Genet Resour Crop Evol
https://doi.org/10.1007/s10722-022-01501-8

RESEARCH ARTICLE

Genetic diversity and structure of *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. (Arecaceae) using microsatellite DNA markers in Costa Rica

Valeria Navarro-Cascante · Elizabeth Arnáez-Serrano · Mónica Rojas-Gómez · Ileana Moreira González · Guillermo Vargas-Hernández · Nelson A. Zamora · Elemer Briceño-Elizondo · Jonathan Morales-Marroquín · Pamela Sevilla-Cortés · Jesús Antonio Oviedo-Ulate · Emanuel Araya-Valverde

Received: 14 March 2022 / Accepted: 20 November 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract *Acrocomia aculeata* is a tree palm species widely distributed throughout the Neotropics, from Mexico to Argentina including some Caribbean islands. The oil from the fruit is used for biodiesel production, human consumption, and biomass fuel. Despite the need to better understand its genetic diversity for commercial use and conservation, genetic diversity and structure knowledge of this species is scarce in Mesoamerica. We used microsatellite markers to analyze 185 samples of *A. aculeata*, from 18 sampling sites, that essentially covered geographic distribution of this species in Costa Rica. We found low genetic diversity across sampling sites, with He values below 0.50 which, is lower than diversity levels found in South America. Interestingly, samples collected in the national parks Santa Rosa and Rincon de la Vieja National exhibited the highest genetic diversity (He = 0.42 and He = 0.50, respectively). The 18 sampling sites were structured

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10722-022-01501-8.

V. Navarro-Cascante · M. Rojas-Gómez · P. Sevilla-Cortés · J. A. Oviedo-Ulate · E. Araya-Valverde (✉)
Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, Pavas, San José 1174-1200, Costa Rica
e-mail: earaya@cenat.ac.cr

V. Navarro-Cascante
e-mail: vale_nc96@outlook.com

M. Rojas-Gómez
e-mail: mrojas@cenat.ac.cr

P. Sevilla-Cortés
e-mail: sevillacortesp@gmail.com

J. A. Oviedo-Ulate
e-mail: joviedou@earth.ac.cr

E. Arnáez-Serrano · I. M. González · E. Briceño-Elizondo · P. Sevilla-Cortés · E. Araya-Valverde
Escuela de Biología, Instituto Tecnológico de Costa Rica, 159-7050 Cartago, Costa Rica
e-mail: earnaez94@gmail.com

I. M. González
e-mail: ilea2757@gmail.com

E. Briceño-Elizondo
e-mail: ebriceno@uned.ac.cr

G. Vargas-Hernández
Estación Experimental Fabio Baudrit Moreno, Universidad de Costa Rica, Alajuela 11501-2060, Costa Rica
e-mail: guillermo.vargas@ucr.ac.cr

N. A. Zamora
Escuela de Ingeniería Forestal, Instituto Tecnológico de Costa Rica, 159-7050 Cartago, Costa Rica
e-mail: nzamora@itcr.ac.cr

J. Morales-Marroquín
Institute of Biology, Department of Genetics, University of Campinas UNICAMP, P.O. Box: 6109, Campinas 13083-970, Brazil
e-mail: jmorales1089@gmail.com

Published online: 02 December 2022

in three clusters. Cluster A contains samples collected in the lower northwestern Pacific, Central Pacific, and Central Region of Costa Rica. Cluster B consisted of samples from sites in the upper northwestern Pacific. Cluster C contains samples from the southeastern Pacific region, being largely separated from Clusters A and B which was supported by the highest differentiation coefficients. This is the first large genetic diversity study of *A. aculeata* in Mesoamerica; therefore, our results serve as reference for future studies on germplasm diversity studies in this region and a baseline for future conservation and management efforts of *A. aculeata*.

Keywords Coyol · Multiplex PCR · Mesoamerica · Conservation · Differentiation genetic resources · Discontinuous populations

Introduction

Acrocomia aculeata is widely distributed in the Neotropics from Mexico to Argentina, Bolivia, and Paraguay and also in the Caribbean (Henderson et al. 1995; Grayum 2003; Dransfield et al. 2008). Its center of origin is presumed to be in South America and the highest diversity appears to be in southeastern Brazil (Scariot et al. 1995; Lanes et al. 2015; de Lima et al. 2018). This conclusion is also supported by Diaz et al. (2021) that found the highest genetic diversity of *A. aculeata* in Brazil and the lowest diversity in samples from Central America. Moreover, fossil records demonstrated that human introduction of this palm tree to Central America came from South America (Morcote-Ríos and Bernal 2001).

In Costa Rica, *A. aculeata* frequently inhabits the Pacific lowlands, mostly from northwestern Guanacaste to the central Pacific slope and occasionally to the Osa Peninsula and Punta Burica. It also occurs in the Guatuso plains in the northern Caribbean slope (Janzen 1983; Grayum 2003; Cornejo et al. 2012). Particularly, the occurrence of this species was studied in the Guanacaste where individuals are dispersed in the dry forest and pasturelands (Henderson et al. 1995; Esquivel-Mimenza et al. 2011). Cattle movement through pasturelands may facilitated dispersion in Costa Rica (Scariot 1998; Harvey et al. 2011).

The province of Guanacaste suffered accelerated deforestation during some 170 years, converting forest into livestock pastures. This land usage, in turn, promoted pastureland management that inhibited *A. aculeata* regeneration (Harvey et al. 2011; Calvo-Alvarado et al. 2009). In addition, the long-time practice of cutting down adult palm trees to obtain the sap, and fermenting into a beverage, also threatens the species. This practice primarily occurs throughout Central American, including Costa Rica with prevalence in Guanacaste (Janzen 1983; Balick 1990; Chizmar-Fernandez 2009; Díaz et al. 2021).

The *A. aculeata* fruit contains high oil content and good quality vegetable oil (Navarro-Díaz et al. 2014; Da Conceição et al. 2015; Lescano et al. 2015; Lieb et al. 2019) used to produce biodiesel, edible vegetable oil, and pharmaceutical products for the cosmetics industry. It also serves as a novel sustainable biomass feedstock and the post-fruit processing residues can be used as biomass fuel (Evaristo et al. 2016; Plath et al. 2016; Falasca et al. 2017). It has the potential to become a new commercial oilseed crop similar to that of oil palm if managed and bred properly (Colombo et al. 2018).

Studies in Costa Rica have quantified carotenoid and tocopherol content in fruits (Schex et al. 2018), oil content and fatty acid profile (Lieb et al. 2019), and variability of fruit morphology and oil composition in three locations (Alfaro-Solís et al. 2020). Despite these efforts, there are no studies yet that describe the genetic diversity of *A. aculeata* at large scale in Costa Rica. Molecular markers such as microsatellites and SNPs have played an important role in describing genetic diversity, structure and mating system of natural populations and germplasm collections of *A. aculeata* (Abreu et al. 2012; Lanes et al. 2015, 2016; Mengistu et al. 2016b; Silva et al. 2017; Coelho et al. 2018; Díaz et al. 2021; Laviola et al. 2022). Therefore, the aim of this study is to estimate the genetic diversity and to determine the structure of *A. aculeata* from samples collected in the range of geographic occurrence in Costa Rica,

J. Morales-Marroquín
Centro de Pesquisa de Recursos Genéticos Vegetais, Instituto Agronômico-IAC, Campinas, SP, Brazil

J. A. Oviedo-Ulate
Universidad EARTH, P.O. Box: 4442-1000, Limón, Costa Rica

using microsatellite markers to generate a baseline for future use and conservation of this promising species in Costa Rica.

Materials and methods

Biological material and study area

The study area comprised four regions in Costa Rica i.e. Brunca, Central, Chorotega, and Huetar Norte with a total of 18 sampling sites (Supplementary Table 1). We collected 185 samples in the following 18 sites: (1) La Cruz, (2) Guanacaste National Park (hereinafter referred as Guanacaste), (3) Santa Rosa National Park (hereinafter referred as Santa Rosa), (4) Liberia, (5) Rincon de la Vieja National Park (hereinafter referred as Rincon de la Vieja), (6) Caño Negro National Wildlife Refuge (hereinafter referred as Caño Negro), (7) Santa Cruz, (8) Nicoya, (9) Barra Honda, (10) Cañas, (11) Tilaran, (12) Abangares, (13) Chomes, (14) Orotina, (15) Atenas, (16) Turrubares, (17) Paso Real, and (18) Peninsula de Osa (Fig. 1). This sampling essentially covered almost all of Costa Rica's current geographic occurrence of *A. aculeata*. The number of collected individuals per site is displayed in the Supplementary Table 1.

Each sampling site consisted of areas where mature reproductive trees (either flowering or fruiting) were growing relatively dispersed. The minimum distance between trees in each site was ca. 40–100 m, depending on the total number of trees. A leaflet lamina fragment of about 20 cm in length from each palm tree sample was collected in conical plastic tubes of 50 mL containing table salt to avoid sample degradation. Upon arrival at the laboratory, the table salt was carefully removed, and samples were lyophilized (CHRIST BETA 1-8LD plus, Germany) prior to DNA extraction.

DNA isolation and microsatellite genotyping using multiplex PCR

For DNA isolation, 50 mg of dried leaflet lamina were ground in a homogenizer Fastprep-24 (MP Biomedicals, USA) using lysing matrix A (MP Biomedical,

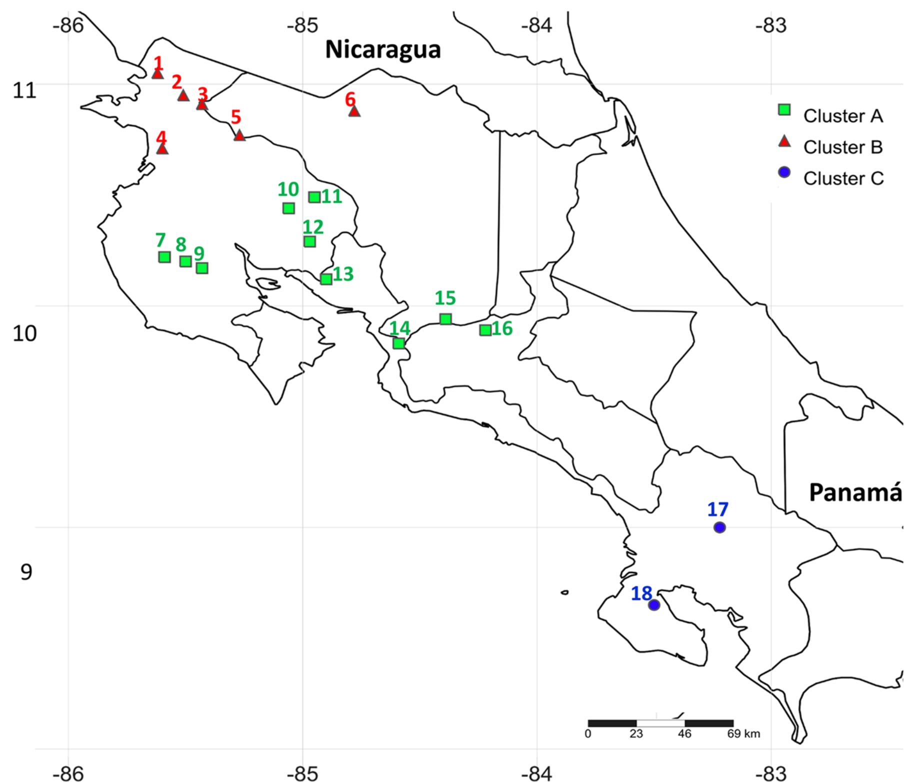

Fig. 1 Distribution of 18 sampled sites of *Acrocomia aculeata,* representing three clusters obtained thorough Structure. Sampled sites 1: La Cruz, 2: Guanacaste, 3: Santa Rosa, 4: Liberia, 5: Rincon de la Vieja, 6: Caño Negro, 7: Santa Cruz, 8: Nicoya, 9: Barra Honda, 10: Cañas, 11: Tilaran, 12: Abangares, 13: Chomes, 14: Orotina, 15: Atenas, 16: Turrubares, 17: Paso Real, 18: Peninsula de Osa. Green squares, red triangles and blue circles represent Cluster A, Cluster B and Cluster C, respectively

USA). A CTAB-based method was employed for DNA isolation (Quirós-Guerrero et al. 2019). The DNA concentration was measured with a spectrophotometer (NanoDrop, Thermo Scientific, USA) and diluted (~30 ng/µL) for PCR amplification.

Twelve microsatellite markers previously developed for *A. aculeata* (Nucci 2007; Nucci et al. 2008) and *Astrocaryum aculeatum* G.Mey. (Ramos et al. 2012) were used. The selection criteria of microsatellites included the level of polymorphism, informativeness previously reported (Lanes et al. 2015, 2016; Mengistu et al. 2016b; Coelho et al. 2018) and expected allele size. The forward primer of each microsatellite was labeled with any of four fluorescent dye labels, i.e., 6-FAM, VIC, NED, and PET for multiplex PCR assays. Several primer combinations and concentrations were iteratively performed to obtain three PCR multiplexes (Table 1). Each multiplex PCR reaction (final volume 25 µL) contained PCR Master Mix (1X DreamTaq Master Mix, ThermoFisher Scientific, USA), bovine serum albumin (0.01 mg, Sigma, USA), DMSO (2%, Sigma, USA), and DNA (~60 ng). Additionally, $MgCl_2$ (0.5 mM, Thermo Scientific, USA) and DNA Taq polimerase (1 Unit, ThermoFisher Scientific, USA) were added. The primer concentrations varied according to the optimized multiplexes (Table 1).

The amplification was performed in a thermal cycler (Veriti™, Applied Biosystems, USA) with the following profile: one cycle at 95 °C for 5 min followed by 35 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, with a final extension at 72 °C for 8 min. Multiplex PCR products were visualized by capillary electrophoresis in the ABI 3130xl Genetic Analyzer (Applied Biosystems, USA) with the combination of 1.5 µL of the multiplex PCR product, 0.4 µL of GeneScan 600LIZ™ (Applied Biosystems, USA), and 8.5 µL of Hi-Di Formamide (Applied Biosystems, USA). After capillary electrophoresis, allele binning and genotyping were manually inspected

Table 1 Primer sequences with its respective fluorescent dye label of 12 microsatellites and the final concentration used for the multiplex PCR amplification in DNA of *A. aculeata*

Microsatellite name (SSR)	Fluorescent dye-Sequence of forward primer (5′–3′)	Sequence of reverse primer (5′–3′)	Final concentration (µM)	Multiplex PCR	References
Aacu10	NED-TGCCACATAGAGTGC TTGCT	CTACCACATCCCCGTGAGTT	0.15	1	Nucci et al. (2008)
Aacu45	6FAM-CAGACTACCAGGCTT CCAGC	TCATCATCGCAGCTTGACTC	0.20	1	Nucci (2007)
Aacu74	VIC-TACTGTTGTGCCAAGTCC CA	GAGCACAAGGGGGATATCAA	0.20	1	Nucci (2007)
Aacu07	VIC- ATCGAAGGCCCTCCA ATACT	AAATAAGGGGACCCTCCAA	0.19	2	Nucci et al. (2008)
Aacu12	NED- GAATGTGCGTGCTCA AAATG	AATGCCAAGTGACCAAGTCC	0.13	2	Nucci et al. (2008)
Aacu26	6FAM- ACTTGCAGCCCCATA TTCAG	CAGGAACAGAGGCAAGTTC	0.19	2	Nucci et al. (2008)
Aacu30	6FAM- TGTGGAAGAAACAGG TCCC	TCGCCTTGAGAAATTATGGC	0.19	2	Nucci et al. (2008)
Aac09	VIC- AGCGTAATAGCATCCCTT GC	AGTCCAGTGTTCTTCCCCTG	0.13	2	Ramos et al. (2012)
Aac06	6FAM-TCTGATCCATCTGGT TGTCTAA	TGCATGGTGCTAGAGTAATCC	0.20	3	Ramos et al. (2012)
Aac14	6FAM-GGCCAGTTGTGTTGA TGAAA	TTAAGCCCCTGGTGAAAACA	0.20	3	Ramos et al. (2012)
Aac04	VIC-GCATTGTCATCTGCAACC AC	GCAGGGGCCATAAGTCATAA	0.20	3	Ramos et al. (2012)
Aac12	NED-GCTCTGTAATCTCGGCTT CCT	TCCAGTTCAAGCTCTCTCAGC	0.20	3	Ramos et al. (2012)

with GeneMapper™ v4.0 software (Applied Biosystems, USA) for further data analysis.

Microsatellite informativeness, genetic diversity and structure of *A. aculeata*

Genotyping data matrix was run in the R package *adegenet* (Jombart 2008) version 2.1.2 to create a genind class object with the data organized according to sampling sites. Each sampling site was considered a putative population. Further, the R package *hierfstat* (Goudet 2005) version 0.04–22 was used to analyze the genetic diversity through the estimation of number of alleles per population, He, Ho, number of effective alleles (Ne), and private alleles (Np). With the aid of the R package *pegas* (Paradis 2010) version 0.12, the inbreeding coefficient through F Statistic (Fst) was estimated (Wright 1965). Finally, the Polymorphic Information Content (PIC) was estimated using the R package *polysat* (Clark and Jasieniuk 2011) version 1.7. The intra- and interpopulation differentiation were also examined with an AMOVA using *poppr* (Kamvar et al. 2014) version 2.8.3 and *ade4* (Dray and Dufour 2007) version 1.7–13.

In order to study the genetic structure, the genotyping table was used in the software Structure version 2.3.4 (Pritchard and Matthew 2000) configured with eight repetitions for each possible number of groups (K from 1 to 8), burn-in of 50,000 and 100,000 repetitions of the Markov Chain Monte Carlo. The Structure results were used as input to estimate the best K number with Structure Harvester (Earl and von Holdt 2012) according to Evanno et al. (2005). This represents the most probable number of populations based on the inferred ancestry of each individual.

The genetic distance matrix of Nei was obtained with the R package *mmod* (Winter 2012) version 1.3.3. Besides, the genetic distance matrix was used in a UPGMA analysis to construct dendrograms in *MEGA6* (Tamura et al. 2013) and *DAPC* using *adegenet* (Jombart 2008) to identify the separation between clusters. Finally, the Mantel test was performed using the package *ade4* (Dray and Dufour 2007) to study the correlation between the matrix of genetic distances and geographic distances which results were visualized with an isolation by distance figure made with *MASS* (Venables and Ripley 2002) version 7.3–51.5.

Results

Microsatellite informativeness

Three of 12 SSRs were monomorphic (Aac12, Aacu45, and Aacu74). Table 2 displays the parameter of informativeness for nine polymorphic SSRs that accounted for a total of 39 different alleles. The number of alleles at each locus ranged from 3 to 7 with an average of 4.33 alleles per locus. Three SSRs (Aac04, Aacu10, and Aac14) exhibited the highest number of alleles (Na) and expected heterozygocity (He). The least number of different alleles was observed in Aacu07, Aacu12, and Aacu26. Five SSRs (Aac04, Aac14, Aac09, Aac06, and Aacu26) were highly informative (PIC > 0.5), while the remaining four (Aacu10, Aacu12, Aacu07, and Aacu30) were reasonably informative (0.5 > PIC > 0.25) (Botstein et al. 1980) (Table 2). On the other hand, Fst values (Table 2) indicate that there is moderate to high levels

Table 2 Genetic diversity parameters for nine polymorphic microsatellites amplified 185 DNA samples of *A. aculeata* samples. NA: number of alleles, He: expected heterozygosity, Ho: observed heterozygosity, PIC: Polymorphic Information Content and Fst as genetic differentiation coefficients

Microsatellite	Allele range (bp)	Na	He	Ho	PIC	Fst
Aac04	199–217	7	0.47	0.38	0.71	0.33
Aac06	130–136	4	0.29	0.29	0.53	0.47
Aac09	329–335	4	0.37	0.29	0.53	0.34
Aac14	336–353	5	0.33	0.33	0.54	0.41
Aacu07	132–138	3	0.25	0.35	0.36	0.18
Aacu10	147–166	6	0.33	0.37	0.43	0.10
Aacu12	168–172	3	0.19	0.26	0.35	0.26
Aacu26	264–268	3	0.27	0.23	0.50	0.41
Aacu30	247–252	4	0.15	0.15	0.40	0.50
Average		4.33	0.29	0.29	0.48	0.33

of differentiation among populations. However, typically, a Fst = 0.334, is considered high for most tropical plants.

Genetic diversity and structure of *A. aculeata*

The average number of alleles per population was 18.89. The lowest genetic diversity (He < 0.19) was found in samples collected in Peninsula de Osa, Orotina and Abangares, which also had a low number of effective alleles (Table 3). La Cruz, Atenas, Tilaran, Barra Honda, Chomes, Santa Cruz, and Nicoya registered an intermediate diversity (He ranging 0.23-0.28). The sites with the highest diversity (He > 0.30) were observed in samples from Guanacaste, Liberia, Cañas, Caño Negro, Turrubares, Paso Real, Santa Rosa and Rincon de la Vieja. Santa Rosa and Rincon de la Vieja also displayed a high number of alleles when compared with other sites (Table 3). We only found five private alleles of 147 bp (Aacu10), 136 bp (Aac06), 250 bp (Aacu30), 331 bp (Aac09), and 264 bp (Aacu26), which amplified in samples of Caño Negro, Chomes, Liberia, Santa Cruz, and Tilaran, respectively (Table 3).

Table 3 Genetic diversity parameters estimated in the 18 sampling sites of *A. aculeata* in Costa Rica. He: expected heterozygosity, Ho: observed heterozygosity, Na: number of alleles, Ne: number of effective alleles, Np: number of private alleles

Site of collection	N_a	N_e	H_e	H_o	N_p
Abangares	16	1.30	0.19	0.17	0
Atenas	20	1.35	0.24	0.22	0
Barra Honda	18	1.42	0.26	0.30	0
Cañas	19	1.55	0.32	0.27	0
Caño Negro	17	1.59	0.33	0.36	1
Chomes	19	1.42	0.26	0.22	1
Guanacaste	18	1.59	0.31	0.40	0
La Cruz	15	1.34	0.23	0.31	0
Liberia	25	1.62	0.32	0.28	1
Nicoya	24	1.52	0.28	0.24	0
Orotina	14	1.26	0.18	0.20	0
Paso Real	17	1.75	0.35	0.41	0
Peninsula de Osa	19	1.23	0.17	0.09	0
Rincon de la Vieja	24	1.96	0.50	0.48	0
Santa Cruz	20	1.50	0.28	0.24	1
Santa Rosa	20	1.84	0.42	0.43	0
Tilaran	16	1.50	0.26	0.28	1
Turrubares	19	1.65	0.33	0.30	0
Average	**18.89**	**1.52**	**0.29**	**0.29**	

The AMOVA indicated that 20.82% of the total genetic variation was explained between regions (Brunca, Central, Huetar Norte, and Chorotega) and 20.39% was explained between sampling sites within regions. The least variation (3.13%) was observed between samples within each site and the greatest differentiation (55.65%) occurred within samples (Supplementary Table 2). Bayesian analysis revealed that the genetic structure among *A. aculeata* populations is best described by three distinct clusters according to the highest ΔK value (182.86) (Supplementary Fig. 1). Cluster A is comprised of samples collected at Barra Honda, Nicoya, Santa Cruz, Cañas, Tilaran, Bagaces, Chomes, Orotina, Atenas, and Turrubares. Cluster B comprised samples from La Cruz, Guanacaste, Santa Rosa, Liberia, Rincon de la Vieja, and Caño Negro. Cluster C grouped samples of Paso Real and Peninsula de Osa (Fig. 2). The assignment of each individual within each of the clusters in Fig. 2 is supported by the percentage assignment given by the *Structure* software (Supplementary Table 3). Figure 1 also shows the clustering observed in the structure analysis.

The clustering obtained with *Structure* was the same with that found with DPCA analysis (Supplementary Fig. 2). Cluster C largely separated into Peninsula de Osa and Paso Real from the remaining sites (Supplementary Fig. 2A). This result is also consistent with the UPGMA that showed a separated branch for Peninsula de Osa and Paso Real with a high bootstrap value (Fig. 3) and observed Fst values. Cluster C registered an Fst value of 0.128 compared to Clusters A and B. The differentiation of Peninsula de Osa was moderate to high with Fst values between 0.104 and 0.198, with an average of 0.145. The differentiation of Paso Real was low to moderate (mean Fst = 0.111) (Supplementary Tables 4 and 5).

When Cluster C was not included in the DPCA analysis, a clearer separation of Clusters A and B was observed (Supplementary Fig. 2B). The UPGMA (Fig. 3) also shows a clear separation of Clusters A and B (bootstrap = 98) consistent with the DPCA and structure analyses. This clustering pattern was also supported by a moderate differentiation in both clusters with average Fst values of 0.065 and 0.076, respectively (Supplementary Tables 4 and 5). In Cluster A, a subcluster with samples from Barra Honda,

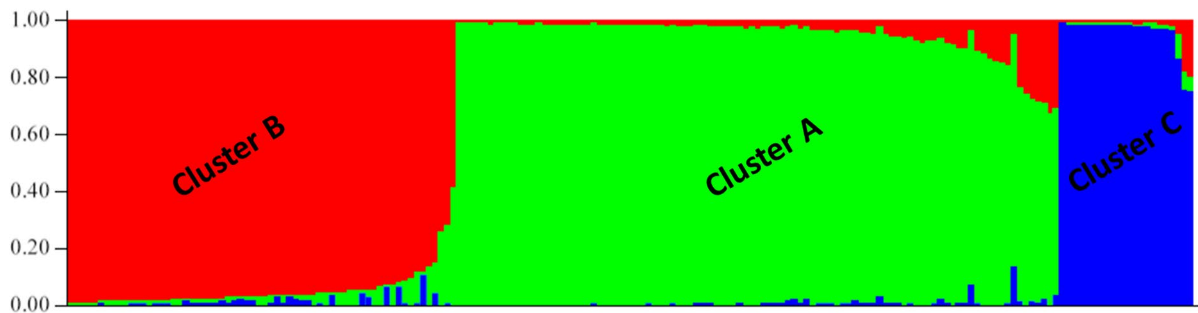

Fig. 2 Population structure of the 185 samples of *A. aculeata* collected in Costa Rica. Clusters assignment of samples (k = 3) is based on the Bayesian analysis on Structure Software using microsatellite markers genotyping

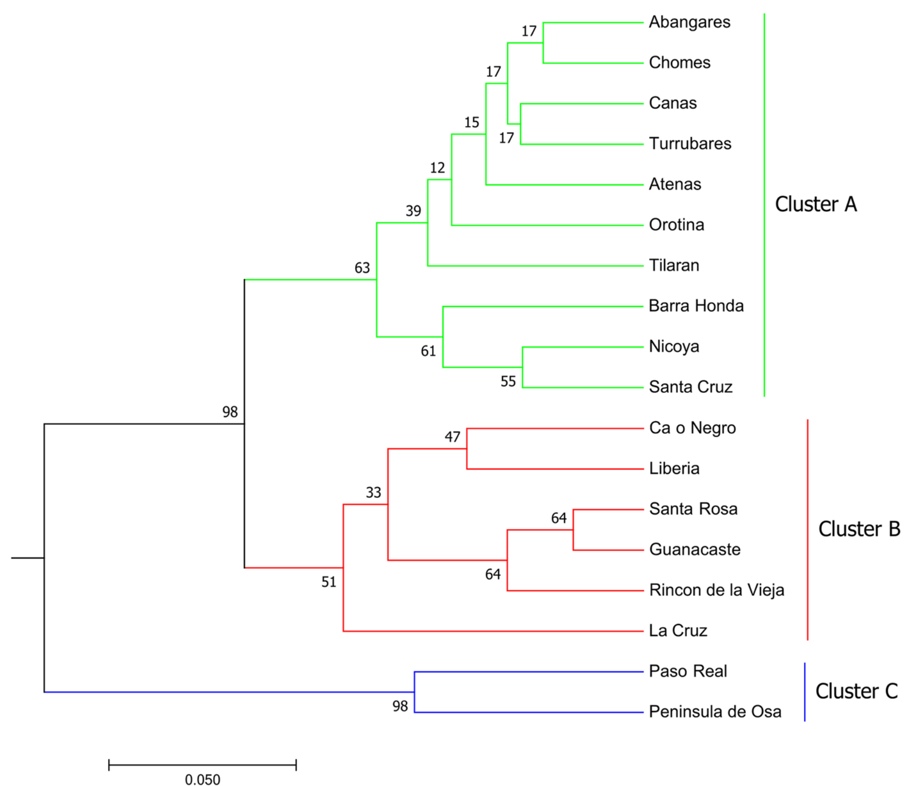

Fig. 3 UPGMA dendrogram of *Acrocomia aculeata*, estimated according to Nei's genetic distance. The dendrogram reflects the same distribution of 18 sampled sites represented in Fig. 1. Sampled sites 1: La Cruz, 2: Guanacaste, 3: Santa Rosa, 4: Liberia, 5: Rincon de la Vieja, 6: Caño Negro, 7: Santa Cruz, 8: Nicoya, 9: Barra Honda, 10: Cañas, 11: Tilaran, 12: Abangares, 13: Chomes, 14: Orotina, 15: Atenas, 16: Turrubares, 17: Paso Real, 18: Peninsula de Osa

Nicoya, and Santa Cruz was observed (Fig. 3). These sampling sites are geographically close together (Fig. 1). Inside Cluster B a separation of samples from Caño Negro was resolved with the DPCA and the UPGMA; this site is relatively closer to Liberia (Fig. 3). This proximity was supported with the lowest Fst value (0.035) between Caño Negro and Liberia (Supplementary Table 4). Also, inside Cluster B, samples collected in three national parks (Santa Rosa, Guanacaste, and Rincon de la Vieja) seem to group into a subcluster. Interestingly, in the UPGMA this subcluster also yields a bootstrap value of 64.

The correlation of genetic distances and geographical locations was explored with an isolation by distance (IBD) analysis (Supplementary Fig. 3). A positive and significant correlation ($r = 0.398$; $p < 0.05$) was obtained when all 185 samples were included in the analysis. Similarly, we found a significant but weak positive correlation ($r = 0.200$; $p < 0.05$) when samples from Peninsula de Osa and Paso Real were

excluded. In both cases, genetic and geographical distances displayed dependency on each other suggesting IBD of samples collected in this study (Supplementary Fig. 3B).

Discussion

Genetic diversity of *A. aculeata*

The observed values of parameters such as He, Ne, and Ho in our study suggest that the genetic diversity of *A. aculeata* in Costa Rica is relatively low (Table 3). Studies of samples from an ex situ germplasm bank have shown higher genetic diversity, with almost the same set of SSRs as our work did. Lanes et al. (2015) found that with six SSRs the genetic diversity was between 0.55 and 0.74, while Mengistu et al. (2016a) reported a diversity range of 0.49–0.63 with ten SSRs. Comparison of our study results with those of *A. aculeata* studied in natural Brazilian populations support the conclusion that Costa Rica has low genetic diversity of *A. aculeata*. Higher levels of genetic diversity (He > 0.50) have been reported in natural populations in Brazil (Coelho et al. 2018; de Lima et al. 2020).

Recently, Díaz et al. (2021) observed that with SNPs markers, samples from Brazil had the highest genetic diversity when compared with samples from Central America. They also found two groups of *A. aculeata*. The northern group included samples from Costa Rica, Trinidad and Tobago, Puerto Rico, Mexico, and Colombia which had the lowest genetic diversity. Despite using microsatellite markers, our results agree with those of Díaz et al. (2021).

Some researchers have proposed that the center of origin and diversification of *A. aculeata* is in South America (Lanes et al. 2015). Based on *A. aculeata* remains found from Brazil to Mexico (Morcote-Ríos and Bernal 2001), and consistent with Janzen's (1983) hypothesis that pre-Columbian Indians introduced *A. aculeata* into Costa Rica, humans may indeed have facilitated dispersion of this species from South to Central America. *A. aculeata* may also have been dispersed from the Caribbean islands, where some samples have been collected to estimate genetic diversity (Díaz et al. 2021). The low genetic diversity found in our study supports this south to northward dispersion hypothesis.

A factor that may further explain the low genetic diversity is that this species has not been found to grow naturally in any other specific habitats other than human-disturbed areas such as paddocks, abandoned land, along roadsides, and pasturelands (Janzen 1983; Harvey et al. 2011). Despite that, *A. aculeata* is essentially the most abundant tree palm species in the Guanacaste area, where several samples were collected, its density remain very low (0.75 individuals ha^{-1}) (Harvey et al. 2011), This contrasts strongly with the 300 individuals ha^{-1} found naturally in Brazil, where *A. aculeata* grows as aggregated clusters forming continuous forests stands (Coelho et al. 2018). In fact during the field trips, it proved difficult to locate sampling sites with groups of palm trees that met the selection criteria. This low density and fragmented distribution of *A. aculeata*, may restrict gene flow among dispersed individuals which may then explain the decreasing genetic diversity and differentiation among studied sites.

Since intense cattleranching and timber extraction in Guanacaste were princpal drivers of deforestation in this region (Calvo-Alvarado et al. 2009), it is interesting to note that both Santa Rosa and Rincon de la Vieja recorded the highest genetic diversity. Both these national parks were established in Guanacaste since 1970 when cattleraising and forest clearing were prohibited. More sampling would be required to corroborate this interesting hypothesis.

Genetic structure of *A. aculeata*

The structure analysis revealed three clusters (Fig. 2) supported by the observed groupings in the UPGMA (Fig. 3) and DPCA (Supplementary Fig. 2). This structure had a moderate differentiation for Cluster A and Cluster B while for Cluster C the differentiation was higher. Furthermore, a significant IBD was observed between the three clusters (Supplementary Fig. 3A). The observed structure in this study may be explained by the dispersion pattern of *A. aculeata* associated with cattle ranching in Costa Rica (Harvey et al. 2011). This anthropogenic activity does not favor a continuous distribution of *A. aculeata* which probably limits gene flow.

Cluster A includes sampling sites in the lower northwestern Pacific in the Nicoya Peninsula area, Central Pacific and, Central Region of Costa Rica. Although there are mountainous areas and river

basins that delimit these regions (Gabb 2007), no differentiation was observed within sampling sites of Cluster A. Despite this, in the UPGMA (Fig. 3) a subcluster containing Barra Honda, Nicoya, and Santa Cruz was identified. These three sites are located in the lower northwestern Pacific region of Nicoya Peninsula, while sites of Central Pacific are separated by the Guanacaste Mountain Range foothills and the Tilaran Mountain Range (Solano-Quintero 2001).

As shown in Figs. 2 and 3, Cluster A and Cluster B are clearly separated. Sampling sites of Cluster B such as Liberia, Santa Rosa, Guanacaste, Rincon de la Vieja, and La Cruz are located in upper northwestern Pacific. This region is geographically located in the subregion of the Tempisque River Northeast Basin (Solano-Quintero 2001; Arroyo-Mora et al. 2005; Calvo-Alvarado et al. 2009). The first Spanish colonizers discovered that soil fertility was favorable for livestock raising in this basin. On the contrary, they found that soil characteristics and topology in the Nicoya Peninsula were not suitable for this activity (Calvo-Alvarado et al. 2009).

During the 1800s, in fact, landowners established large farms (>5000 ha) used for livestock intensification until 1972 (Hall 1984; Calvo-Alvarado et al. 2009). In the 1970s, most forests in Guanacaste were cleared due to the accelerated deforestation and converted into pastures for livestock, rice and sugar cane crops and some timberwood forestry. Therefore, the land use of these activities favored a discontinuos occurrence of *A. aculeata* which may explain the separation of both Cluster A and B.

It is interesting to note, moreover, that the DPCA (Supplementary Fig. 2) displayed a separation of Caño Negro from other sites inside this cluster. Caño Negro is located on the Caribbean slope in the lowland northern zone of Guatusos plains far enough from the rest of the sites which possibly explain genetic separation or differences between them. Finally, Cluster C (Peninsula de Osa and Paso Real) was largely separated from the remaining sites, an observation supported by the highest differention coefficients (Supplementary Table 4). Even when more sampling is needed to clearly elucidate Cluster C's genetic structure, it is important to mention that the cultural use of *A. aculeata* in this region is traditionally much less than that in the northwestern Pacific region; this results in the species becoming less common and individuals more restricted or scattered across the region.

Need for *A. aculeata* genetic characterization for uses and conservation uses

Our results are consistent with those of Diaz et al. (2021) that demonstrate low genetic diversity of *A. aculeata* in Costa Rica and other countries in Central America relative to genetic diversity in Brazil. Our study, therefore, serves as a useful input to conservation efforts in Costa Rica, as well as in other Mesoamerican countries. For example, in Costa Rica a germplasm bank is being established by our research group with seedlings found close to mature trees (when available) in our sampling sites. New seed and seedlings collections should consider our results to increase the number of accessions in germplasm bank collections. Once palm trees reach reproductive age, the genetic diversity impact of future outcrossings among different accessions should be estimated.

Acrocomia aculeata has tremendous potential for the biofuel and food industries as well as for other products. More understanding is needed in Mesoamerica, nonetheless, to develop commercial and conservation strategies for this species. This increased understanding depends on molecular, phenotypic, and biochemical description. For the latter to occur, traits related to oil content and the fatty acids profile require further study. Our results also contribute to future genetic diversity studies of oil-related traits in *A. aculeata* fruits in Costa Rica, similar to those studies by Da Conceição et al. (2015). A much deeper understanding of *A. aculeata* germplasm diversity will greatly enhance the species´s future conservation and management.

Acknowledgements The technical and administrative support of the personnel in the Instituto Tecnológico de Costa Rica and CENIBiot-CeNAT is acknowledged.

Author contributions Conceptualization: EAS, EAV, IM, NAZ, Methodology: EAV, EAS, VNC, NAZ, Validation: EAV, VNC, MRG, GVH, EBE, JMM, Formal Analysis: VNC, MRG, EAV, GVH, Investigation: VNC, PSC, JAOU, NAZ, EAS, EBE, GVH, JMM, Resources: EAS, IM, NAZ, GVH, EBE, JMM, Data Curation: VNC, MRG, EAV, Writing—Original Draft Preparation: VNC, EAV, MRG, Writing—Review and Editing: NAZ, EAV, EAS, MRG, IM, JMM, PSC, JAOU, Visualization: VNC, MRG, GVH, Supervision: EAV, EAS, IM, Project Administration: EAS, EAV, IM, Funding Acquisition: EAS, IM.

Funding This research was funded by Vicerrectoría de Investigación y Extensión (VIE) of the Instituto Tecnológico de Costa Rica under the project 1510136 named "Proceso de domesticación del coyol (*Acrocomia aculeata*) como alternativa bioenergética en Costa Rica, Fase II". The Centro Nacional de Innovaciones Biotecnológicas (CENIBiot-CeNAT) also funded the project (FP-026-2020).

Data availability The results presented in this study were obtained from the genotyping data available in the Supplementary Table 6.

Declarations

Conflict of interest The authors declare no conflict of interest and they have no relevant financial or non-financial interests to disclose.

Ethical statement The collection of samples in this research had the permission of the National Commission for the Management of Biodiversity in Costa Rica, under the permit number CM-ITCR-004-2021.

References

Abreu AG, Priolli RHG, Azevedo-Filho JA et al (2012) The genetic structure and mating system of *Acrocomia aculeata* (Arecaceae). Genet Mol Biol 35:119–121. https://doi.org/10.1590/s1415-47572012005000002

Alfaro-Solís JD, Montoya-Arroyo A, Jiménez VM et al (2020) *Acrocomia aculeata* fruits from three regions in Costa Rica: an assessment of biometric parameters, oil content and oil fatty acid composition to evaluate industrial potential. Agrofor Syst 94:1913–1927. https://doi.org/10.1007/s10457-020-00511-8

Arroyo-Mora JP, Sánchez-Azofeifa GA, Rivard B et al (2005) Dynamics in landscape structure and composition for the Chorotega region, Costa Rica from 1960 to 2000. Agric Ecosyst Environ 106:27–39. https://doi.org/10.1016/j.agee.2004.07.002

Balick MJ (1990) Production of coyol wine from *Acrocomia mexicana* (Arecaceae) in Honduras. Econ Bot 44:84–93. https://doi.org/10.1007/BF02861070

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Calvo-Alvarado J, McLennan B, Sánchez-Azofeifa A, Garvin T (2009) Deforestation and forest restoration in Guanacaste, Costa Rica: putting conservation policies in context. For Ecol Manage 258:931–940. https://doi.org/10.1016/j.foreco.2008.10.035

Chizmar-Fernandez C (2009) Plantas comestibles de Centroamerica. Editorial INBIO, Heredia

Clark LV, Jasieniuk M (2011) polysat: An R package for polyploid microsatellite analysis. Mol Ecol Resour 11:562–566. https://doi.org/10.1111/j.1755-0998.2011.02985.x

Coelho NHP, Tambarussi EV, Aguiar BI et al (2018) Understanding genetic diversity, spatial genetic structure, and mating system through microsatellite markers for the conservation and sustainable use of *Acrocomia aculeata* (Jacq.) Lodd. Ex Mart. Conserv Genet 19:879–891. https://doi.org/10.1007/s10592-018-1061-z

Colombo CA, Berton LHC, Diaz BG, Ferrari RA (2018) Macauba: a promising tropical palm for the production of vegetable oil. OCL Oilseeds Fats, Crop Lipids. https://doi.org/10.1051/ocl/2017038

Cornejo X, Mori SA, Aguilar R et al (2012) Phytogeography of the trees of the Osa Peninsula, Costa Rica. Brittonia 64:76–101. https://doi.org/10.1007/s12228-011-9194-0

Da Conceição LDHCS, Antoniassi R, Junqueira NTV et al (2015) Genetic diversity of macauba from natural populations of Brazil. BMC Res Notes 8:1–9. https://doi.org/10.1186/s13104-015-1335-1

de Lima NE, Carvalho AA, Meerow AW, Manfrin MH (2018) A review of the palm genus Acrocomia: neotropical green gold. Org Divers Evol 18:151–161. https://doi.org/10.1007/s13127-018-0362-x

de Lima NE, Meerow AW, Manfrin MH (2020) Genetic structure of two acrocomia ecotypes (Arecaceae) across brazilian savannas and seasonally dry forests. Tree Genet Genomes. https://doi.org/10.1007/s11295-020-01446-y

Díaz BG, Zucchi MI, Alves-Pereira A et al (2021) Genome-wide SNP analysis to assess the genetic population structure and diversity of Acrocomia species. PLoS ONE 16:1–24. https://doi.org/10.1371/journal.pone.0241025

Dransfield J, Uhl NW, Asmussen CB et al (2008) Genera palmarum: the evolution and classification of palms. Royal Botanic Gardens, London

Dray S, Dufour AB (2007) The ade4 package: Implementing the duality diagram for ecologists. J Stat Softw 22:1–20. https://doi.org/10.18637/jss.v022.i04

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Esquivel-Mimenza H, Ibrahim M, Harvey CA et al (2011) Dispersed trees in pasturelands of cattle farms in a tropical dry ecosystem. Trop Subtrop Agroecosyst 14:933–941

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Evaristo AB, Grossi JAS, de Cássia Oliveira Carneiro A et al (2016) Actual and putative potentials of macauba palm as feedstock for solid biofuel production from residues. Biomass Bioenergy 85:18–24. https://doi.org/10.1016/j.biombioe.2015.11.024

Falasca S, Ulberich A, Pitta-Alvarez S (2017) Development of agroclimatic zoning model to delimit the potential growing areas for macaw palm (*Acrocomia aculeata*). Theor Appl Climatol 129:1321–1333. https://doi.org/10.1007/s00704-016-1850-6

Gabb WM (2007) On the geology of the republic of Costa Rica. Rev Geológica América Cent 37:103–118. https://doi.org/10.15517/rgac.v0i37.4210

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186. https://doi.org/10.1111/j.1471-8278

Grayum M. (2003) Arecaceae. In: Hammel BE, Grayum MH, Herrera C, Zamora N (eds) Manual de Plantas de Costa Rica. Gimnospermas y Monocotiledóneas (Agavaceae--Musaceae). Missouri Botanical Garden Press, Missouri, USA, pp 201–293

Hall C (1984) Costa Rica: una interpretación geo-gráfica con perspectiva historia. Editorial Costa Rica, San Jose

Harvey CA, Villanueva C, Esquivel H et al (2011) Conservation value of dispersed tree cover threatened by pasture management. For Ecol Manage 261:1664–1674. https://doi.org/10.1016/j.foreco.2010.11.004

Henderson A, Galeano G, Bernal R (1995) Field guide to the palms of the Americas. Princeton University Press, New Jersey

Janzen D (1983) Costa Rican natural history. The University of Chicago Press Chicago, Chicago

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2014:1–14. https://doi.org/10.7717/peerj.281

Lanes ÉCM, SéY M, Kuki KN et al (2015) Molecular characterization and population structure of the Macaw Palm, *Acrocomia aculeata* (Arecaceae), ex situ Germplasm collection using microsatellites markers. J Hered 106:102–112. https://doi.org/10.1093/jhered/esu073

Lanes ÉCM, Motoike SY, Kuki KN et al (2016) Mating system and genetic composition of the macaw palm (*Acrocomia aculeata*): implications for breeding and genetic conservation programs. J Hered 107:527–536. https://doi.org/10.1093/jhered/esw038

Laviola BG, dos Santos A, Rodrigues EV et al (2022) Structure and genetic diversity of macauba [*Acrocomia aculeata* (Jacq.) Lodd. ex Mart.] approached by SNP markers to assist breeding strategies. Genet Resour Crop Evol 69:1179–1191. https://doi.org/10.1007/s10722-021-01295-1

Lescano CH, Oliveira IP, Silva LR et al (2015) Nutrients content, characterization and oil extraction from *Acrocomia aculeata* (Jacq.) Lodd. fruits. African J Food Sci 9:113–119. https://doi.org/10.5897/ajfs2014.1212

Lieb VM, Schex R, Esquivel P et al (2019) Fatty acids and triacylglycerols in the mesocarp and kernel oils of maturing Costa Rican *Acrocomia aculeata* fruits. NFS J 14–15:6–13. https://doi.org/10.1016/j.nfs.2019.02.002

Mengistu FG, Motoike SY, Caixeta ET et al (2016a) Cross-species amplification and characterization of new microsatellite markers for the macaw palm, *Acrocomia aculeata* (Arecaceae). Plant Genet Resour Characterisation Util 14:163–172. https://doi.org/10.1017/S1479262115000179

Mengistu FG, Motoike SY, Cruz CD (2016b) Molecular characterization and genetic diversity of the macaw palm Ex situ germplasm collection revealed by microsatellite markers. Diversity. https://doi.org/10.3390/d8040020

Morcote-Ríos G, Bernal R (2001) Remains of palms (palmae) at archaeological sites in the new world: a review. Bot Rev 67:309–350. https://doi.org/10.1007/BF02858098

Navarro-Díaz HJ, Gonzalez SL, Irigaray B et al (2014) Macauba oil as an alternative feedstock for biodiesel: characterization and ester conversion by the supercritical method. J Supercrit Fluids 93:130–137. https://doi.org/10.1016/j.supflu.2013.11.008

Nucci SM (2007) Development, characterization and analyses of the use of microsatellite markers in macaw palm population genetics. MSc. Dissertation. University of Sao Paulo

Nucci SM, Azevedo-Filho JA, Colombo CA et al (2008) Development and characterization of microsatellites markers from the macaw. Mol Ecol Resour 8:224–226. https://doi.org/10.1111/j.1471-8286.2007.01932.x

Paradis E (2010) Pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26:419–420. https://doi.org/10.1093/bioinformatics/btp696

Plath M, Moser C, Bailis R et al (2016) A novel bioenergy feedstock in Latin America? Cultivation potential of *Acrocomia aculeata* under current and future climate conditions. Biomass Bioenerg 91:186–195. https://doi.org/10.1016/j.biombioe.2016.04.009

Pritchard J, Matthew DP (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1007/s10681-008-9788-0

Quirós-Guerrero L, Albertazzi F, Araya-Valverde E et al (2019) Phenolic variation among *Chamaecrista nictitans* subspecies and varieties revealed through UPLC-ESI(-)-MS/MS chemical fingerprinting. Metabolomics. https://doi.org/10.1007/s11306-019-1475-8

Ramos SLF, De Macêdo JLV, Lopes MTG et al (2012) Microsatellite loci for tucumã of Amazonas (*Astrocaryum aculeatum*) and amplification in other Arecaceae. Am J Bot 99:2011–2013. https://doi.org/10.3732/ajb.1100607

Scariot A (1998) Seed dispersal and predation of the Palm *Acrocomia aculeata*. Principes 42:5–8

Scariot A, Lleras E, Hay JD (1995) Flowering and fruiting phenologies of the Palm *Acrocomia aculeata*: patterns and consequences. Biotropica 27:168. https://doi.org/10.2307/2388992

Schex R, Lieb VM, Jiménez VM et al (2018) HPLC-DAD-APCI/ESI-MSn analysis of carotenoids and α-tocopherol in Costa Rican *Acrocomia aculeata* fruits of varying maturity stages. Food Res Int 105:645–653. https://doi.org/10.1016/j.foodres.2017.11.041

Silva LCC, do Lemos RC, Carvalho CGI et al (2017) Genetic diversity and structure of macaw palm: implications for genetic variability sampling. Rev Árvore. https://doi.org/10.1590/1806-90882017000500008

Solano-Quintero J, Villalobos-Flores R (2001) Aspectos fisiográficos aplicados a un bosquejo de regionalización geográfico climático de Costa Rica. Top Meteor Oceanog 8:26–39

Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197

Venables WN, Ripley BD (2002) Modern applied statistics with S, Fourth. Springer, New York

Winter DJ (2012) MMOD: an R library for the calculation of population differentiation statistics. Mol Ecol Resour 12:1158–1160. https://doi.org/10.1111/j.1755-0998.2012.03174.x

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. J Evol 19:395–420. https://doi.org/10.2307/2406450

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.