1399-3054.1992.tb04764.x

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

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

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

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

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

Ferreira ME, Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análise genética. Embrapa Cenargen, Brasília

Ferreira TN, Borges EF, Ferreira TFC, Rocha AJG, Karsten J (2015) Características físicas do cajuí (*Anacardium* sp.) do Cerrado Baiano. CBPM 1:1–8

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

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

Fitmawati, Harahap SP, Sofiyanti N (2017) Phylogenetic analysis of mango (*Mangifera*) in Northern Sumatra based on gene sequences of cpDNA *trnL-F* intergenic spacer. Biodiversitas 18:715–719. https://doi.org/10.13057/biodiv/d180238

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

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

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

Gill BS, Bir SS, Singhal VK (1979) In IOPB Chromosome number reports LXV. Taxon 28:630

Gill BS, Singhal VK, Bedi YS, Bir SS (1990) Cytological evolution in the woody taxa of *Pachmarhi* Hills. J Cytol Genet 25:308–320

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

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

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

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 Int 431487:1–14. https://doi.org/10.1155/2015/431487

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

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

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

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

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

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

IBPGR (1986) International Board for Plant Genetic Resources: Cashew Descriptors. IBPGR, Rome

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

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

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

Las Peña ML, Bernardello GLM, Steibel P, Troiani H (2006) Cytogenetic studies in *Schinus* (Anacardiaceae) Missouri Botanical Garden. Arnaldoa 13:270–275

Lima VPMS, Ramos AD, Franca FMC (1988) A cultura do cajueiro no Nordeste do Brasil. Banco do Nordeste do Brasil, Fortaleza

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

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

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

Machado O (1944) Estudos novos sobre uma planta velha – o cajueiro (*Anacardium occidentale* L.). Rodriguesia 8:19–48

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 Evol 293:227–235. https://doi.org/10.1007/s00606-011-0434-2

Mitchell JD, Mori SA (1987) The cashew and its relatives (*Anacardium*: Anacardiaceae), Memoirs of the New York Botanical Garden-NYBG Press, New York

Mitchell JD, Pel SK, Bachelier JB, Warschefsky EJ, Joyce EM, Canadell LC, Silva-Luz CL, Coiffard C (2022) Neotropical Anacardiaceae (cashew family). Braz J Bot 45:139–180. https://doi.org/10.1007/s40415-022-00793-5

Mojena R (1977) Hierarchical grouping methods and stopping rules: an evaluation. Comput J 20:359–363. https://doi.org/10.1093/comjnl/20.4.359

N'Danikou S, Tchokponhoue DA (2019) Plant domestication for enhanced food security. In: Leal Filho W, Azul A, Brandli L, Özuyar 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

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 Biol 41:878–883. https://doi.org/10.1590/1678-4685-GMB-2017-0256

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

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

Paiva FFA, Garruti DS, Silva Neto RM (2000) Aproveitamento Industrial do caju. Embrapa-CNPAT, Fortaleza

Pedrosa PH, Gitaí J, Silva AEB, Felix LP, Guerra M (1999) Citogenética de angiospermas coletadas em Pernambuco. Acta Bot Bras 13:49–60. https://doi.org/10.1590/S0102-33061999000100006

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

Pell SK (2004) Molecular systematics of the cashew family (Anacardiaceae). Dissertation, Louisiana State University

Pereira LD (2018) Caracterização e diversidade genética de frutos de cajuzinho-do-cerrado. Dissertation, Federal University of Goiás

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

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

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

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

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

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

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

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

Rufino MSM, Vasconcelos LFL, Corrêa MPF, Ribeiro VQ, Soares EB, Souza VABDE (2002) Caracterização física e química do fruto e pseudofruto de genótipos de cajuí (*Anacardium* spp.). CBF17:5

Rufino MSM (2004) Qualidade e potencial de utilização de cajuís (*Anacardium* spp.) oriundos da vegetação litorânea do Piauí. Dissertation, Federal University of Piauí

Salgotra RK, Chauhan BS (2023) Genetic diversity, conservation, and utilization of plant genetic resources. Genes Acre14:1–20. https://doi.org/10.3390/genes14010174

Sang T, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am J Bot 84:1120–1136. https://doi.org/10.2307/2446155

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

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

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 Rep 12:1–13. https://doi.org/10.1038/s41598-022-22600-7

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 Res 6: 1–8. https://doi.org/10.1016/j.jafr.2021.100219

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

Silva-Júnior JF, Souza FVD, Pádua JG (2021) A arca de Noé das frutas nativas brasileiras. Embrapa, Brasília

Singh D (1981) The relative importance of characters affecting genetic divergence. Indian J Genet Plant Breed 41:237–245

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

Sokal RR, Rohlf FJ (1962) The comparison of dendrograms by objeticve methods. Taxon 11:30–40. https://doi.org/10.2307/1217208

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

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

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 Biol 10:1–22. https://doi.org/10.1186/1471-2148-10-352

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

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

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three noncoding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109. https://doi.org/10.1007/BF00037152

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

Turchetto-Zolet AC, Turchetto C, Zanella CM, Passaia G (2017) Marcadores moleculares na era genômica: metodologias e aplicações. Sociedade Brasileira de Genética, Ribeirão Preto

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 Soc 188:269–280. https://doi.org/10.1093/botlinnean/boy054

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

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

Weeks A, Zapata F, Pell SK, Daly DC, Mitchell JD, Fine PVA (2014) To move or to evolve: contrasting patterns of intercontinental connectivity and climatic niche evolution in "Terebinthaceae" (Anacardiaceae and Burseraceae). Front Genet 5:1–20. https://doi.org/10.3389/fgene.2014.00409

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

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

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

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

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

Yonemori K, Honsho C, Kanzaki S, Eiadtong W, Sugiura A (2002) Phylogenetic relationships of *Mangifera* species revealed by ITS sequences of nuclear ribosomal DNA and a possibility of their hybrid origin. Plant Syst Evol 231:59–75. https://doi.org/10.1007/s006060200011

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

Supplementary Material – Tables

Table S1 *Anacardium* accessions analysed from the Active Germplasm Bank of the Professor Cinobelina Elvas Campus, Federal University of Piauí, in Bom Jesus-PI (AN: *A. humile* or *A. occidentale*), and provided by Embrapa Tropical Agroindustry (CNPAT), Fortaleza-CE (BRS: *A. occidentale*; and BGC: *A. occidentale*, previously identified as *A. othonianum*). BRS 226 is a cultivar, and clones from this accession planted in Fortaleza-CE and Parnaíba-PI were evaluated. The abbreviations PI, CE, and GO refer to the states of Piauí, Ceará, and Goiás, respectively, while DF stands for the Federal District

Species	Germplasm ID	Elevation (m)	Municipality	State	Voucher	Geographic Coordinates	Morphological Description	Type of analysis
A. occidentale	AN-700	271	Bom Jesus	PI	88.968	09°05’03.1’’S, 44°20’23.3” W	Tree, red peduncle	Agronomy, cytometry and molecular
A. occidentale	AN-701	264	Bom Jesus	PI	88.969	09°05’03.5” S, 44°20’23.3” W	Tree, yellow peduncle	Agronomy, cytogenetics, cytometry and molecular
A. occidentale	AN-702	270	Bom Jesus	PI	88.970	09°05'51.5"S, 44°20'33.1" W	Tree, yellow peduncle	Agronomy and molecular
A. occidentale	AN-703	263	Bom Jesus	PI	88.971	09°05’53.5” S, 44°20’34.8” W	Tree, red peduncle	Agronomy and molecular
A. occidentale	AN-704	262	Bom Jesus	PI	88.972	09°05’51.7” S, 44°20’34.7” W	Tree, red peduncle	Agronomy, cytogenetics,

								cytometry and molecular
A. occidentale	AN-705	221	Alvorada do Gurguéia	PI	88.973	08°22’35.0” S, 43°51’27.1” W	Tree, red peduncle	Agronomy, cytometry and molecular
A. occidentale	AN-706	117	Alvorada do Gurguéia	PI	88.974	08°22’35.5” S, 43°51’30.5” W	Tree, yellow peduncle	Agronomy, cytogenetics and cytometry
A. occidentale	AN-707	114	Alvorada do Gurguéia	PI	88.975	08°22’32.4” S, 43°51’31.2” W	Tree, yellow peduncle	Agronomy and cytogenetics
A. occidentale	AN-708	103	Alvorada do Gurguéia	PI	88.976	08°22’53.3” S, 43°51’41.9” W	Tree, yellow peduncle	Agronomy, cytogenetics and molecular
A. occidentale	AN-709	202	Alvorada do Gurguéia	PI	88.977	08°22’43.1” S, 43°51’38.2” W	Tree, yellow peduncle	Agronomy, cytometry and molecular
A. humile	AN-710	366	Corrente dos Matões- Bom Jesus	PI	88.978	09°08’31.7” S, 44°40’36.7” W	Shrub, orange peduncle	Agronomy

A. humile	AN-711	364	Corrente dos Matões- Bom Jesus	PI	88.979	09°08’30.3” S, 44°40’36.5” W	Shrub, yellow peduncle	Agronomy, Cytogenetics and molecular
A. humile	AN-712	361	Corrente dos Matões-Bom Jesus	PI	88.980	09°08’31.3” S, 44°40’36.0” W	Shrub, orange peduncle	Cytogenetics, Cytometry and molecular
A. occidentale	AN-714	340	Corrente dos Matões- Bom Jesus	PI	88.982	09°09’06.9” S, 44°37’23.4” W	Tree, orange/yellowish peduncle	Cytogenetics and molecular
A. occidentale	AN-715	366	Currais	PI	88.983	09°01’05.6” S, 44°23’52.0” W	Tree, orange peduncle	Agronomy, cytogenetics and molecular
A. occidentale	AN-716	377	Currais	PI	88.984	09°00'52.7" S, 44°23'45.8" W	Tree, orange/yellowish peduncle	Agronomy
A. occidentale	AN-717	368	Currais	PI	88.985	09°00’52.6” S, 44°23’46.5” W	Tree, red/orange peduncle	Agronomy, cytometry and molecular
A. occidentale	AN-718	378	Currais	PI	88.986	09°00'51.4"S 44°23'48.5" W	Tree, yellow peduncle	Agronomy and molecular
A. occidentale	AN-719	379	Currais	PI	88.987	09°00’52.3” S, 44°23’47.4” W	Tree, yellow peduncle	Agronomy, cytometry and molecular

A. occidentale	AN-800	10	Parnaíba	PI	90.237	02°51'5.26"S 41°46'0.03"W	Tree, yellow peduncle	Agronomy and molecular
A. occidentale	AN-801	6	Parnaíba	PI	90.238	02°51'6.85" S, 41°45'54.21" W	Tree, yellow peduncle	Agronomy, cytogenetics, cytometry and molecular
A. occidentale	AN-802	2	Parnaíba	PI	90.239	02°51'4.49" S, 41°45'53.95" W	Tree, yellow peduncle	Agronomy
A. occidentale	AN-803	10	Parnaíba	PI	90.240	02°50'4.86" S 41°44'21.85" W	Tree, yellow peduncle	Agronomy, cytogenetics and molecular
A. occidentale	AN-804	9	Parnaíba	PI	90.241	02°50'7.56" S, 41°44'20.04" W	Tree, yellow peduncle	Agronomy
A. occidentale	AN-808	48	Buriti dos Lopes	PI	90.245	03°11'34.64" S, 41°52'43.70" W	Tree, yellow peduncle	Agronomy
A. occidentale	AN-809	46	Buriti dos Lopes	PI	90.246	03°12'27.43" S, 41°52'02.51" W	Tree, yellow peduncle	Agronomy
A. occidentale	AN-810	45	Buriti dos Lopes	PI	90.247	03°12'2.41" S, 41°52'02.27" W	Tree, yellow peduncle	Agronomy

A. humile	AN-830	-	Uruçuí	PI	-	08°59'28.9" S 45°05'24.7" W	Shrub, yellow peduncle	Molecular
A. occidentale	BGC-45	-	CNPAT	GO	-	16°13'44.6" S 47°58'15.2" W	-	Cytogenetics and cytometry
A. occidentale	BGC-45.2	-	CNPAT	GO	-	16°13'44.6" S 47°58'15.2" W	-	Molecular
A. occidentale	BGC-45.4	-	CNPAT	GO	-	16°13'44.6" S 47°58'15.2" W	-	Molecular
A. occidentale	BGC-46	-	CNPAT	GO	-	16°13'44.6" S 47°58'15.2" W	-	Molecular
A. occidentale	Clone BRS 226	-	CNPAT	CE	-	-	Tree, red/orange peduncle	Cytogenetics and molecular
A. occidentale	Clone BRS 226	-	Parnaíba	PI	-	09°04'54.5" S 44°19'40.3" W	Tree, red/orange peduncle	Agronomy
A. humile	663	-	Brasília	DF	-	-	without fruits and pseudofruits	Molecular
A. humile	670	-	Brasília	DF	-	-	without fruits and pseudofruits	Molecular
A. occidentale	Goiás	-	Santa Rosa de Goiás	GO	-	16°01'54.8" S 49°26'40.9" W	Tree, red peduncle	Molecular

Table S2 PCR (*Polymerase Chain Reaction*) programs used for amplification of nuclear and plastid loci

Locus		35 cycles			
	Denaturation	Denaturation	Annealing	Extension	Final Extension
matK	4 min at 94°C	1 min at 94°C	1 min at 55°C	2 min at 72°C	8 min at 65°C
rps16, trnL-F and ycf1	5 min at 95°C	1 min at 95°C	1 min at 56°C	1 min at 72°C	10 min at 72°C
ITS	5 min at 95°C	1 min at 95°C	1 min at 55°C	1 min at 72°C	10 min at 72°C

Table S3 The samples used in the phylogenetic analyses. Species sampled, voucher information, and GenBank accession numbers. Sequences taken from GenBank are listed with the corresponding publication. New sequences were generated during this study

Taxa	Accession	*ITS*	*matK*	*rps16*	*trnL_trnF*	*ycf1*	Reference (s)
Spondias mombin Jacq	-	AF445882	NC_035973	NC_035973	KR081860	NC_035973	Becerra (2003); Santos and Almeida (2019)
Fegimanra africana (Oliv.) Pierre	-	-	AY594489	AY594599	AY594515	-	Pell (2004)
Mangifera indica L.	-	MF678509	-	-	-	NC_035239	Fitmawati et al. (2017); Rabah et al. (2017)
Anacardium excelsum (Bertero & Balb. ex Kunth) Skeels	Montiel 32769; BioBot02020; Daly 13970	KF664193	JQ586468	KP055362	KP055484	-	Xie et al. (2014); Weeks et al. (2014)

Anacardium parvifolium Ducke	Reserva Ducke (INPA)	-	-	KP055364	KP055485	-	Weeks et al. (2014)
Anacardium spruceanum Beth.ex Engl.	INPA2527	-	-	KP055365	KP055486	-	Weeks et al. (2014)
Anacardium occidentale L.	Zhang L sn (KUN)	KF664192; AB071690	NC_035235	KF664442	AY594497	NC_035235	Yonemori et al. (2002); Pell (2004); Xie et al. (2014); Rabah et al. (2017)
Anacardium occidentale	AN-700	-	-	-	-	PRJNA1224782	This study
Anacardium occidentale	AN-701	PV089646	-	-	-	PRJNA1224782	This study
Anacardium occidentale	AN-702	-	-	-	-	PRJNA1224782	This study
Anacardium occidentale	AN-703	PV089647	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium occidentale	AN-704	PV089648	-	-	-	PRJNA1224782	This study
Anacardium occidentale	AN-705	PV089649	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study

Anacardium occidentale	AN-708	-	-	-	-	PRJNA1224782	This study
Anacardium occidentale	AN-709	PV089650	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium humile A.St.-Hil.	AN-711	-	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium occidentale	AN-714	PV089651	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium humile	AN-712	PV089703	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium occidentale	AN-715	PV089652	-	-	-	PRJNA1224782	This study
Anacardium occidentale	AN-717	PV089653	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium occidentale	AN-718	-	-	-	-	PRJNA1224782	This study
Anacardium occidentale	AN-719	PV089654	-	-	-	-	This study
Anacardium occidentale	AN-800	PV089655	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study

Anacardium occidentale	AN-801	PV089656	-	-	-	PRJNA1224782	This study
Anacardium occidentale	AN-803	PV089657	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium occidentale	BGC-45.4	PV089661	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium occidentale	BRS 226	PV089659	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium occidentale	BGC-45.2	PV089660	-	-	-	PRJNA1224782	This study
Anacardium occidentale	BGC-46	-	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium humile A.St.-Hil.	830	-	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium occidentale	Goiás	PV089658	PRJNA1224782	PRJNA1224782	PRJNA1224782	PRJNA1224782	This study
Anacardium humile	663	-	PRJNA1224782	PRJNA1224782	PRJNA1224782	-	This study
Anacardium humile	670	-	-	-	-	PRJNA1224782	This study

Table S4 Qualitative data of pseudofruits (peduncles) from *Anacardium occidentale* L. and *A. humile* A.St.-Hil. accessions of the Active Germplasm Bank

BAG ID[1]	CAS[2]	SCAB[3]	RCA[4]	CAA[5]	GACA[6]	CACA[7]	SCA[8]
AN-700	Conical to obovate	Angular	Broken	Level	Shallow	Shallow	Smooth and glossy
AN-701	Cylindrical	Flattened	Absent	Level	Shallow	Shallow	Smooth and glossy
AN-702	Round	Obliquely flattened	Absent	Level	Absent	Shallow	Smooth and glossy
AN-703	Round	Obliquely flattened	Absent	Level	Deep	Shallow	Smooth and glossy
AN-704	Cylindrical	Angular	Absent	Level	Shallow	Deep	Smooth and glossy
AN-705	Conical to obovate	Angular	Absent	Level	Absent	Shallow	Smooth and glossy
AN-706	Round	Obliquely flattened	Broken	Level	Absent	Shallow	Smooth and glossy
AN-707	Conical to obovate	Obliquely flattened	Absent	Level	Shallow	Deep	Smooth and glossy
AN-708	Pyriform	Obliquely flattened	Broken	Level	Shallow	Deep	Smooth and glossy
AN-709	Round	Obliquely flattened	Entire	Level	Deep	Shallow	Smooth and glossy
AN-710	Round	Obliquely flattened	Absent	Level	Absent	Shallow	Smooth and glossy
AN-711	Conical to obovate	Obliquely flattened	Absent	Level	Absent	Deep	Smooth and glossy
AN-715	Conical to obovate	Angular	Absent	Level	Shallow	Shallow	Smooth and glossy
AN-716	Round	Rounded	Absent	Level	Absent	Shallow	Smooth and glossy
AN-717	Round	Obliquely flattened	Absent	Level	Deep	Absent	Smooth and glossy
AN-718	Conical to obovate	Angular	Absent	Level	Shallow	Deep	Rough and dull
AN-719	Conical to obovate	Angular	Broken	Level	Shallow	Shallow	Smooth and glossy

AN-800	Round	Obliquely flattened	Entire	Level	Absent	Shallow	Smooth and glossy
AN-801	Conical to obovate	Angular	Entire	Level	Shallow	Shallow	Smooth and glossy
AN-802	Conical to obovate	Rounded	Entire	Level	Shallow	Shallow	Smooth and glossy
AN-803	Round	Rounded	Absent	Oblique	Shallow	Shallow	Smooth and glossy
AN-804	Round	Rounded	Absent	Level	Absent	Shallow	Smooth and glossy
AN-808	Conical to obovate	Angular	Broken	Level	Absent	Shallow	Smooth and glossy
AN-809	Conical to obovate	Obliquely flattened	Broken	Level	Shallow	Shallow	Smooth and glossy
AN-810	Conical to obovate	Obliquely flattened	Broken	Level	Shallow	Shallow	Smooth and glossy
BRS 226	Cylindrical	Flattened	Entire	Level	Absent	Shallow	Smooth and glossy

BAG ID[1] - Identification of the accessions in the Germplasm Bank

CAS[2] – Cashew apple shape

SCAB[3] – Shape of cashew apple base

RCA[4] – Ridges on cashew apple

CAA[5] – Cashew apple apex

GACA[6] – Grooves on apex cashew apple

CACA[7] – Cavity at apex cashew apple

SCA[8] – Skin of cashew apple

Table S5 Scott-Knott analysis based on agronomic data of pseudofruits (peduncles) from accessions of the Active Germplasm Bank of *A. occidentale* L. and *A. humile* A.St.-Hil.

BAG ID[2]	LeCA[3] (cm)	WiCA[4] (cm)	DiCA[5] (cm)	WeCA[6] (g)	TSS[7] (%)	*L*[8]	*a*[9]	*b*[10]
AN-700	3.48 e[1]	2.28 e	2.10 f	8.52 e	15.2 c	61.33 d	23.79 c	30.03 g
AN-701	2.20 i	1.92 g	1.91 f	5.02 f	15.9 b	68.31 b	13.57 f	50.32 c
AN-702	2.15 i	2.24 e	2.04 f	6.92 f	14.0 c	69.61 b	18.24 d	48.67 d
AN-703	2.24 i	2.04 f	1.91 f	5.44 f	16.2 b	57.61 e	34.08 a	33.29 f
AN-704	5.04 b	2.21 e	2.17 e	12.13 e	16.5 b	68.86 b	20.78 d	47.60 d
AN-705	2.13 i	1.84 h	1.83 g	3.91 f	13.6 d	57.90 e	33.66 a	34.41 f
AN-706	2.44 h	2.57 d	2.52 d	9.19 e	14.0 c	75.03 a	12.35 g	48.68 d
AN-707	2.45 h	2.42 e	2.32 e	9.14 e	15.1 c	70.10 b	17.25 e	55.71 b
AN-708	2.70 g	2.35 e	2.50 d	9.16 e	13.4 d	71.94 b	14.60 e	54.92 b
AN-709	2.27 i	2.27 e	2.24 e	7.42 e	12.3 e	67.64 b	13.57 f	51.44 c
AN-710	1.92 j	2.14 f	2.33 e	6.46 f	17.4 b	62.57 d	16.39 e	36.71 e
AN-711	2.21 i	1.68 i	1.82 g	4.33 f	19.6 a	65.31 c	24.51 c	49.82 c
AN-715	2.96 f	1.92 g	1.79 g	5.22 f	17.3 b	59.45 e	24.83 c	38.18 e
AN-716	1.50 k	1.62 i	1.57 h	3.31 f	20.0 a	61.66 d	23.22 c	50.93 c
AN-717	2.01 j	2.05 f	1.98 f	5.20 f	14.8 c	58.80 e	27.24 b	37.53 e
AN-718	2.30 i	1.95 g	1.81 g	5.36 f	16.0 b	59.79 e	19.34 d	47.31 d
AN-719	2.86 g	2.01 g	2.06 f	6.91 f	13.0 d	59.63 e	30.66 a	39.53 e
AN-800	3.14 f	3.74 b	3.55 b	29.07 c	13.4 d	74.93 a	10.93 g	56.36 b

AN-801	4.34 c	3.56 b	3.43 b	31.37 c	13.4 d	69.93 b	15.24 e	60.83 a
AN-802	3.94 d	3.68 b	3.51 b	34.06 b	14.0 c	75.41 a	5.48 g	55.08 b
AN-803	2.29 i	2.63 d	2.62 d	10.14 e	15.4 b	69.21 b	14.09 f	49.12 c
AN-804	2.88 g	3.08 c	3.04 c	18.81 d	14.3 c	70.19 b	11.41 g	58.42 a
AN-808	3.01 f	2.13 f	1.82 g	7.54 e	13.5 d	76.17 a	10.05 g	55.16 b
AN-809	1.98 j	2.10 f	2.00 f	6.22 f	16.1 b	71.08 b	12.11 g	52.67 b
AN-810	2.00 j	1.86 h	1.79 g	5.86 f	13.1 d	67.27 b	16.09 e	53.48 b
BRS 226	6.21 a	5.29 a	5.08 a	104.02 a	11.1 e	59.71 e	32.51 a	45.79 d
CV (%)	10.41	9.25	9.95	36.6	13.86	5.91	20.43	8.35

[1]Means followed by the same letter in the column do not differ significantly from each other ($p \leq 0.01$)

BAG ID[2] – Identification of the accessions in the Germplasm Bank

LeCA[3] – Length cashew apple

WiCA[4] – Width cashew apple

DiCA[5] – Diameter cashew apple

WeCA[6] – Weight cashew apple

TSS[7] – Total soluble solids content

L[8] – Lightness

a[9] – Red/green coordinate (^{+}a indicates red and ^{-}a indicates green)

b[10] – Yellow/blue coordinate (^{+}b indicates yellow and ^{-}b indicates blue)

Table S6 Qualitative data of fruits (cashew nuts) from accessions of the Active Germplasm Bank of *A. occidentale* L. and *A. humile* A.St.-Hil.

BAG ID[1]	RPSA[2]	SNA[3]	SN[4]	SNB[5]	NS[6]	FN[7]	SSN[8]
AN-700	In front of apex	Intermediate	Rounded	Flattened	Oblong-ellipsoid	Rounded	Large
AN-701	In front of apex	Round	Angular	Flattened	Kidney-shaped	Rounded	Small
AN-702	In front of apex	Round	Angular	Flattened	Kidney-shaped	Flattened	Large
AN-703	In front of apex	Round	Rounded	Flattened	Kidney-shaped	Flattened	Large
AN-704	In front of apex	Round	Rounded	Flattened	Kidney-shaped	Rounded	Large
AN-705	In front of apex	Round	Angular	Flattened	Kidney-shaped	Flattened	Small
AN-706	In front of apex	Pointed	Angular	Obliquely flattened	Kidney-shaped	Rounded	Large
AN-707	In front of apex	Round	Angular	Flattened	Kidney-shaped	Flattened	Large
AN-708	In front of apex	Intermediate	Rounded	Flattened	Kidney-shaped	Flattened	Small
AN-709	In front of apex	Intermediate	Angular	Flattened	Oblong-ellipsoid	Flattened	Large
AN-710	In front of apex	Intermediate	Angular	Flattened	Oblong-ellipsoid	Flattened	Large
AN-711	In front of apex	Round	Angular	Flattened	Oblong-ellipsoid	Flattened	Large
AN-715	In front of apex	Round	Angular	Flattened	Oblong-ellipsoid	Flattened	Large
AN-716	In front of apex	Intermediate	Angular	Flattened	Kidney-shaped	Rounded	Large
AN-717	In front of apex	Round	Rounded	Flattened	Kidney-shaped	Flattened	Large
AN-718	In front of apex	Round	Angular	Flattened	Kidney-shaped	Flattened	Large
AN-719	In front of apex	Round	Rounded	Flattened	Oblong-ellipsoid	Bulging	Small
AN-800	In front of apex	Intermediate	Angular	Rounded	Kidney-shaped	Rounded	Large
AN-801	In front of apex	Intermediate	Rounded	Rounded	Kidney-shaped	Flattened	Large
AN-802	In line with apex	Round	Angular	Rounded	Kidney-shaped	Flattened	Large

AN-803	In front of apex	Intermediate	Angular	Flattened	Kidney-shaped	Flattened	Large
AN-804	In front of apex	Round	Rounded	Flattened	Kidney-shaped	Flattened	Small
AN-808	In front of apex	Round	Angular	Rounded	Kidney-shaped	Flattened	Small
AN-809	In front of apex	Round	Angular	Rounded	Kidney-shaped	Flattened	Small
AN-810	In front of apex	Round	Angular	Rounded	Kidney-shaped	Flattened	Large
BRS 226	In line with apex	Round	Angular	Rounded	Kidney-shaped	Flattened	Large

BAG ID[1] – Identification of the accessions in the Germplasm Bank

RPSA[2] – Relative position of suture and apex

SN[3] – Suture of nut

SNA[4] – Shape of nut apex

SNB[5] – Shape of nut base

NS[6] – Nut shape

FN[7] – Flanks of nut

SSN[8] – Stylar scar on nut

Table S7. Scott-Knott analysis based on agronomic data of fruits (cashew nuts) from accessions of the Active Germplasm Bank of *A. occidentale* L. and *A. humile* A.St.-Hil.

BAG ID[2]	NL[3] (cm)	NW[4] (cm)	NT[5] (cm)	WeN[6] (g)	*L*[7]	*a*[8]	*b*[9]
AN-700	1.51 h [1]	1.69 d	1.05 f	1.20 f	57.99 a	-2.53 f	19.21 a
AN-701	1.62 g	1.49 f	1.08 e	1.46 e	54.95 b	-0.62 e	16.87 b
AN-702	1.76 f	1.54 e	1.02 f	1.56 e	50.97 c	0.54 d	20.09 a
AN-703	1.60 g	1.29 h	0.92 h	1.10 f	52.48 c	-1.61 f	14.96 c
AN-704	1.82 f	1.56 e	1.12 e	1.74 e	51.06 c	1.16 c	20.15 a
AN-705	1.84 f	1.40 g	0.98 g	1.38	52.82 c	1.19 c	19.61 a
AN-706	1.69 g	1.45 f	0.97 g	1.55 e	58.29 a	-1.28 e	17.84 b
AN-707	1.43 h	1.48 f	0.99 g	1.50 e	50.20 c	-0.20 d	15.26 c
AN-708	1.67 g	1.62 e	1.10 e	1.52 e	48.74 d	3.49 a	16.52 b
AN-709	2.01 e	1.58 e	0.92 h	1.68 e	47.55 d	-2.16 f	13.53 d
AN-710	1.59 g	1.48 f	1.10 e	1.36 f	41.66 f	-1.87 f	14.16 c
AN-711	1.61 g	1.61 e	1.04 f	1.33 f	55.67 b	-1.06 e	20.19 a
AN-715	1.58 g	1.18 i	0.83 i	1.10 f	59.43 a	-2.50 f	16.58 b
AN-716	1.77 f	1.75 d	1.18 d	1.81 e	50.21 c	1.45 c	14.86 c
AN-717	1.60 g	1.46 f	1.03 f	1.37 f	48.88 d	1.43 c	15.74 c
AN-718	1.55 g	1.39 g	0.90 h	1.03 f	44.68 e	0.93 c	18.37 b
AN-719	1.5 g	1.38 g	0.93 h	1.44 e	48.02 d	1.54 c	13.12 d
AN-800	2.36 d	1.74 d	1.18 d	3.48 c	41.79 f	0.46 d	13.07 d

AN-801	2.64 b	1.98 c	1.48 b	4.67 b	35.30 h	1.20 c	7.84 e
AN-802	2.50 c	2.23 b	1.20 d	4.37 b	38.49 g	0.89 c	11.70
AN-803	2.06 e	1.59 e	1.12 e	2.77 d	39.13 g	1.87 b	12.67 d
AN-804	2.09 e	1.67 d	1.37 c	3.00 d	45.09 e	1.96 c	12.42 d
AN-808	1.61 g	1.35 g	1.02 f	1.24 f	41.31 f	-1.64 f	13.08 d
AN-809	1.86 f	1.36 g	1.07 e	1.75 e	44.33 e	1.32 c	13.01 d
AN-810	1.85 f	1.36 g	1.08 e	1.60 e	42.04 f	-1.98 f	13.08 d
BRS 226	3.63 a	2.99 a	2.22 a	10.93 a	57.13 a	0.19 d	13.34 d
CV (%)	6.68	7.50	7.74	21.37	8.40	6450	19.25

[1]Means followed by the same letter in the column do not differ significantly from each other ($p \leq 0.01$)

BAG ID[2] – Identification of the accessions in the Germplasm Bank

NL[3] – Nut length

NW[4] – Nut width

NT[5] – Nut thickness

WeN[6] – Weight of nut

L[7]– Lightness

a[8] – Red/green coordinate (^{+}a indicates red and ^{-}a indicates green)

b[9] – Yellow/blue coordinate (^{+}b indicates yellow and ^{-}b indicates blue)

Table S8 Nucleotide diversity index (π) between populations for nuclear and plastidial datasets

Populations	Nucleotide Diversity (π)	
	Nuclear Region	Plastidial Region
A. occidentale cultivated	0.028231 +/- 0.017896	0.089319 +/- 0.051699
A. occidentale Bom Jesus	0.009852 +/- 0.007094	0.000976 +/- 0.001127
A. occidentale Alvorada	0.032895 +/- 0.033707	0.002598 +/- 0.002756
A. occidentale Currais	0.002204 +/- 0.002264	0.000000 +/- 0.000000
A. occidentale Parnaíba	0.052893 +/- 0.040156	0.082844 +/- 0.061894
A. occidentale Santa Rosa	0.000000 +/- 0.000000	0.000000 +/- 0.000000
A. humile Bom Jesus	0.000000 +/- 0.000000	0.003910 +/- 0.004070
A. humile Brasília		0.000000 +/- 0.000000
A. humile Uruçuí		0.000000 +/- 0.000000

Table S9 Population differentiation based on allele frequency variance, pairwise F_{ST} for the nuclear dataset

Populations	*A. occidentale* cultivated	*A. occidentale* Bom Jesus	*A. occidentale* Alvorada	*A. occidentale* Currais	*A. occidentale* Parnaíba	*A. occidentale* Santa Rosa	*A. humile* Bom Jesus
A. occidentale cultivated	0.00000						
A. occidentale Bom Jesus	0.17088	0.00000					
A. occidentale Alvorada	0.12387	0.16418	0.00000				
A. occidentale Currais	0.14793	-0.11392	0.20755	0.00000			

A. occidentale Parnaíba	-0.00236	0.23596	0.06746	0.18702	0.00000		
A. occidentale Santa Rosa	-0.26415	0.46667	-0.28000	0.80000	-0.36735	0.00000	
A. humile Bom Jesus	-0.45652	-0.33333	-0.77778	0.20000	-0.67500	1.00000	0.00000

Table S10 Population differentiation based on allele frequency variance, pairwise F_{ST} for the plastidial dataset

Populations	*A. occidentale* cultivated	*A. occidentale* Bom Jesus	*A. occidentale* Alvorada	*A. occidentale* Currais	*A. occidentale* Parnaíba	*A. occidentale* Santa Rosa	*A. humile* Bom Jesus	*A. humile* Brasília	*A. humile* Uruçuí
A. occidentale cultivated	0.00000								
A. occidentale Bom Jesus	-0.19191	0.00000							
A. occidentale Alvorada	-0.20337	0.00000	0.00000						
A. occidentale Currais	-0.78225	0.33333	-0.33333	0.00000					
A. occidentale Parnaíba	0.01187	-0.09512	-0.08190	-0.81592	0.00000				
A. occidentale Santa Rosa	-0.95350	0.92000	0.68000	1.00000	-0.82957	0.00000			
A. humile Bom Jesus	0.61078	0.99755	0.99020	1.00000	0.74021	1.00000	0.00000		
A. humile Brasília	-0.21322	0.65714	0.50000	0.37500	-0.03496	0.37500	0.98783	0.00000	
A. humile Uruçuí	-0.73897	0.91667	0.68000	1.00000	-0.58351	1.00000	1.00000	0.41176	0.00000

Table S11 Significance analysis based on allele frequency variance, pairwise F_{ST} for the nuclear dataset

Populations	*A. occidentale* cultivated	*A. occidentale* Bom Jesus	*A. occidentale* Alvorada	*A. occidentale* Currais	*A. occidentale* Parnaíba	*A. occidentale* Santa Rosa	*A. humile* Bom Jesus
A. occidentale cultivated	*						
A. occidentale Bom Jesus	0.12355+-0.0031	*					
A. occidentale Alvorada	0.20404+-0.0035	0.33997+-0.0048	*				
A. occidentale Currais	0.14909+-0.0038	0.99990+-0.0000	0.19226+-0.0043	*			
A. occidentale Parnaíba	0.55430+-0.0053	0.02891+-0.0017	0.39244+-0.0055	0.10256+-0.0028	*		
A. occidentale Santa Rosa	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	*	
A. humile Bom Jesus	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	*

Table S12 Significance analysis based on allele frequency variance, pairwise F_{ST} for the plastidial dataset

Populations	*A. occidentale* cultivated	*A. occidentale* Bom Jesus	*A. occidentale* Alvorada	*A. occidentale* Currais	*A. occidentale* Parnaíba	*A. occidentale* Santa Rosa	*A. humile* Bom Jesus	*A. humile* Brasília	*A. humile* Uruçuí
A. occidentale cultivated	*								
A. occidentale Bom Jesus	0.28245+-0.0043	*							
A. occidentale Alvorada	0.31343+-0.0044	0.99990+-0.0000	*						
A. occidentale Currais	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	*					
A. occidentale Parnaíba	0.34670+-0.0045	0.50460+-0.0050	0.50411+-0.0048	0.49856+-0.0051	*				
A. occidentale Santa Rosa	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	*			
A. humile Bom Jesus	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	*		
A. humile Brasília	0.26819+-0.0042	0.33215+-0.0048	0.33106+-0.0044	0.33759+-0.0046	0.51104+-0.0057	0.34086+-0.0049	0.33521+-0.0042	*	
A. humile Uruçuí	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	0.99990+-0.0000	*

Table S13 Analysis of Molecular Variance (AMOVA) for seven populations of the nuclear dataset

Source of variation	*d.f.*	Sum of squares	Variance components	Percentage of variation (%)
Among groups	2	10.613	-0.36860 Va	-6.63
Among populations within groups	4	27.026	0.55368 Vb	9.96
Within populations	12	64.467	5.37222 Vc	96.67
Total	18	102.105	5.55730	
Fixation Indices				
F_{SC}[1]	0.108560			
F_{ST}[2]	0.03330			
F_{CT}[3]	-0.06633			

F_{SC}[1] – Fixation index among subpopulations within clusters

F_{ST}[2] – Fixation index among subpopulations relative to the total population

F_{CT}[3]– Fixation index among clusters relative to the total population

Table S14 Analysis of Molecular Variance (AMOVA) for nine populations of the plastidial dataset

Source of variation	*d.f.*	Sum of squares	Variance components	Percentage of variation (%)
Among groups	1	123.007	-0.06342 Va	-0.06
Among populations within groups	7	865.967	15.05904 Vb	13.82
Within populations	10	939.500	93.95000 Vc	86.24

Total	18	1928.474	108.94561
Fixation Indices			
F_{SC}[1]	0.13814		
F_{ST}[2]	0.13764		
F_{CT}[3]	-0.00058		

F_{SC}[1] – Fixation index among subpopulations within clusters

F_{ST}[2] – Fixation index among subpopulations relative to the total population

F_{CT}[3]– Fixation index among clusters relative to the total population

Supplementary Material - Figures

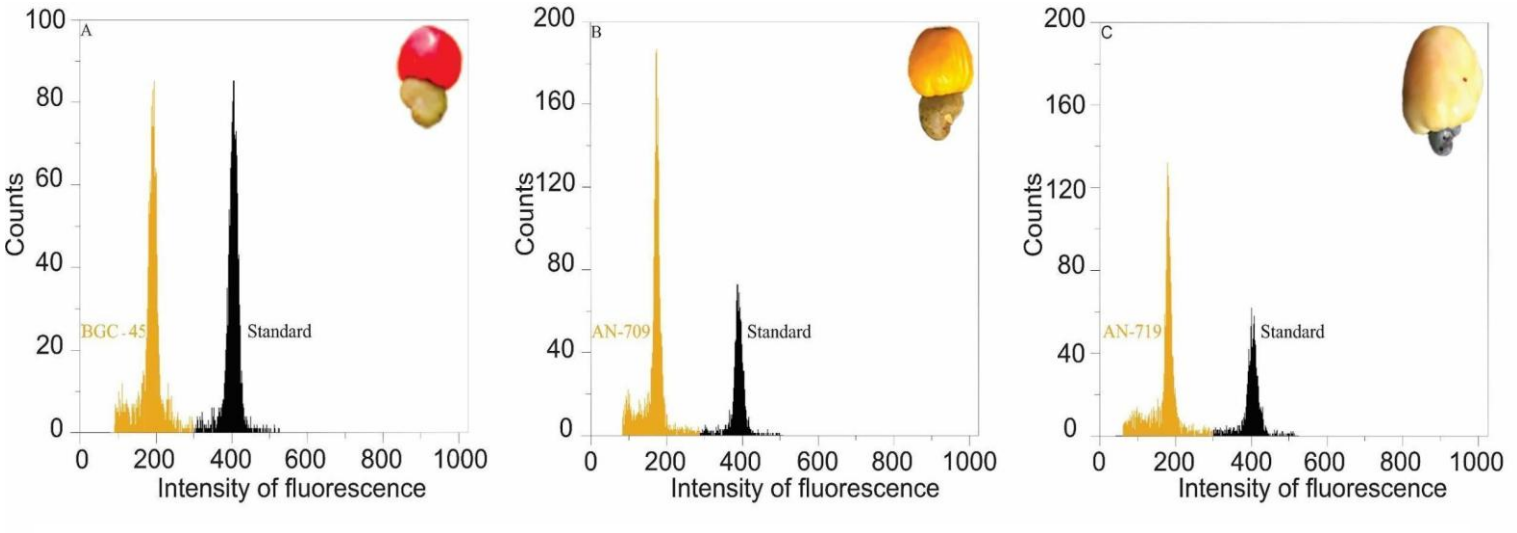


Figure S1 Histograms of relative fluorescence intensities of nuclei isolated from samples of *Anacardium*. A) *A. occidentale* previously identified as *A. othonianum* (BGC-45), B) *A. occidentale* (AN-709), and C) *A. occidentale* (AN-719) against the internal standard *Solanum lycopersicum*, var. Stupicke (control)

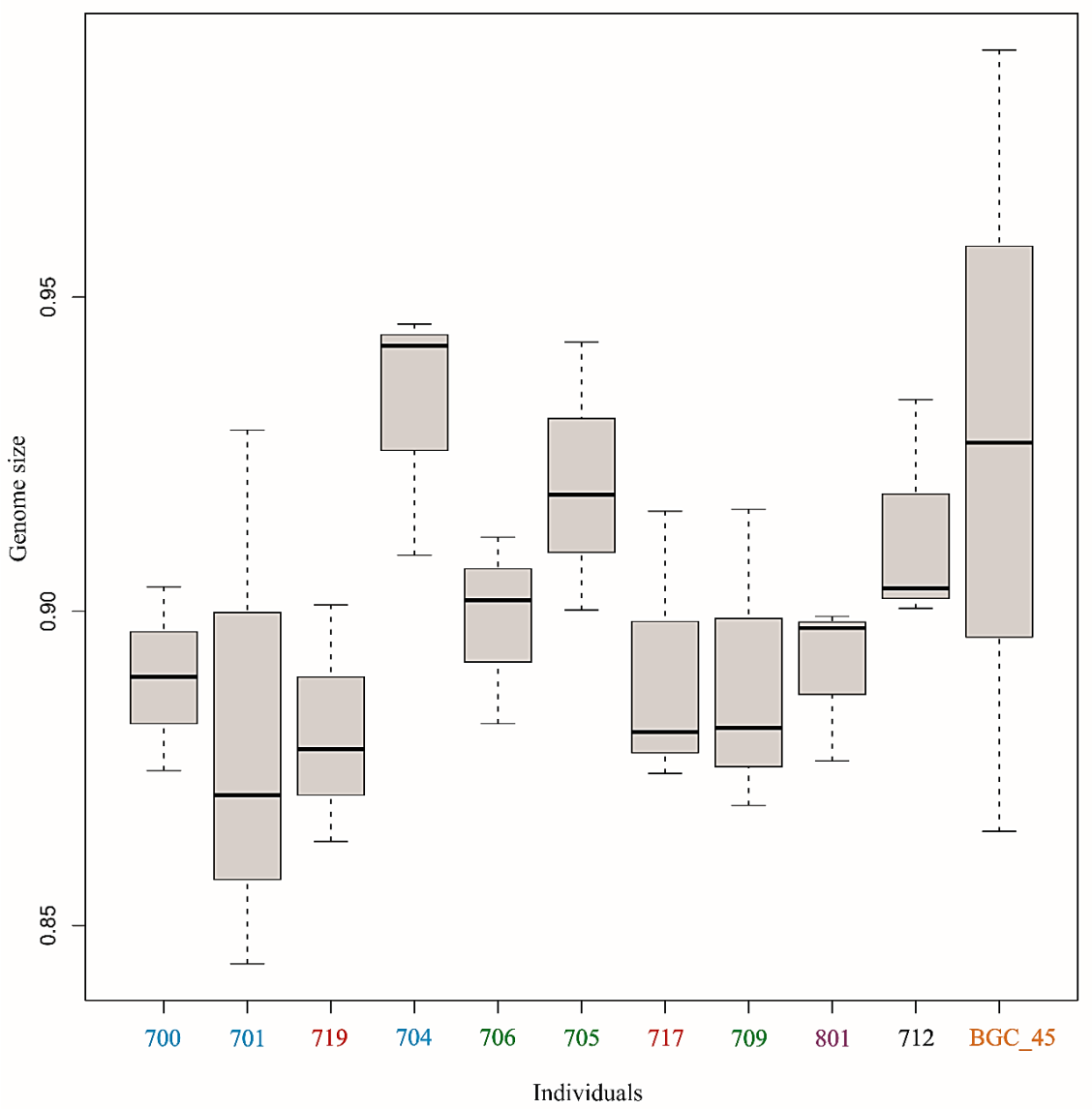


Figure S2 Distribution of genome size (pg/2C) of *A. occidentale* and *A. humile* accessions across five distinct locations: Bom Jesus highlighted in blue; Alvorada do Gurguéia in green; Currais in red; Parnaíba in purple; Luziânia in orange

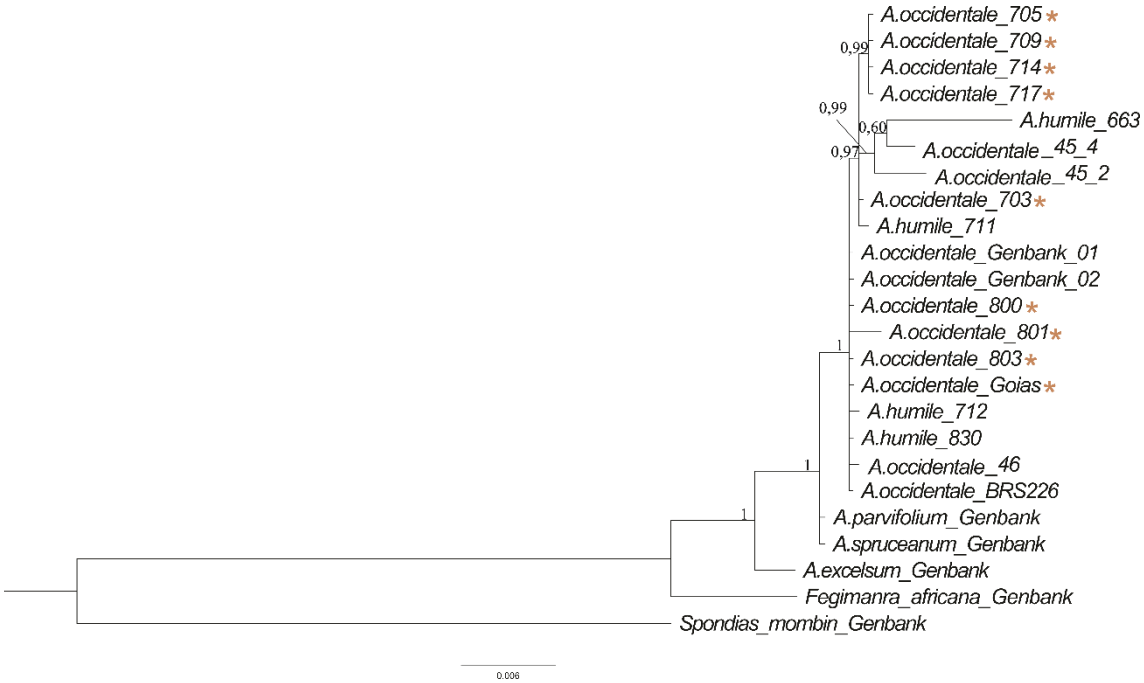


Figure S3 Bayesian tree of *Anacardium* derived from the analysis of plastidial regions *matK, rps16*, and *trnL-F*. The posterior probability (PP) values are shown for each node, and asterisks indicate tree accessions of *Anacardium occidentale* with cashew-type fruits and cashew apples, i.e., the possible hybrids

Figure S4 Alignment of the cashew reference plastome (*A. occidentale*) with NGS sequences

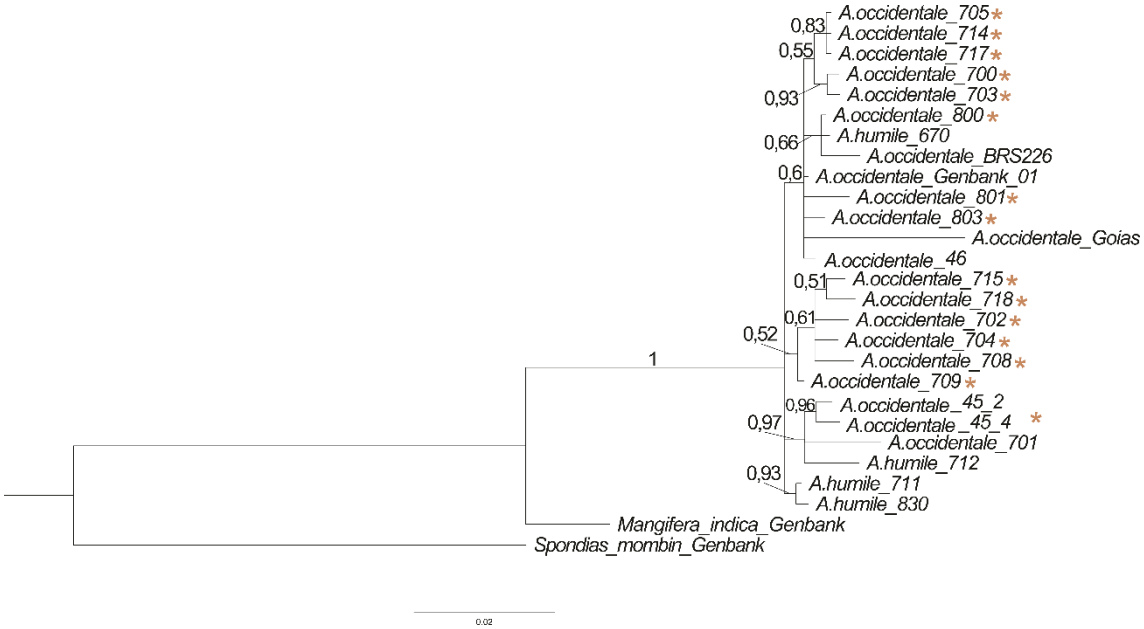


Figure S5 Bayesian tree of *Anacardium* derived from the analysis of the plastidial region *ycf1*. The posterior probability (PP) values are shown for each node, and asterisks indicate tree accessions of *Anacardium occidentale* with cashew-type fruits and cashew apples, i.e., the possible hybrids

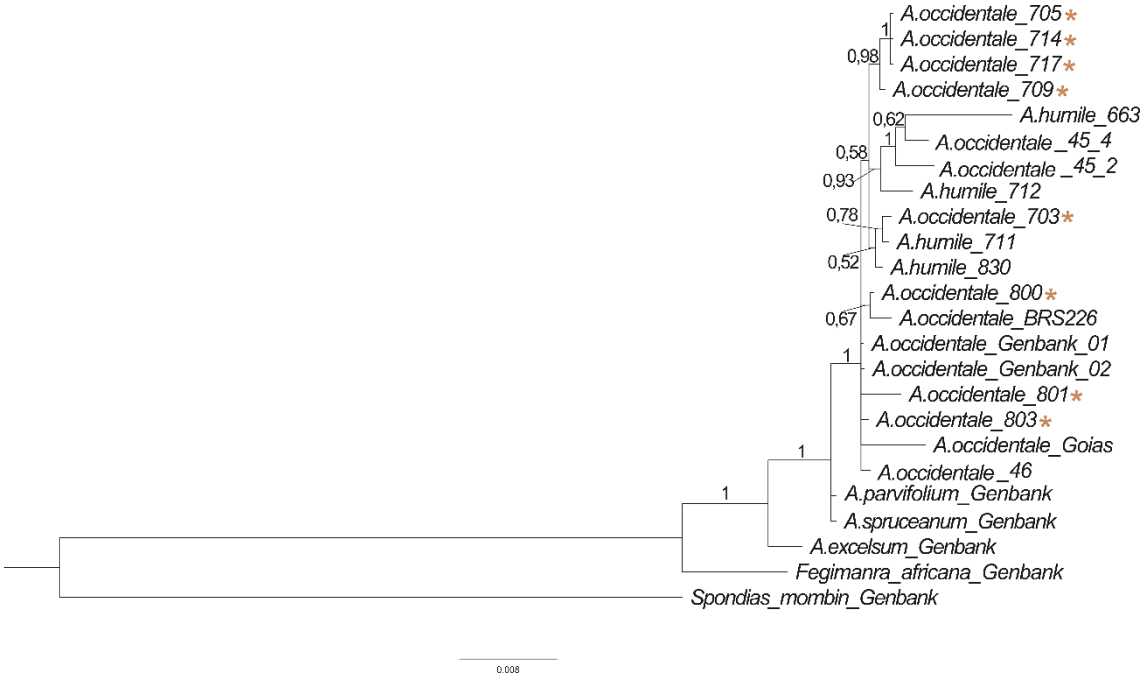


Figure S6 Bayesian tree of *Anacardium* derived from the analysis of plastidial regions *matK, trnL-F, rps16,* and *ycf1*. The posterior probability (PP) values are shown for each node, and asterisks indicate tree accessions of *Anacardium occidentale* with cashew-type fruits and cashew apples, i.e., the possible hybrids

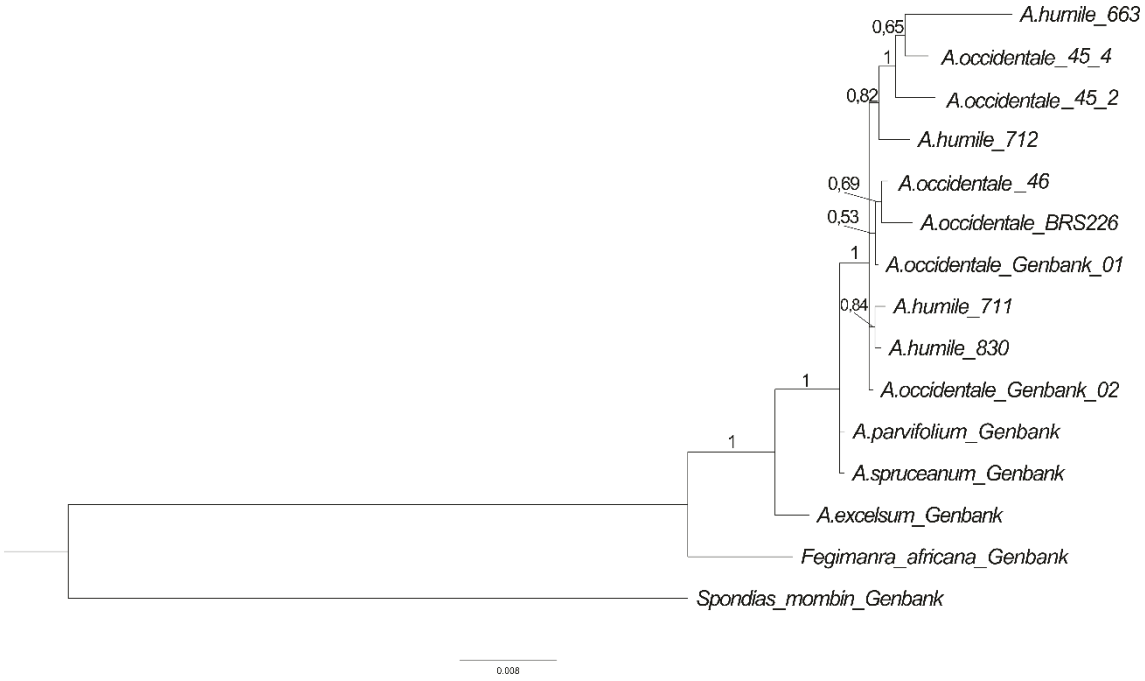


Figure S7 Bayesian tree of *Anacardium* derived from the analysis of plastidial regions *matK, trnL_trnF*, *rps16*, and *ycf1*, excluding putative hybrids. Posterior probability (PP) values are shown on each branch node.

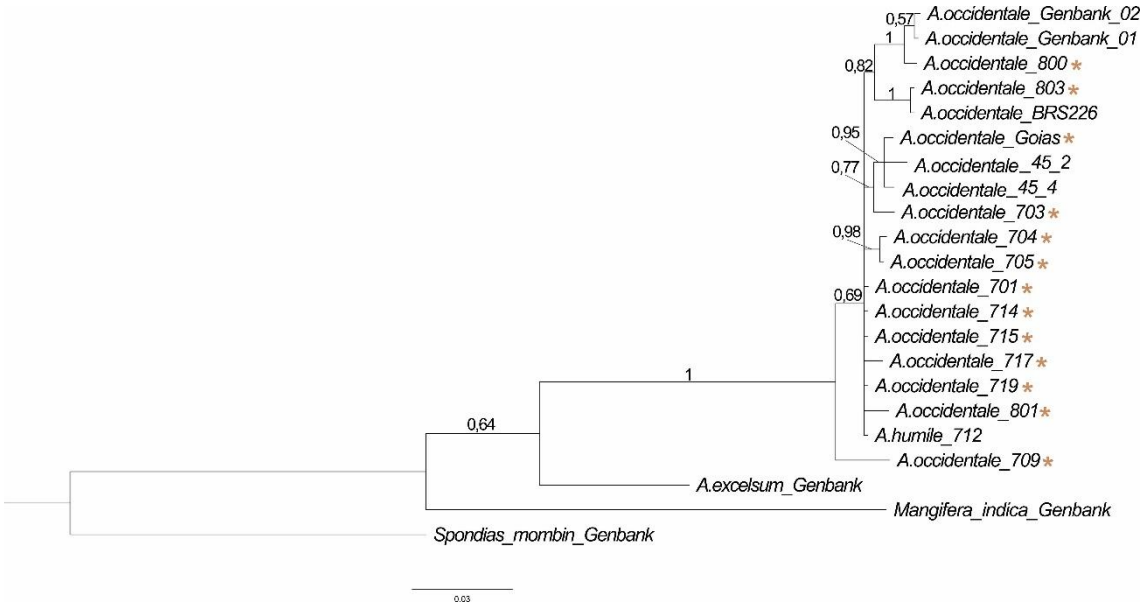


Figure S8 Bayesian tree of *Anacardium* derived from the analysis of the nuclear *ITS* region. The posterior probability (PP) values are shown for each node, and asterisks indicate tree accessions of *Anacardium occidentale* with cashew-type fruits and cashew apples, i.e., the possible hybrids

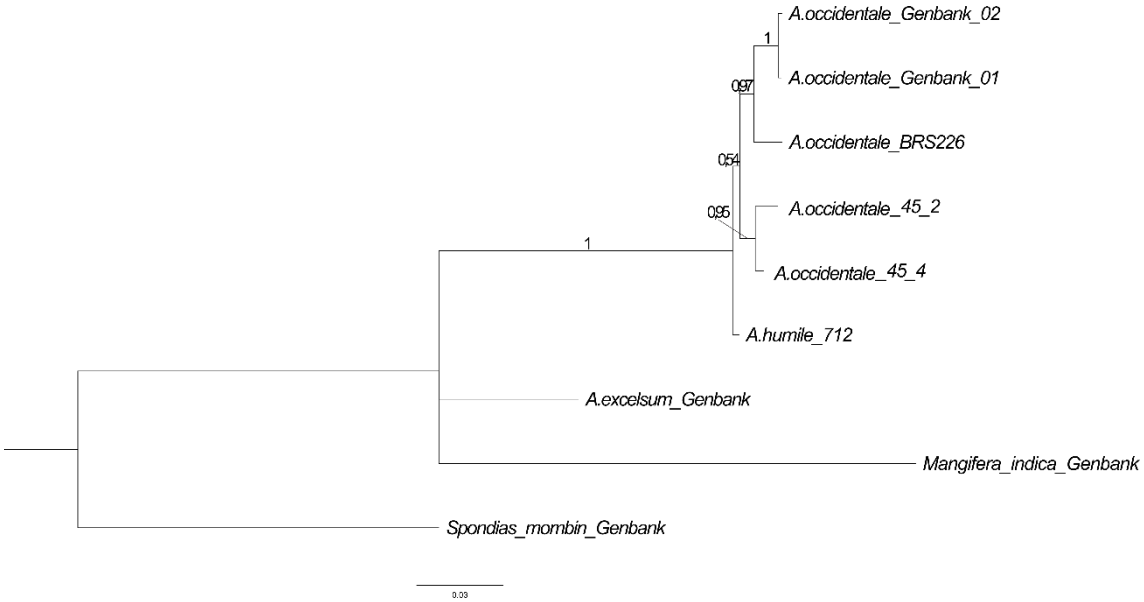


Figure S9 Bayesian tree of *Anacardium* derived from the analysis of the nuclear *ITS* region, excluding putative hybrids. Posterior probability (PP) values are shown on the branches

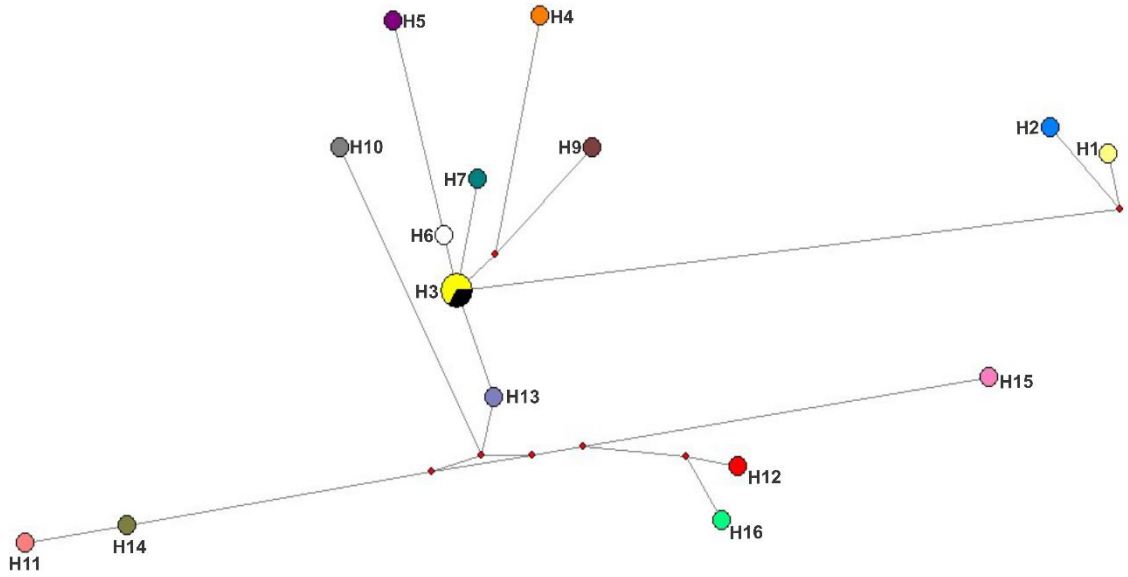


Figure S10 Haplotype distribution patterns of *A. humile* (H13) and *A. occidentale* (H1-H12 and H14-H16) for nuclear data. The colours in each pie chart correspond to the haplotypes, and the size of the circle represents the proportion of each haplotype

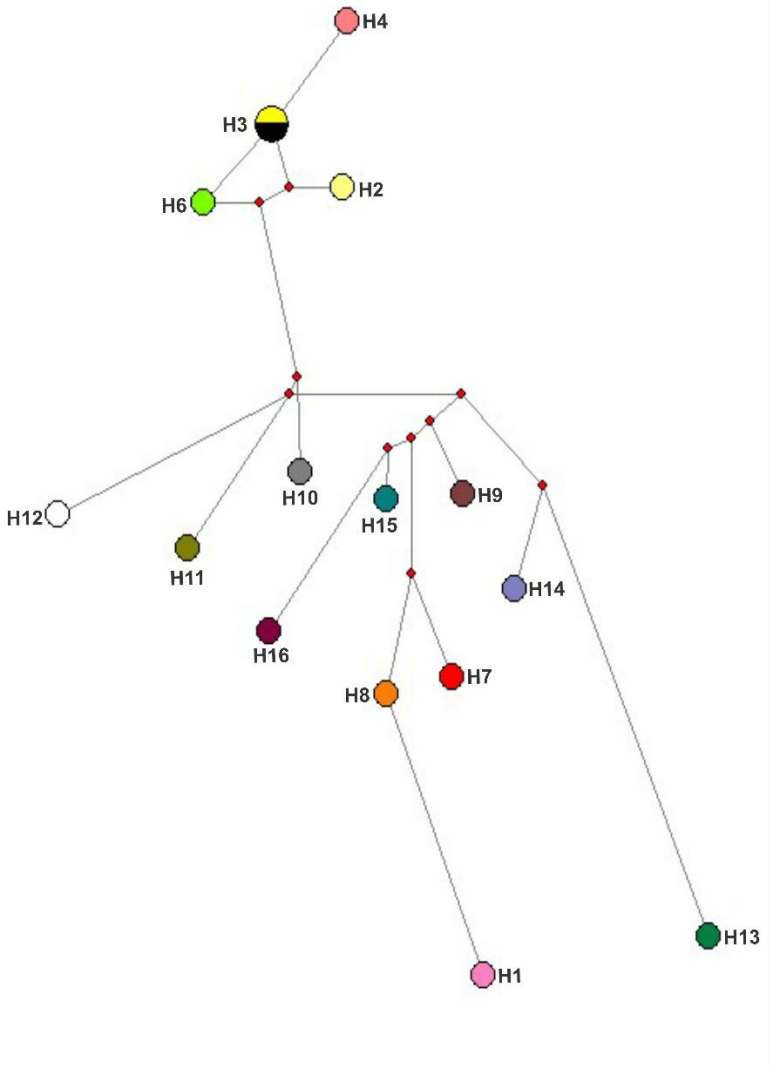


Figure S11 Haplotype distribution patterns of *A. humile* (H10, H11 and H12) and *A. occidentale* (H1-H9 and H13-H16) for plastidial data. The colours in each pie chart correspond to the haplotypes, and the size of the circle represents the proportion of each haplotype

4. CONSIDERAÇÕES FINAIS

A avaliação agronômica dos acessos avaliados de *Anacardium occidentale* e *A. humile*, empregando descritores quali- e quantitativos para o pedúnculo (pseudofruto) e a castanha (fruto verdadeiro), revelou a existência de variabilidade morfológica e genética.

Os acessos avaliados apresentaram sequências relativamente conservadas para as regiões plastidiais e nuclear.

As relações filogenéticas para as espécies de *Anacardium* avaliadas neste estudo permanecem incertas, uma vez que não foi possível separar taxonomicamente as espécies e confirmar a existência de possíveis híbridos.

As espécies aqui analisadas não diferiram citogeneticamente no número de cromossomos, padrão de bandas CMA/DAPI, distribuição de sítios de DNAr e conteúdo de DNA.

A similaridade genômica e morfológica entre *Anacardium occidentale* e *A. humile* sugere especiação insipiente ou hibridização natural entre elas e talvez o papel da domesticação na diferenciação previamente reportada entre elas.

REFERÊNCIAS

ALIYU, O. M.; AWOPETU, J. A. Chromosome studies in Cashew (*Anacardium occidentale* L.) **African Journal of Biotechnology**, v.6, n.2, p.131-136, 2007. DOI:10.4314/AJB.V6I2.56120.

ALIYU, O. M. Analysis of absolute nuclear DNA content reveals a small genome and intra-specific variation in Cashew (*Anacardium occidentale* L.), Anacardiaceae. **Silvae Genetica**, v.63, p.285-292, 2014. DOI: 10.1515/sg-2014-0036.

ALMEIDA, C. C. S.; CARVALHO, P. C. L.; GUERRA, M. Karyotype differentiation among *Spondias* species and the putative hybrid Umbu-cajá (Anacardiaceae), **Botanical Journal of the Linnean Society***,* v.155, n.4, p.541-547, 2007. DOI:10.1111/j.1095-8339.2007. 00721.x.

ALLSHIRE, R. C.; MADHANI, H. D. Ten principles of heterochromatin formation and function. **Nature Reviews Molecular Cell Biology**, v.19, n.4, p. 229-244, 2018. DOI: 10.1038/nrm.2017.119.

AMABILE, R. F.; VILELA, M. S.; PEIXOTO, J. R. Melhoramento de plantas: variabilidade genética, ferramentas e mercado. Brasília, DF: **Proimpress**; Sociedade Brasileira de Melhoramento de Plantas, 2018. p.108.

AMORIM, B. S.; VASCONCELOS, T. N. C.; SOUZA, G.; ALVES, M.; ANTONELLI, A.; LUCAS, E. Advanced understanding of phylogenetic relationships, morphological evolution and biogeographic history of the mega-diverse plant genus Myrcia and its relatives (Myrtaceae: Myrteae). **Molecular Phylogenetics and Evolution,** v.138, p.65-88, 2019. DOI:10.1016/j.ympev.2019.05.014.

ARIYARATHNE, M.; YAKANDAWALA, D.; BARFUSS, M.; HECKENHAUER, J.; SAMUEL, R. Molecular phylogeny and chromosomal evolution of endemic species of Sri Lankan Anacardiaceae. **Journal of the National Science Foundation of Sri Lanka**, v.48, n.3, p.289-303, 2020. DOI: 10.4038/jnsfsr.v48i3.9368.

BADAEVA, E. D.; RUBAN, A. S.; ALIYEVA-SCHNORR, L.; MUNICIO, C.; HESSE, S.; HOUBEN, A. In Situ Hybridization to Plant Chromosomes. In: Shakoori, A. R. Fluorescence In Situ Hybridization (FISH) and Its Applications. **Chromosome Structure and Aberrations**, 2017. p.343-367. DOI: 10.1007/978-81-322-3673-3_16.

BARROS-SILVA, A. E.; GUERRA, M. CMA/DAPI Banding of Plant Chromosomes. In: Heitkam, T.; Garcia, S. **Plant Cytogenetics and Cytogenomics**. Methods in Molecular Biology, 2023.

BORGES, A. N. C.; LOPES, A. C. A.; BRITTO, F. B.; VASCONCELOS, L. F. L.; LIMA, P. S. C. Genetic diversity in a cajuí (*Anacardium* spp.) germplasm bank as determined by ISSR markers. **Genetics and Molecular Research**, v.17, n.4, 2018.

BORÉM, A.; MIRANDA, G. V.; FRITSCHE-NETO, R. **Melhoramento de plantas**. 8. ed. São Paulo: Oficina de Textos, 2021.

BRAMMER, S. P.; JÚNIOR, N. A. Estudos citogenéticos em triticale: caracterização de germoplasma e apoio ao melhoramento genético. In: BARBOSA, F. C. **Genética**: hereditariedade e características. Piracanjuba: Conhecimento Livre, 2022. p.12-26.

BRAINER, M. S. DE C. P.; VIDAL, M. DE F. Cajucultura. **Caderno Setoria ETENE**. Fortaleza: Banco do Nordeste do Brasil, n.114, 2020.

CARNEIRO, L. A.; SILVA, L. S.; GOMES, M. F. C.; SANTOS, M. F.; VALENTE, S. E. S.; GOMES, R. L. F.; COSTA, M. F. Morphological characterization and genetic divergence of a cashew population in Floriano, Piauí, Brazil. **Genetics and Molecular Research**, v.18, n.3, 2019. DOI: 10.4238/gmr18348.

COLLI, G. R.; VIEIRA, C. R.; DIANESE, J. C. Biodiversity and conservation of the Cerrado: recent advances and old challenges. **Biodiversity and Conservation**, v.29, p.1465-1475, 2020. DOI: 10.1007/s10531-020-01967-x.

CONAB-COMPANHIA NACIONAL DE ABASTECIMENTO. **Boletim da Sociobiodiversidade**, Brasília, DF, v. 7, n. 1, 2023.

CORDEIRO, J. M. P.; KAEHLER, M.; SOUZA, L. G.; FELIX, L. P. Heterochromatin and numeric chromosome evolution in Bignoniaceae, with emphasis on the Neotropical clade Tabebuia alliance. **Genetics and Molecular Biology**, v.27, n.43, 2020. DOI: 10.1590/1678-4685-GMB-2018-0171.

COSTA, M. F.; SILVA, L. DE S. S.; CARNEIRO, L. A.; GOMES, M. F. C.; DOS SANTOS, M.F.; VALENTE, S. E. S. Characterization and genetic diversity of the cashew natural population through morphological descriptors. **Revista Brasileira De Agropecuária Sustentável**, v.10, n.1, p. 205–212, 2020. DOI: https://doi.org/10.21206/rbas.v10i1.9782

COTA, L. G.; MOREIRA, P. A.; BRANDÃO, M. M.; ROYO, V. A.; MELO-JUNIOR, A. F.; MENEZES, E. V. O. Structure and genetic diversity of *Anacardium humile* (Anacardiaceae): a tropical shrub. **Genetics and Molecular Research**, v.16, p.1-13, 2017. DOI: 10.4238/gmr16 03977 8.

CUNHA, R. M. S. **Filogenia Molecular em *Anacardium* (Anacardiaceae): utilizaçãodo gene da subunidade pequena do RNA ribossômico (SSU rRNA).** 2002. 78f. Dissertação (Mestrado em Bioquímica) - Universidade Federal do Ceará, Fortaleza, 2002.

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

CRITICAL ECOSYSTEM PARTNERSHIP FUND (CEPF), "Ecosystem Profile: Cerrado Biodiversity Hotspot", 2016. Disponível em: < https://www.cepf.net/>. Acesso em 30 de abril de 2023.

CRESPO, M. F. V.; SOUZA, L. I DE. **Cajuí: boas práticas e manejo sustentável**. Parnaíba: Sieart, 2014.

CRUZ, C. D.; FERREIRA, M. F.; PESSONI, A. L. **Biometria aplicada ao estudo da diversidade genética**, v.1, p.620, 2011.

DANIELL, H.; LIN, C.S.; YU, M.; CHANG, W.J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. **Genome Biology**, v.23, p.1-29, 2016. DOI: https://doi.org/10.1186/s13059-016-1004-2

DANTAS, L. G.; GUERRA, M. Chromatin differentiation between *Theobroma cacao* L. and *T. grandiflorum* Schum. **Genetics and Molecular Biology***,* v.33*,* p.94 -98, 2010.

DARLINGTON, C. D.; JANAKI-AMMAAL, E. K. **Chromosome Atlas of Cultivated Plants**. London, 1945.

DE BARROS, D.; MONTENEGRO, C.; GOMES, M.; FERRAZ, M. E.; MIOTTO, S. T. S.; PEDROSA-HARAND, A. Cytogenetic characterization and karyotype evolution in six *Macroptilium* species (Leguminosae). **Genome**. e-First, 2023. DOI: 10.1139/gen-2022-0101.

DJOLOSSÈ, N. K.; ADOUKONOU-SAGBADJA, H.; GBÈMAVO, C. D. S. J.; KODJO, S.; BADOU, A.; MALIKI, R.; AHOYO-ADJOVI, N. R. Agro-morphological characterization of preselected cashew (*Anacardium occidentale* L.) mother trees in Benin farmer's plantations.

Journal of Agriculture and Environment for International Development, v.113, n.1, p.17-34, 2019. DOI: 10.12895/jaeid.20191.841.

DOS SANTOS, J. O.; MAYO, S. J.; BITTENCOURT, C. B.; DE ANDRADE, I. M. Genetic diversity in wild populations of the restinga ecotype of the cashew (*Anacardium occidentale*) in coastal Piauí, Brazil**. Plant Systematics and Evolution**, v.305, p.913-924, 2019. DOI: 10.1007/s00606-019-01611-4.

ELLEGREN, H., GALTIER, N. Determinants of genetic diversity. **Nature Reviews Genetics**, v.17, p.422-433, 2016. DOI: 10.1038/nrg.2016.58.

FLORA DO BRASIL, 2020. Jardim Botânico do Rio de Janeiro. Disponível em: < http://ala-bie.sibbr.gov.br/>. Acesso em 30 julho 2021.

FU, YB. Understanding crop genetic diversity under modern plant breeding. **Theoretical and Applied Genetics**, v.128, p.2131-2142, 2015. DOI: 10.1007/s00122-015-2585-y.

GIARETTA, A.; AMORIM, B. S.; SANO, P. T.; SOUZA, G.; LUCAS, E. Phylogenetic Placement of New Species with Fused Calyx Reveals Homoplastic Character in *Eugenia* (Myrtaceae). **Systematic Botany**, v.44, n.1, p.66-73, 2019. DOI:10.1600/036364419X697903.

GILL, B. S.; BIR, S. S.; SINGHAL, V. K. In IOPB Chromosome number reports LXV. **Taxon**, v.28, p.630, 1979.

GILL, B. S.; SINGHAL, V. K.; BEDI, Y. S.; BIR, S. S. Cytological evolution in the woody taxa of Pachmarhi Hills. **Journal of Cytology and Genetics**, v.25, p.308-320, 1990.

GOMES, J. G.; OLIVEIRA, N. C.; NASCIMENTO, P. S. N.; PEREIRA, M. B. Componentes Principais no estudo da diversidade de cajuí na Sub-bacia do Rio dos Cochos. **Cadernos de Agroecologia**, v.6, n.2, p.1-5, 2011.

GOVINDARAJ, M.; VETRIVENTHAN, M.; SRINIVASAN, M. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. **Genetics Research International**, 2015. DOI: 10.1155/2015/431487.

GUERRA, M. Reviewing the chromosome nomenclature of Levan et al. **Brazilian Journal of genetics**, v.9, p.741-743,1986.

GUERRA, M. **Introdução à Citogenética Geral**. Rio de Janeiro: Guanabara, 1988.

GUERRA, M. Patterns of heterochromatin distribution in plant chromosomes. **Genetics and Molecular Biology**, v.23, p.1029-1041, 2000. DOI:10.1590/S1415-47572000000400049.

GUERRA, M. **FISH, Conceitos e Aplicações na Citogenética**. Ribeirão Preto: Sociedade Brasileira de Genética, 2004. p. 1-32.

GUERRA, M. Cytotaxonomy: The end of childhood, Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology: **Official Journal of the Societa Botanica Italiana**, v.146, n.3, p.703-710**,** 2012. DOI: 10.1080/11263504.2012.717973.

GUERRA, S.; GUERRA, M.; MENDES, S.; SOARES-FILHO, W. S.; PEDROSA-HARAND, A. Karyotype variability of sour orange (*Citrus* × *aurantium* L.) and the origin of its heteromorphic karyotype. **Tree Genetics & Genomes**, v.16, p.78, 2020. DOI: 10.1007/s11295-020-01471-x.

GUO, C.; LUO, Y.; GAO, L. M.; YI, T. S.; LI, H. T.; YANG, J. B.; LI, DE. Z. Phylogenomics and the flowering plant tree of life. **Journal of Integrative Plant Biology**, v.65, n.2, p.299-323, 2023.DOI:10.1111/jipb.13415.

HE, L.; ZHAO, H.; HE, J.; YANG, Z.; GUAN, B.; CHEN, K.; HONG, Q.; WANG, J.; LIU, J.; JIANG, J. Extraordinarily conserved chromosomal synteny of *Citrus* species revealed by chromosome-specific painting. **The Plant Journal**, v.103, n.6, p.2225-2235,2020. DOI: 10.1111/tpj.1489.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE).2023 Disponível em: <http://www.inpe.br/institucional/sobre_inpe/missao.php>. Acesso em: 23 jan. 2024.

INTERNATIONAL UNION FOR CONSERVATION OF NATURE. IUCN. 2023. Banco de dados online: Red List of threatened species. Disponível em: <https://www.iucnredlist.org/>. Acesso em: 29 mar. 2024.

JIANG J. Fluorescence in situ hybridization in plants: recent developments and future applications. **Chromosome research**, v.27, n.3, p.153-165, 2019. DOI: https://doi.org/10.1007/s10577-019-09607-z.

JOHNSON, D. The botany, origin and spread of cashew (*Anacardium occidentale* L.). **The Journal of Plantation Crops**, Kerala, v.1, n.2, p.1-7, 1973.

KADLEC, M.; BELLSTEDT, DU.; LE MAITRE, N. C.; PIRIE, M. D. Targeted NGS for species level phylogenomics: “made to measure” or “one size fits all”? **PeerJ**, v.5, p. e3569, 2017. DOI: 10.7717/peerj.3569.

KAPLI, P.; YANG, Z.; TELFORD, M. J. Phylogenetic tree building in the genomic age. **Nature Reviews Genetics**, v.21, p.428-444, 2020. DOI: 10.1038/s41576-020-0233-0.

KEW SCIENCE, 2017. *Anacardium* L. Plants of the World Online, Board of Trustees of the Royal Botanic Gardens, Kew. Disponível em: < https://powo.science.kew.org/>. Acesso em 30 abril 2022.

KOUAKOU, C. K.; KONAN, A.; ADOPO, A. N.; N'DA, H. A.; DJAHA, A. J.; MINHIBO, M.Y.; DJIDJI, A. H.; DOSSO, M.; N'GUESSAN, A. E. Agro-morphological characterization of cashew (*Anacardium occidentale* L.) ecotypes from North-Central of Côte d'Ivoire. **Fruits***,* 2018.

KUBALOVÁ, I.; CÂMARA, A. S.; CÁPAL, P.; BESEDA, T.; ROUILLARD, J. M.; KRAUSE, G. M.; HOLUŠOVÁ, K.; TOEGELOVÁ, H.; HIMMELBACH, A.; STEIN, N.; HOUBEN, A.; DOLEŽEL, J.; MASCHER, M.; ŠIMKOVÁ, H.; SCHUBERT, V.; Helical coiling of metaphase chromatids, **Nucleic Acids Research***,* v.51, p.2641-2654, 2023. DOI: 10.1093/nar/gkad028.

LAMBERS, H.; DE BRITTO COSTA, P.; OLIVEIRA, R.S.; SILVEIRA, F. A. O. Towards more sustainable cropping systems: lessons from native Cerrado species. **Theoretical and Experimental Plant Physiology,** v. 32, p.175-194 (2020). DOI: https://doi.org/10.1007/s40626-020-00180-z

LEVIN, D. A. The role of chromosomal change in plant evolution. Oxford series in **Ecology and Evolution**. Ed. PH Harvey and RM May. New York. p.228, 2002.

LIMA, J. R. Caracterização físico-química e sensorial de hambúrguer vegetal elaborado à base de caju**. Ciência e Tecnologia de Alimentos**, v.32, 2008. DOI: https://doi.org/10.1590/S1413-70542008000100028

MACHADO, O. Estudos novos sobre uma planta velha – o cajueiro (*Anacardium occidentale* L.) **Rodriguesia**, v.8, p.19-48,1944.

MITCHELL, J. D.; MORI, S. A. The cashew and its relatives (*Anacardium*: Anacardiaceae), Memoirs of the New York Botanical Garden. 42 stedn. **NYBG Press,** New York, 1987.

MATOS-FILHO, C.H.A.; DE CARVALHO-JUNIOR, J.E.V.; COSTA, G. DO N. et al. Estimation of the extent of phenotypic diversity among cashew (Anacardium occidentale L.) genotypes based on agro-morphological and physicochemical traits. **Plant Genetic Resources**, v.20, n.3, p.230-238, 2023. DOI:10.1017/S1479262123000102. 116

MITCHELL, J. D.; PELL, S. K.; BACHELIER, J. B.; WARSCHEFSKY, E. J.; JOYCE, E. M.; CANADELL, L. C.; SILVA-LUZ, C. L.; COIFFARD, C. Neotropical Anacardiaceae (cashew family). **Brazilian jornal of botany**, v.45, p.139-180, 2022. DOI: 10.1007/s40415-022-00793-5.

MONTEIRO, A. O.; JUNQUEIRA, A. M. R. **Valorização do cajuzinho-do-cerrado: memória involuntária e memória gustativa**. 2018, 120 f. Dissertação. (Mestrado em Agronegócio) – Faculdade de Agronomia e Medicina Veterinária, Universidade de Brasília, Brasília, 2018.

MONTENEGRO, C., ROA, F., SOARES-FILHO, W. BARROS-SILVA, A. E. Heterochromatic patterns of *Citrus* revisited: a new look at species origins and karyotype evolution. **Tree Genetics & Genomes**, v.19, n.36, 2023. DOI: 10.1007/s11295-023-01610-0.

MORRISON, O.; THAKUR, J. Molecular Complexes at Euchromatin, Heterochromatin and Centromeric Chromatin. **International journal of molecular sciences**, v.22, n.13, 2021. DOI: 10.3390/ijms22136922.

NEVES, N.; DELGADO, M.; SILVA, M.; CAPERTA, A.; MORAIS-CECÍLIO, L.; VIEGAS, W. Ribosomal DNA heterochromatin in plants. **Cytogenetic and genome research,** v.109, n.1-3, p.104-111, 2005. DOI: 10.1159/000082388.

NOBRE, L. L. M.; SANTOS, J. D. O. D.; LEITE, R.; ALMEIDA, C. Phylogenomic and single nucleotide polymorphism analyses revealed the hybrid origin of *Spondias bahiensis* (family Anacardiaceae): de novo genome sequencing and comparative genomics. **Genetics and Molecular Biology**, v.41, n.4, p.878-883, 2018. DOI: 10.1590/1678-4685-GMB-2017-0256.

OLIVEIRA, V. F. **Diversidade genética e avaliação do conteúdo de DNA nuclear entre matrizes de caju-de-árvore-do-cerrado (*Anacardium othonianum* Rizzini).** 2018. Dissertação (Mestrado e Agroquímica) Instituto Federal de Educação, Ciência e Tecnologia Goiano, Rio Verde, 2018.

ORTIZ, A. M.; ROBLEDO, G.; SEIJO, G.; VALLS, J. F. M.; LAVIA, G. I. Cytogenetic evidences on the evolutionary relationships between the tetraploids of the section *Rhizomatosae* and related diploid species (*Arachis*, Leguminosae). **Journal of Plant Research,** v.130, p.791-807, 2017. DOI: 10.1007/s10265-017-0949-x.

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

PAULSON, J. R.; HUDSON, D. F.; CISNEROS-SOBERANIS, F.; EARNSHAW, W. C. Mitotic chromosomes, **Seminars in Cell & Developmental Biology**, v.117, p.7-29, 2021. DOI: 10.1016/j.semcdb.2021.03.014.

PEDROSA, P. H.; GITAÍ, J.; SILVA, A. E. B.; FELIX, L. P.; GUERRA, M. Citogenética de angiospermas coletadas em Pernambuco. **Acta Botanica Brasilica,** v.13, n.1, p.49-60, 1999.

PENAGOS-PUIG, A.; FURLAN-MAGARIL, M. Heterochromatin as an Important Driver of Genome Organization**. Frontiers in Cell and Developmental Biology**, v.8, n. 579137, 2020. DOI: 10.3389/fcell.2020.579137.

PEREIRA, L. P; DA SILVA, D.F. P; DE SOUZA, L.K.F.; PEREIRA, E.T.L.; DA ASSUNÇÃO, H.F.; COSTA, M.M. Genetic diversity of bushy cashew (*Anacardium humile* A. St.-Hil.) based on characteristics of fruits. **Revista Brasileira Fruticultura**, v.41, n.5, 2019. DOI: 10.1590/0100-29452019065.

PESSOA, E. M.; CORDEIRO, J. M. P.; FELIX, L. P.; ALMEIDA, E. M.; COSTA, L.; NEPOMUCENO, A.; SOUZA, G.; CHASE, M. W.; ALVES, M.; VAN DEN BERG, C. Too many species: morphometrics, molecular phylogenetics and genome structure of a Brazilian species complex in *Epidendrum* (Laeliinae; Orchidaceae) reveal fewer species than previously thought. **Botanical Journal of the Linnean Society,** v.195, n.2, p. 161-188, 2021. DOI: 10.1093/botlinnean/boaa071.

POWO (2023). Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Disponível em http://www.plantsoftheworldonline.org. Acesso; 19 setembro de 2023.

QUERINO, B. C.; FERRAZ, M. E.; MATA-SUCRE, Y.; SOUZA. G.; FELIX, L.P. Cytomolecular diversity of the subtribe Laeliinae (Epidendroidae, Orchidaceae) suggests no relationship between genome size and heterochromatin abundance**. Plant Systematics and Evolution, v.306, n.19,2020. DOI:** 10.1007/s00606-020-01650-2.

RABAH, S.O.; LEE, C.; HAJRAH, N.H.; MAKKI, R.M.; ALHARBY, H.F.; ALHEBSHI, A.M.; SABIR, J.S.M.; JANSEN, R.K.; RUHLMAN, T.A.; Plastome Sequencing of Ten Nonmodel Crop Species Uncovers a Large Insertion of Mitochondrial DNA in Cashew. **Plant Genome**, v.10, n.3, 2017. DOI:10.3835/plantgenome2017.03.0020.

REIS, A. F.; SCHMIELE, M. Characteristics and potentialities of Savanna fruits in the food industry. **Brazilian Journal of Food Technology**, v. 22, 2019. DOI: 10.1590/1981-6723.15017.

ROA, F.; GUERRA, M. Distribution of 45S rDNA sites in chromosomes of plants: Structural and evolutionary implications. **BMC Evolutionary Biology (Online**), v.12, p. 225, 2012.

ROA, F.; GUERRA, M. Non-Random Distribution of 5S rDNA Sites and Its Association with 45S rDNA in Plant Chromosomes. **Cytogenetic and Genome Research (Printed ed.)**, v.146, p.243-249, 2015.

RUFINO, M. S. M.; CORRÊA, M. P. F.; ALVES, R. E.; LEITE, L. A. S. Utilização atual do cajuí nativo da vegetação litorânea do Piauí, Brasil. **Proccedings of the Tropical Region - American Society for Horticultural Science**, v.52, p.147-149, 2008.

RUFINO, M. S. M. **Qualidade e potencial de utilização de cajuís (*Anacardium* spp.) oriundos da vegetação litorânea do Piauí.** 2004. 92f. Dissertação (Mestrado em Desenvolvimento e meio ambiente) Universidade Federal do Piauí, Teresina, 2004.

RUFINO, M. S. M.; CORRÊA, M. P. F.; ALVES, R. E.; BARROS, L. M.; LEITE, L. A. S. **Suporte tecnológico para a exploração racional do cajuizeiro**. Fortaleza: Embrapa Agroindústria Tropical, (Embrapa Agroindústria Tropical, Comunicado Técnico, 107), p.30, 2007.

SALGOTRA, R. K.; CHAUHAN, B. S. Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. **Genes**, v.14, n.1, 2023. DOI: 10.3390/genes14010174.

SANTOS, R. DA C.; SANTOS-JÚNIOR, J. E. Divergência genética por análise multivariada de caracteres fenotípicos de *Anacardium humile* (St. Hilaire). **Revista Ceres**, v.62, n.6, p.553-560, 2015. DOI: 10.1590/0034-737X201562060007.

SANTOS, A.; NASCIMENTO, T.; FELIX, L. P.; GUERRA, M. Cytomolecular characterization of early and late condensing chromosome regions in some Annonaceae species. **Botanical Journal of the Linnean Society**, v.203, n.2, p.180-192, 2023. DOI: 10.1093/botlinnean/boad020.

SAVADI, S.; SOWMYA, K.; MEGHA, V.S.; MURALIDHARA, B. M.; MOHANA, G. S. Genetic diversity and identification of interspecific hybrids of *Anacardium* species using microsatellites. **Brazilian Journal of Botany**, v.44, p.139-148, 2020. DOI:10.1007/s40415-020-00678-5.

SILVA, D. B. DA; SILVA, J. A. DA; JUNQUEIRA, N. T. V.; ANDRADE, L. R. M. DE. **Frutas do cerrado**. Brasília: Embrapa Informações Tecnológica, 2001. p.179.

SILVA, M. R.; SILVA, M. S.; OLIVEIRA, J. S. Estabilidade de ácido ascórbico em pseudofrutos de caju-do-cerrado refrigerados e congelados. **Pesquisa Agropecuária Tropical**, v.34, p.9-14, 2004.

SILVA-JÚNIOR, J.; SOUZA, F. V. D.; PÁDUA, J. G. **A arca de Noé das frutas nativas brasileiras**. Embrapa, Brasília, 2021.

SILVA-LUZ, C. L. S. **Filogenia e sistemática de *Schinus L*. (Anacardiaceae), com revisão de um clado endêmico das matas nebulares andinas**. 2017. Tese (Doutorado em Botânica) - Instituto de Biociências, Universidade de São Paulo, São Paulo, 2017. DOI:10.11606/T.41.2017.

SCHWEIGGERT, R.M.; VARGAS, E.; CONRAD, J.; HEMPEL, J.; GRAS, C.C.; ZIEGLER, J.U.; MAYER, A.; JIMÉNEZ, V.; ESQUIVEL, P.; CARLE, R. Carotenoids, carotenoid esters, and anthocyanins of yellow-, orange-, and red-peeled cashew apples (*Anacardium occidentale* L.). **Food Chem**, v.200, n.1, p.274-82, 2016. Doi: 10.1016/j.foodchem.2016.01.038.

STEPANENKO, A.; CHEN, G.; HOANG, P. T. N.; FUCHS, J.; SCHUBERT, I.; BORISJUK, N. The Ribosomal DNA Loci of the Ancient Monocot *Pistia stratiotes* L. (Araceae) Contain Different Variants of the 35S and 5S Ribosomal RNA Gene Units. **Frontiers in plant science**, v.13, 2022. DOI: 10.3389/fpls.2022.819750.

STIMPSON, K. M.; SULLIVAN, B. A. Centromere: In: MALOY, S.; HUGHES, K. **Brenner's Encyclopedia of Genetics**: Academic Press, p.500-502, 2013. DOI:10.1016/B978-0-12-374984-0.00219-9.

TAKEHANA, C. L. I.; OHASHI, S.T.; JARDIM, M. A. G.; DOS SANTOS, J. U. M. Biologia floral e visitantes florais de *Anacardium giganteum* W. Hancock ex Engl. (Anacardiaceae) no município de Bragança, Pará. Rev. **Ciências Agrárias**, v.56, n.3, p. 202-211, 2013. DOI: 10.4322/rca.2013.030.

TURCHETTO-ZOLET, A. C.; TURCHETTO, C.; ZANELLA, C. M.; PASSAIA, G. **Marcadores Moleculares na Era genômica**: Metodologias e Aplicações. Ribeirão Preto: Sociedade Brasileira de Genética, 181 p. 2017.

VAN-LUME, B.; ESPOSITO, T.; DINIZ-FILHO, J.A.; GAGNON, E.; LEWIS, G.P.; SOUZA, G. Heterochromatic and cytomolecular diversification in the Caesalpinia group (Leguminosae):

relationships between phylogenetic and cytogeographical data**. Perspectives in Plant Ecology Evolution and Systematics**, v.29, p.51-63, 2017. DOI: 10.1016/J.PPEES.2017.11.004.

VESELINYOVÁ, D.; MASLANKOVÁ, J.; KALINOVA, K.; MIČKOVÁ, H.; MAREKOVÁ, M.; RABAJDOVÁ, M. Selected in situ hybridization methods: principles and application. **Molécules**, v.26*,* 2021. DOI: 10.3390/molecules26133874.

VIEIRA, R. F. V.; COSTA, T. S. A.; DA SILVA, D. B.; FERREIRA, F. R.; SANO, S. M. Embrapa Recursos Genéticos e Biotecnologia. **Frutas Nativas da Região Centro – Oeste do Brasil.** Brasília, 2006.p.320.

VIEIRA, M.; MAYO, S.J.; DE ANDRADE, I.M. Geometric morphometrics of leaves of *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the coastal region of Piauí, Brazil. **Brazilian Journal Botany**, v.37, p.315-327, 2014.DOI: https://doi.org/10.1007/s40415-014-0072-3

WU, L.; NIE, L.; XU, Z.; LI, P.; WANG, Y.; HE, C.; SONG, J.; YAO, H. Comparative and Phylogenetic Analysis of the Complete Chloroplast Genomes of Three *Paeonia* Section *Moutan* Species (Paeoniaceae). **Frontiers in Genetics**, v.11, n.980, 2020. DOI:10.3389/fgene.2020.00980.

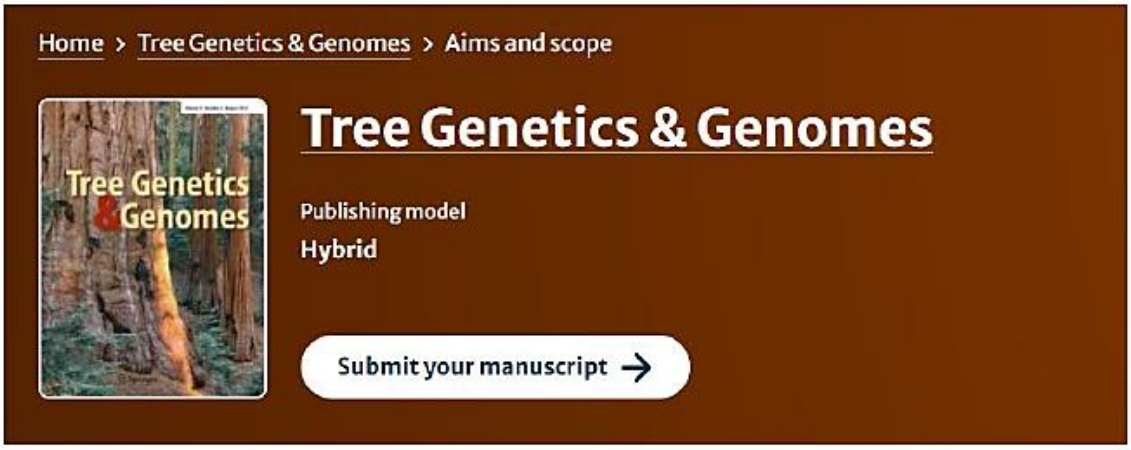
XIE, L.; YANG, Z.Y.; WEN, J.; LI, D.Z.; YI, T.S. Biogeographic history of *Pistacia* (Anacardiaceae), emphasizing the evolution of the Madrean-Tethyan and the eastern Asian-Tethyan disjunctions. **Molecular Phylogenetics and Evolution,** v.77, p.136-46, 2014. DOI: 10.1016/j.ympev.2014.04.006.

YI, T.; MILLER, A. J.; WEN, J. Phylogenetic and biogeographic diversification of *Rhus* (Anacardiaceae) in the Northern Hemisphere. **Molecular Phylogenetics and Evolution**, v.33, n.3, p.861-79, 2004. DOI: 10.1016/j.ympev.2004.07.0

YOUNG, A. D.; GILLUNG, J. P. Phylogenomics - principles, opportunities and pitfalls of big-data phylogenetics. **Systematic Entomology**, 2019. DOI: 10.1111/syen.12406.

YU, X.; YANG, D.; GUO, C.; GAO, L. Plant phylogenomics based on genome-partitioning strategies: Progress and prospects. **Plant Diversity**, v. 40, p.158-164, 2018. DOI: https://doi.org/10.1016/j.pld.2018.0 121

ANEXO - Normas de submissão da revista Tree Genetics & Genomes



Aims and scope

Tree Genetics and Genomes is an international, peer-reviewed journal, which provides for the rapid publication of high quality papers covering the areas of forest and horticultural tree genetics and genomics.

Manuscript Submission

Manuscript Submission

Submission of a manuscript implies: that the work described has not been published before; that it is not under consideration for publication anywhere else; that its publication has been approved by all co-authors, if any, as well as by the responsible authorities – tacitly or explicitly – at the institute where the work has been carried out. The publisher will not be held legally responsible should there be any claims for compensation.

Permissions

Authors wishing to include figures, tables, or text passages that have already been published elsewhere are required to obtain permission from the copyright owner(s) for both the print and online format and to include evidence that such permission has been granted when submitting their papers. Any material received without such evidence will be assumed to originate from the authors.