PLOS ONE

RESEARCH ARTICLE

Genome-wide SNP analysis to assess the genetic population structure and diversity of Acrocomia species

Brenda Gabriela Díaz[1], Maria Imaculada Zucchi[1,2], Alessandro Alves-Pereira[1], Caléo Panhoca de Almeida[2], Aline Costa Lima Moraes[1], Suelen Alves Vianna[2], Joaquim Azevedo-Filho[2], Carlos Augusto Colombo[2]*

1 Biology Institute, University of Campinas UNICAMP, Campinas-SP, Brazil, **2** Centro de Pesquisa de Recursos Genéticos Vegetais, Instituto Agronômico-IAC, Campinas-SP, Brazil

* carlos.colombo@sp.gov.br

OPEN ACCESS

Citation: Díaz BG, Zucchi MI, Alves-Pereira A, de Almeida CP, Moraes ACL, Vianna SA, et al. (2021) Genome-wide SNP analysis to assess the genetic population structure and diversity of Acrocomia species. PLoS ONE 16(7): e0241025. https://doi.org/10.1371/journal.pone.0241025

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: October 6, 2020

Accepted: July 5, 2021

Published: July 20, 2021

Data Availability Statement: All relevant data are within the paper and its Supporting information files.

Funding: This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo-FAPESP https://fapesp.br/ under doctoral scholarship conceded to BGD author (process 2017/22987-2) and research grants (2014/23591-7) conceded to CAC author, and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) https://www.gov.br/capes/pt-br under the

Abstract

Acrocomia (Arecaceae) is a genus widely distributed in tropical and subtropical America that has been achieving economic interest due to the great potential of oil production of some of its species. In particular *A. aculeata*, due to its vocation to supply oil with the same productive capacity as the oil palm (*Elaeis guineenses*) even in areas with water deficit. Although eight species are recognized in the genus, the taxonomic classification based on morphology and geographic distribution is still controversial. Knowledge about the genetic diversity and population structure of the species is limited, which has limited the understanding of the genetic relationships and the orientation of management, conservation, and genetic improvement activities of species of the genus. In the present study, we analyzed the genomic diversity and population structure of Acrocomia genus, including 172 samples from seven species, with a focus on *A. aculeata* with 117 samples covering a wide geographical area of occurrence of the species, using Single Nucleotide Polymorphism (SNP) markers originated from Genotyping By Sequencing (GBS).The genetic structure of the Acrocomia species were partially congruent with the current taxonomic classification based on morphological characters, recovering the separation of the species *A. aculeata*, *A. totai*, *A. crispa* and *A. intumescens* as distinct taxonomic groups. However, the species *A. media* was attributed to the cluster of *A. aculeata* while *A. hassleri* and *A. glauscescens* were grouped together with *A. totai*. The species that showed the highest and lowest genetic diversity were *A. totai* and *A. media*, respectively. When analyzed separately, the species *A. aculeata* showed a strong genetic structure, forming two genetic groups, the first represented mainly by genotypes from Brazil and the second by accessions from Central and North American countries. Greater genetic diversity was found in Brazil when compared to the other countries. Our results on the genetic diversity of the genus are unprecedented, as is also establishes new insights on the genomic relationships between Acrocomia species. It is also the first study to provide a more global view of the genomic diversity of *A. aculeata*. We also highlight the applicability of genomic data as a reference for future studies on genetic diversity, taxonomy, evolution and phylogeny of the Acrocomia genus, as well as to

doctoral scholarship conceded to BGD author (Project Biocomputacional 23038.001614/2016-52). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

support strategies for the conservation, exploration and breeding of Acrocomia species and in particular *A. aculeata*.

Introduction

The genus Acrocomia is endemic to tropical and subtropical America. This genus is one of the most taxonomically complex concerning species in the family Arecaceae [1]. Taxonomic classifications of Acrocomia are mostly limited to the description of species based on morphological and geographical distribution information. However, extensive morphological plasticity, especially for species with wide geographical distribution, has hindered the taxonomic resolution of species. Since the description of the genus Acrocomia by Martius in 1824 [2], many species have been included and removed from the genus. From the most recent classifications, Henderson et al. [3] attributed only two species to the genus. One is *A. aculeata* (Jacq.) Lodd. ex Mart., which is large (arboreal) and widely distributed throughout Central, North, and South America. The other is *A. hassleri* (Barb. Rodr.) WJ Hahn, which is small size and is restricted to the Cerrado savanna in Brazil and part of Paraguay. Lorenzi et al. [4] recognized seven species for the genus. Six of these are found in Brazil: *A. aculeata*, *A. intumescens*, and *A. totai* have an arboreal size and are mainly differentiated by the stipe characteristics. *A. hassleri*, *A. glaucescens*, and *A. emensis* are small size and are differentiated by their height. The seventh species, *A. crispa*, has an arboreal size and is endemic to Cuba. The Plant List [5] and The Palmweb [6] recognized *A. media* as the eighth species. It is endemic to Puerto Rico. Therefore, the systematics of the genus Acrocomia remain controversial, with the number of species not well resolved and very few studies having addressed species delimitation, population genetic diversity and structure, and inter-species relationships.

As the exclusive use of morphology may not be enough for accurate species delimitation, molecular markers can provide additional information to species classification [7–9]. However, genetic differentiation can be influenced by forces acting in different DNA regions. Natural selection also can promote the population genetic structure acting in non-neutral loci (loci under selection). While gene flow and genetic drift are stochastic processes driving the population genetic structure at neutral loci [10,11].

In a genus as Acrocomia with a wide geographic distribution, some of the morphological differentiation at the species level may be due to responses to local adaption. Nevertheless, morphological traits may experience similar selective pressures and evolve convergently could bias morphological-based taxonomy. In addition, the lack of resolution in the genus could be a result of gene flow between species mainly in areas of co-occurrence [12], independently of selective forces. In this sense, neutral molecular markers are interesting to assess the genetic diversity of populations because they provide unbiased estimates of random processes such as genetic drift [10].

A. aculeata, *A. totai*, and *A. intumescens* are the species of greatest economic interest, mainly due to their many applications and products obtained, with practically all parts of the palms used. The fruits are important for the production of vegetable oil as a bioenergy source and flour for human and animal consumption [13] as well as for medicinal uses [13,14]. Of these three species, *A. aculeata* is distinguished by its high productive capacity and oil quality [15]. The oil production of 4,000 oil L/ha/year estimated in Brazil far surpasses soybean (*Glycine max)* (400 L/ha) [15] and equals the oil palm (*Elaeis guineenses*), which is considered the oilseed with the highest oil yield per area, with an oil production volume of up to 6,000 L/ha [16,17].

A. *aculeata* is an arborescent heliophile and monoecious. This species produces unisexual flowers in the same inflorescence [3,4]. According to Abreu et al. [18], the species has a mixed reproduction system. Nonetheless, [19,20] indicate that the species is preferably allogamous.. It is a diploid species (2n = 30), with a genome size of 2.8 Gbp [21]. *A*. *aculeata* has a wide geographic distribution, occurring naturally from northern Mexico and the Antilles to southern Brazil [3,4,22,23]. It is commonly found in savanna areas, but also is found in tropical and subtropical forests, and in the dry forests of Caatinga [3,4,24] and has adapted to sandy soils and regions with low water availability [25]. Besides being a perennial species, it is beneficial for soil management and conservation since its useful life can exceed 50 years. Colombo et al. [15] identified *A*. *aculeata* as a promising resource for sustainable large-scale production of vegetable oil.

Although the economic interest in some Acrocomia species is growing, little is known about infrageneric relationships, levels of genetic diversity and structure, and patterns of gene flow at the genus level. The population genetics approach can assist in species delimitation and provide reference information on the genetic diversity and structure within and between species. Such knowledge is essential for more efficient management and economic exploration of the species and can guide strategies for domestication and conservation of these genetic resources [26–28]. *A*. *aculeata* is an emerging crop with incipient domestication. The analysis of genetic diversity of *A*. *aculeata* is crucial to guide the selection of the most promising materials for crop use, to maximize genetic gains, and to more effectively contribute to the creation of commercial cultivars.

In this context, molecular markers have been broadly adopted in plants as an essential tool to investigate genetic diversity in ecological, phylogenetic, and evolutionary studies. In addition, they have been widely used for direct management, conservation, and genetic breeding of several species [29]. Next generation sequencing (NGS) has facilitated the identification of single nucleotide polymorphisms (SNPs) and has become the most used molecular marker due to the abundance, wide genomic coverage, access to neutral variations and loci under selection, offering a fast and high-yield genotyping, with low error rates and ability to identify SNP without the need for reference genomes [26,30–32]. The applications of this marker are quite wide and in different areas of science mainly in the use of high density panels for the purposes of genomic selection, conservation genetics, genomic landscape and breeding [26,33–35]. However, although very advantageous, SNPs markers have not been used so far in genetic studies of Acrocomia species.

In Acrocomia, microsatellites or simple sequence repeats (SSR) have been the most used molecular markers, with the main objective of evaluating the genetic diversity and structure of natural populations and germplasm banks [18,19,36–38]. Other approaches include the use of internal transcribed ribosomal 18S-26S spacer (ITS region) [39] and random amplification of polymorphic DNA (RAPD) markers [40]. However, most studies have focused on *A*. *aculeata* [18,20,36–38]. Only one study has analyzed the genetic diversity of *A*. *totai* [12].

Genetic studies of species occurring in large geographic areas associated with analyzes of a large number of neutral loci provide accurate information on how stochastic events such as genetic drift and gene flow are acting on diversity and genetic structure [10]. Essential information for breeding, management of germplasm resources and conservation in the different areas where species occur [41]. However, the genetic diversity studies carried out in *A*. *aculeata* refer to a very small geographical sample, with Brazilian genotypes mainly from the states of São Paulo and Minas Gerais [18,20,36,37,40] and only a single study has evaluated the genetic diversity of natural populations of *A*. *aculeata (called A*. *mexicana)* from another country besides Brazil, that being Mexico [42]. Therefore, there is a huge gap in information about the

extent of genomic variation in the Acrocomia species in other areas of occurrence besides Brazil.

In view of the controversies regarding the taxonomic resolution of the genus Acrocomia and the lack of information on the genetic diversity of its species, we applied a population genomic approach based on genome-wide SNPs to assess the genetic population structure and diversity of Acrocomia species from wide heterogeneity of environmental conditions where their occur. We hypothesized that the current taxonomic designation of the Acrocomia species is supported by population genetic structure based on genomic neutral data.

The present study is unprecedented because it was conducted using seven Acrocomia species and a wide sampling of *A*. *aculeata* from several countries in the American continent. This is the first study carried out with SNP markers for the genus.

Material and methods

Plant material and DNA extraction

In the present study, we considered 172 samples to represent seven from eight Acrocomia species: *A*. *aculeata*, *A*. *totai*, *A*. *intumescens*, *A*. *media*, *A*. *crispa*, *A*. *hassleri*, *A*. *glaucescens*. This study was carried out in accordance with the Ministério do Meio Ambiente do Brazil and registered in the Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado(SISGEN) within the number A69E071 and was approved by the Comissão de Ética Cientifica of Instituto Agronômico de Campinas (CEtIAC). The samples were obtained from different locations in order to represent the entire geographic distribution described in the literature for the respective species [3,4]. The species *A*. *aculeata*, with a greater distribution in America, was represented by samples from five countries (Fig 1 and S1 Table).

The total genomic DNA was extracted from leaf material using the Doyle & Doyle [43] protocol. We evaluated the quality and quantity of DNA on a 1% agarose gel, on the NanoVue ™ Plus spectrophotometer (GE Healthcare), and through fluorescence using the QubitTM dsDNA BR Assay (Qubit—Life Technologies). Based on the obtained reading, we standardized the DNA to a concentration of 30ng.μl-1.

GBS library preparation and high-throughput sequencing

To obtain SNPs, we developed genomic libraries using the two-enzyme genotyping-by-sequencing (GBS) technique according to the protocol described by Poland et al. [44], with modifications. We digested 7 μl of the genomic DNA [30ng.μl-1] from each sample at 37˚ C for 12 h with the enzymes *NsiI* and *MspI*. Subsequently, 0.02 μM of specific adapters for the Illumina technology (containing the barcode sequences and complementary to the Illumina ™ primers for sequencing) were connected to the fragments ends generated in the digestion. The ligation reaction was carried out at 22˚C for 2 h; 65˚C for 20 min; 10˚C indefinitely.

After adapters ligation, we purified the samples using QIAquick PCR Purification Kit (Qiagen). The library was enriched by PCR. We performed eight replicates, each one containing 10 μL of purified and amplified ligation, using 12.5 μL of Phusion® High-Fidelity PCR Master Mix NEB (New England Biolabs Inc.), and 2 μl of Illumina forward and reverse [10 μM] primers ™, in a final volume of 25 μL, using the following amplification program: 95˚C for 30 s, followed by 16 cycles of 95˚C for 10 s, 62˚C for 20 s, 72˚C for 30 s, ending at 72˚C for 5 min. Finally, we purified the library using QIAgen's QIAquick PCR Purification Kit.

The verification of average size of the DNA fragments using the Agilent DNA 12,000 kit and the 2100 Bioanalyzer System (Agilent) equipment. The libraries were quantified by qPCR using the CFX 384 real-time thermocycler (BioRad) with the aid of the KAPA Library

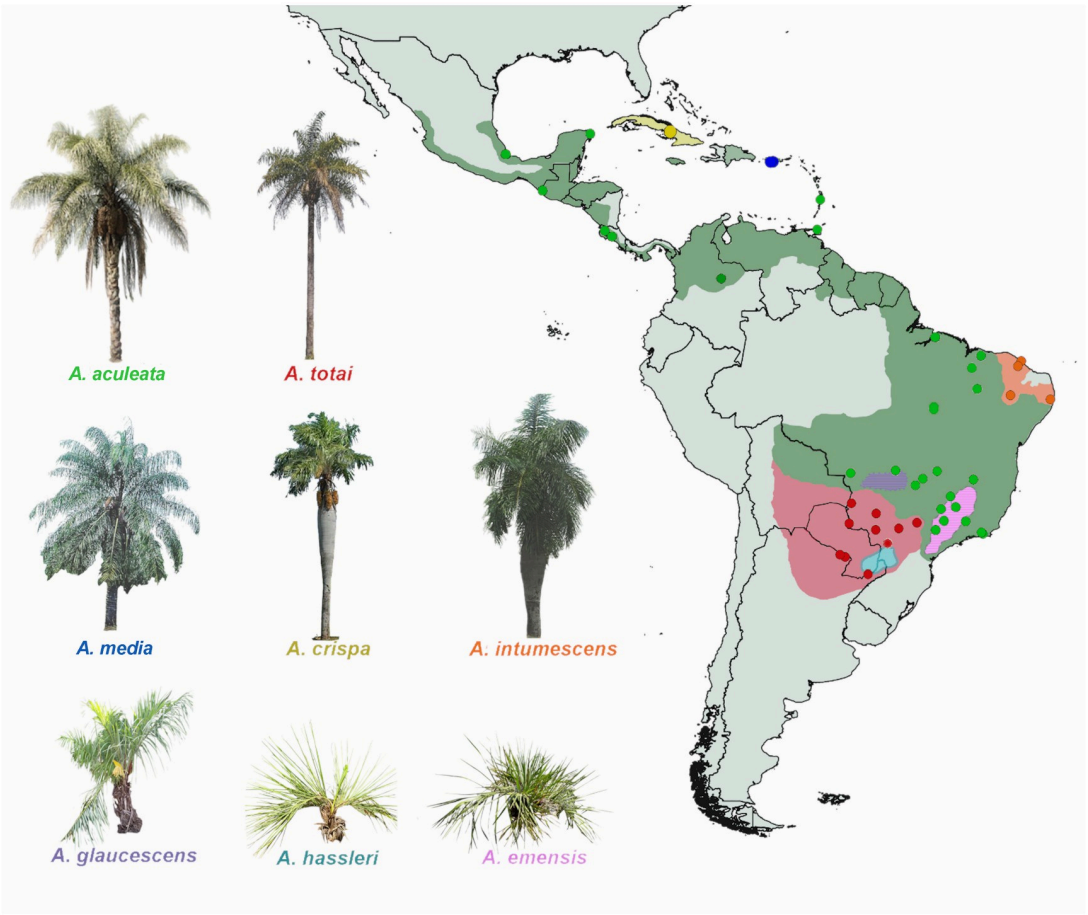

Fig 1. Schematic map of Acrocomia species distribution and geographic location and origin of samples. Data used to generate the species distribution (Colored shading) are based on occurrence record data from GBIF (Global Biodiversity Information Facility www.gbif.org) and [4]. Circles represent geographical location and origin of samples in this study. Image sources: *A. aculeata*, *A. totai*, *A. hassleri*, *A. glaucescens*. *A. emensis* (B. G. Díaz); *A. intumescens*; *A. media* and *A. crispa* (S. A. Vianna).

https://doi.org/10.1371/journal.pone.0241025.g001

Quantification kit (KAPA Biosystems). We prepared two libraries of 96 samples each, which were sequenced using Illumina's NextSeq 500/550 Mid Output Kit v2.5 (150-cycle), on the NextSeq550 platform (Illumina Inc., San Diego, CA).

SNP identification

We performed the identification of SNP markers using the Stacks *v*. 1.42 *pipeline* [45]. We used the *process_radtags* module to demultiplex the samples and we remove the low-quality reads (reads that either dropped the Phred score of 10, contained Illumina adapters, uncalled bases" Ns "or without restriction sites). As there is no reference genome for Acrocomia, we used the *denovo* pipeline of Stacks starting with the *ustacks* module to identify groups of putatively homologous reads (putative loci). This step was performed for each sample separately with the following parameters: the minimum sequencing depth (-m) $\geq$ 3, the maximum distance between *stacks* (-M) = 2; and the maximum distance between primary and secondary sequences [30] = 2. Subsequently, a locus catalog was built using the *cstacks* module, allowing a maximum of 2 differences between stacks [30] from different individuals. We eliminated loci with lower values of probability (lnl_lim -10) by the *rxstacks* correction module. The SNPs

were filtered using the *populations* module, retaining only one SNP per sequence, with a minimum depth of 3X sequencing, minor allele frequency ≥ 0.01, and minimum occurrence in 75% of individuals in each location/population. After filtering, we identified 3269 SNPs (S1 File) considering all the samples.

Neutral loci identification

Three approaches were applied to identify outlier loci, putatively under selection. For the first approach, a based population structure using PCA analyses, with no a priori information about the number of populations, was employed using the *pcadapt* package [46], on the R platform [47]. In the second approach, we used the *fsthet* package [48] based on Wright's FST fixation index [49] to identify the loci with deviation from the expected relationship between FST and heterozygosity (HE), using the island migration model [50]. For the third approach we adopted to test the association of environmental variables with the genetic variation of SNP with the LFMM (Latent Factor Mixed Models) [51], using the LEA package [52]. Nineteen bioclimatic variables were obtained from WorldClim database [53]. However, the analysis was performed with the 13 variables (correlation ≤ 0.8) (S2 File). The number of latent factors was determined, using the SNMF function [51] to estimate the number of ancestral populations with ten repetitions of K = 1 to 15. For the LFMM function, five replicates were performed with 200,000 MCMC interactions after 50,000 burn-ins, considering K = 8 (species) and 6 (*A*. *aculeata*), identified by the SNMF function. The p-values were adjusted using the genomic inflation factor (λ) and the false discovery rates (FDR) were defined using the Benjamini-Hochberg algorithm, considering FDR = 0.1 (see S2 File for more methods details).

The identification of outlier loci was performed independently for the following groups: 1) In the genus Acrocomia, considering the species as groups, and 2) within *A*. *aculeata*, considering as groups the samples' countries of origin. We considered as outlier loci those shared between the three identification methods (S2 Table). Consequently, we adopted the remaining SNPs considered neutral for the analysis of population genomic diversity and structure.

Population structure

We used all samples (S1 Table) to perform the analysis of the genomic structure for de Acrocomia genus and to infer the number of the most likely groups using the software Structure v.2.3.4 [54], considering only neutral SNPs (3227). We also used the same software to assess the genomic structure of *A*. *aculeata* separately, considering 3259 neutral SNPs identified for the species. Each analysis in Structure was performed with a burn-in of 100,000 interactions, followed by 500,000 repetitions of the Markov Chain Monte Carlo (MCMC) in 10 independent simulations, and without prior information to define the clusters. The number of clusters (K) was determined using the average likelihood values of the ΔK method [55] implemented in the program Structure Harvester [56]. The ancestry coefficients of each sample was given by the alignment of five repetitions of the best *K* through the CLUMPP method [57] by the software CLUMPAK [58].

To visualize the genetic relationships among Acrocomia species and within the *A*. *aculeata*, we obtained the Nei genetic distance [59] between the individuals of each data set, and the Neighbor-Joining [60] hierarchical classification method with 20000 bootstrap repetitions, using the poppr package [61] on R [47].

In addition, the Principal Component Analysis (PCoA) was also carried out through the ade4 package [62] to explore the genetic structure of the different groups using only neutral SNPs, and was visualized graphically by the ggplot2 package [63].

Finally, given the wide distribution of *A*. *aculeata* populations we used the Monmonier's function [64] in the R package adegenet 1.3-1 [65] to identify possible barriers to gene flow between populations associated with geography. Random noise was eliminated using Principal Coordinate Analysis with the dudi.pco function from the ade4 package [62]. The connection network was calculated using a Gabriel graph [66].

Analysis of genomic diversity

We conducted the population diversity analysis only with the SNP data set identified as neutral for two groups or taxonomic levels: 1) The genus Acrocomia (except the species *A*. *hassleri* and *A*. *glaucescens* as they contain only one individual for each species), and 2) *A*. *aculeata*. Population estimates of allelic richness, number of total alleles by locus, observed heterozygosity, expected heterozygosity, and inbreeding coefficient were calculated using the diveRsity [67], poppr [61], and the PopGenKit packages [68] on R platform [47]. To minimize the effect of differences in the number of samples of each population, we calculated the allelic richness (Ar) and the richness of private alleles (ap) for populations of each group or taxonomic level, by the rarefaction method implemented in the software HP-Rare v.1.1 [69].

Contemporary effective population size (*Ne*) was estimated using NeEstimator v2.01 [70]. The PGDSpider software [71] was used to convert the *vcf* file into GENEPOP input file. Estimation was assessed based on LD method [72], assuming random mating and setting a critical value of a minimum allele frequency of 0.05 as is a common value used in SNP-based studies.

Results

Genomic structure of Acrocomia spp.

The neutral datasets for the different groups were constructed by removing the outliers. After the removal of outlier loci (S2 Table), genus Acrocomia (all species) and *A*. *aculeata* contained, respectively, 3227 and 3259 neutral loci, that were used for the analyses of genetic structure and diversity.

The Bayesian analyses to access the genomic population structure of 172 samples of Acrocomia species based on 3227 neutral SNPs suggested the existence of seven genetic groups (Fig 2) based on the ΔK (S1 Fig). Samples with an ancestry coefficients > 0.75 and < 0.75 were assigned to the "pure group" and "admixture group", respectively. Based on the classification of Lorenzi [4] and the geographic distribution of the species, we observed a substructure of samples considered to be *A*. *aculeata*. Two well-defined subgroups (clusters 1 and 3) strongly associated with the geographical origin of the samples were evident. Cluster 1 (Fig 2) was composed of 38 samples of *A*. *aculeata* from Central and North America (Costa Rica, Trinidad and Tobago, Puerto Rico, and Mexico) and Colombia. Cluster 2 (Fig 2) comprised 39 samples of *A*. *totai* and five samples considered as *A*. *aculeata*. Of the latter, four were collected in the state of Parana, southeastern Brazil, (XAM, PR) and one in state of Tocantins, northern Brazil (PAL). The samples from Campo Grande (CGR) showed low mixture levels with clusters 1 and 5 of *A*. *aculeata*. Cluster 3 (Fig 2) consisted of 39 samples from Brazil. The majority (n = 34) of these samples were from the southeast region of the country, with five from the north region (BEL population). Cluster 4 was exclusively formed by *A*. *crispa* samples, with a 100% probability of assignment to the cluster.

Based on ancestry coefficients ≤ 0.75, some samples were assigned to an admixture group. Twenty samples of *A*. *aculeata* from the central-west, north, and northeast regions of Brazil, and all samples of *A*. *intumescens* displayed a similar genomic composition, with a median level of assignment (≥ 0.50). A genetic admixture of *A*. *aculeata* samples in cluster 5 (Fig 2) with samples mainly from clusters 1 and 3 was evident. *A*. *intumescens* samples presented a

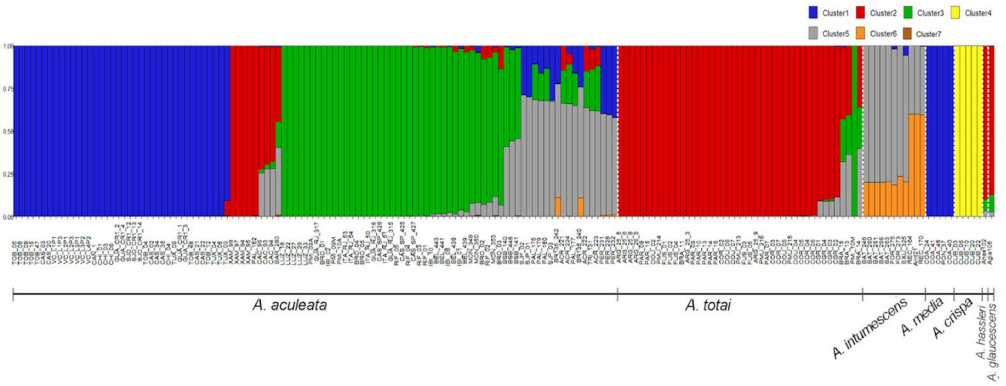

Fig 2. Genomic structure of 172 samples from Acrocomia species based on 3227 neutral SNP loci. The y-axis is the population membership, and the x-axis is the sample. Each vertical bar represents a sample and color represent separate clusters ($K = 7$).

https://doi.org/10.1371/journal.pone.0241025.g002

mixture of clusters 5 and 6, with cluster 6 being practically exclusive to the species. Individuals from Cáceres, MT (CAC), and Braúna, São Paulo (SP) (BRA), with a greater assignment to cluster 2, also showed a significant degree of admixture with clusters 3 and 5.

The NJ and PCoA analyses (Fig 3a and 3b) performed with all the samples showed strong agreement with the results of the Bayesian analysis performed using Structure software. However, the NJ tree showed higher resolution in group/cluster recovery than the PCoA. In both analyses, *A*. *crispa* was clearly separated from the rest of the Acrocomia species. In addition, there is a clear genomic differentiation between *A*. *aculeata* and *A*. *totai*. Similar to the results obtained using the Structure software, the NJ analysis also recovered the substructure within *A*. *aculeata*, separating the Brazilian samples from those from other countries (Fig 3a). This separation did not result from the PCoA (Fig 3b). In agreement with the results obtained using the Structure software, both PCoA and NJ grouped *A*. *media* and *A*. *intumescens* samples into the cluster formed mainly by *A*. *aculeata*, with *A*. *hassleri* and *A*. *glaucescens* grouped into the *A*. *totai* cluster. The results of NJ and PCoA also agreed concerning the allocation of samples from Xambré, PR (XAM) originally considered as *A*. *aculeata* in the cluster of *A*. *totai*. Samples from Braúna, SP (BRA) and Cáceres, MT (CAC), which were identified as an admixture by the Structure software, occupied an intermediate position between the clusters formed mainly by *A*. *aculeata* and *A*. *totai* in the PCoA.

Based on the Structure software results (Fig 2) and NJ and PCoA data (Fig 3a and 3b), the samples from Xambrê, PR (XAM) previously considered *A*. *aculeata* were treated as *A*. *totai* species for further analysis of differentiation and genomic diversity. The F_{ST} values enabled a moderate genetic differentiation between species, with an average value of 0.469. The F_{ST} values between species (Table 1) ranged from 0.083 (*A*. *aculeata* vs. *A*. *totai)* to 0.946 (*A*. *media* vs. *A*. *crispa*). In agreement with the genomic structure analysis findings, all comparisons between *A*. *crispa* and the other species showed higher values of F_{ST}, demonstrating a greater degree of genetic differentiation of *A*. *crispa* with the other species.

Genomic diversity within species

The mean number of alleles per locus of the five Acrocomia species ranged from 0.017 to 0.601. *A*. *aculeata* had the highest mean and *A*. *media* had the lowest mean (Table 2). The genomic diversity based on the average expected heterozygosity (H_E) in the species ranged

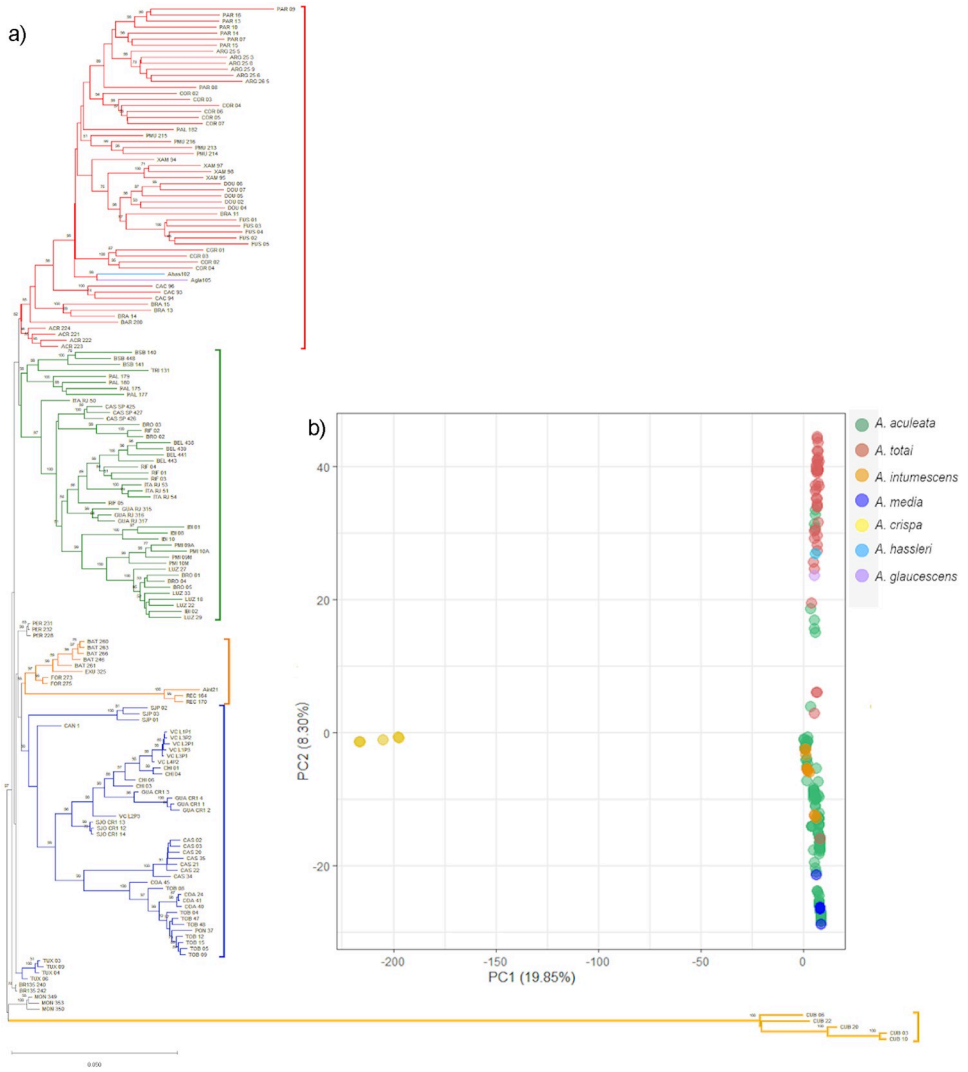

Fig 3. Neighbor-joining [60] tree and principal components analysis (PCoA) of Acrocomia species. a) Scatterplot of the principal components analysis (PCoA) showing the dispersion of samples across the first two principal components and b) Neighbor-Joining dendrogram based on Nei´s genetic distance. Bootstrap support of nodes is shown.

https://doi.org/10.1371/journal.pone.0241025.g003

from 0.106 in *A. totai* to 0.005 in *A. media*. However, *A. crispa* was the species with the highest allelic richness (2.29) and the highest allelic richness of private alleles (0.17), while *A. media* presented the lowest values of allelic richness and allelic richness of private alleles (1.08 and 0.01, respectively). The inbreeding coefficient (F) values were high for all species, indicating

Table 1. Pairwise F_{ST} estimates among five species of Acrocomia.

	A. aculeata	*A. totai*	*A. intumescens*	*A. media*	*A. crispa*
A. aculeata	0.000				
A. totai	0.083	0.000			
A. intumescens	0.128	0.194	0.000		
A. media	0.133	0.235	0.700	0.000	
A. crispa	0.673	0.687	0.912	0.946	0.000

https://doi.org/10.1371/journal.pone.0241025.t001

Table 2. Genetic diversity parameter estimates for Acrocomia species calculated from 3227 neutral loci SNPs.

Species	Na	I	H_O	H_E	Ar	PAr	f
A. aculeata	1.601	0.157	0.031	0.093	1.20	0.03	0.479
A. totai	1.534	0.176	0.074	0.106	1.23	0.07	0.262
A. intumescens	1.053	0.037	0.011	0.025	1.10	0.02	0.483
A. media	0.993	0.007	0.006	0.005	1.08	0.01	-0.145
A. crispa	0.630	0.028	0.006	0.020	2.29	0.21	0.591

Mean of different alleles (Na), Shannon's Index (I), Observed (H_O) and Expected (H_E) Heterozygosity, allelic richness (Ar), private alleles richness (PAr) and Fixation index (f).

https://doi.org/10.1371/journal.pone.0241025.t002

relatively high levels of inbreeding in Acrocomia species, with the exception of *A. media*, which presented negative values (Table 2). The current of effective population size (*Ne*) range from 0.6 to infinite. The lowest values were observed for *A. aculeata*, while *A. totai* and *A. media* presented the highest *Ne*. Nevertheless, ranges of confidence for *A. media*, were extremely high (infinite), making these estimates unreliable (S3 Table).

Genomic structure of *A. aculeata*

The population structure of all the *A. aculeata* samples was evaluated using 3259 hypothetically neutral SNPs. Using the method of Evanno [55] the most probable Δk was $K = 2$ (S2 Fig). This finding supported the presence of two genetically distinct subpopulations previously identified in the structure analysis at the genus level (Fig 2). The two groups were mainly associated with geographical origin, given that samples from Central and North America (Colombia, Costa Rica, Trinidad and Tobago, and Mexico) were grouped in cluster 1, and most of the collected in Brazil were grouped in cluster 2 (Fig 4).

The same two groups identified using the Structure software were also visualized by using the first two PCoA axes as well as the NJ dendrogram. These analyses clearly revealed the formation of two distinct genetic groups within *A. aculeata* (Fig 4b and 4c). In addition, the analysis using Monmonier's algorithm [64] revealed three potential barriers to gene flow separating the Northern Populations from the rest of Sothern populations from Brazil (S3 Fig), which are suggested to be geographically separated by the Amazon Rainforest.

Based on the NJ dendrogram, two large groups were assigned based on geographical origin, separating all individuals from the North and Central America in one node [41] from the Brazilian samples (blue). Two main subgroups were evident in the Northern group. One subgroup contained samples from Peritoró (PER) and São Jose dos Patos (SJP) from Maranhão, Brazil. The other subgroup contained the remaining samples. Interestingly, individuals from Tuxtla Chico, Chiapas (TUX) in Mexico formed a separate cluster from the other samples from Mexico and Colombia, Costa Rica, Trinidad and Tobago, and Puerto Rico (Fig 4b).

The second PCoA axis comprised three samples from Cáceres, MT (CAC). These samples formed a subgroup that was very distant from the other samples of *A. aculeata*. However, the Structure and NJ dendrogram data were not able to discriminate these samples and grouped with individuals from Brazil (Fig 4a and 4c).

The 'South' group (Cluster 2 in Fig 4a) contained most of the samples from Brazil. The samples collected in Maranda formed a different cluster from the other samples. However, most clusters reflected a strong relationship with the samples geographic origins, with the exception of samples collected in Belém, PA (BEL), northern Brazil, which were more closely related to samples from Rio de Janeiro and São Paulo located in southeastern Brazil. It is also noteworthy

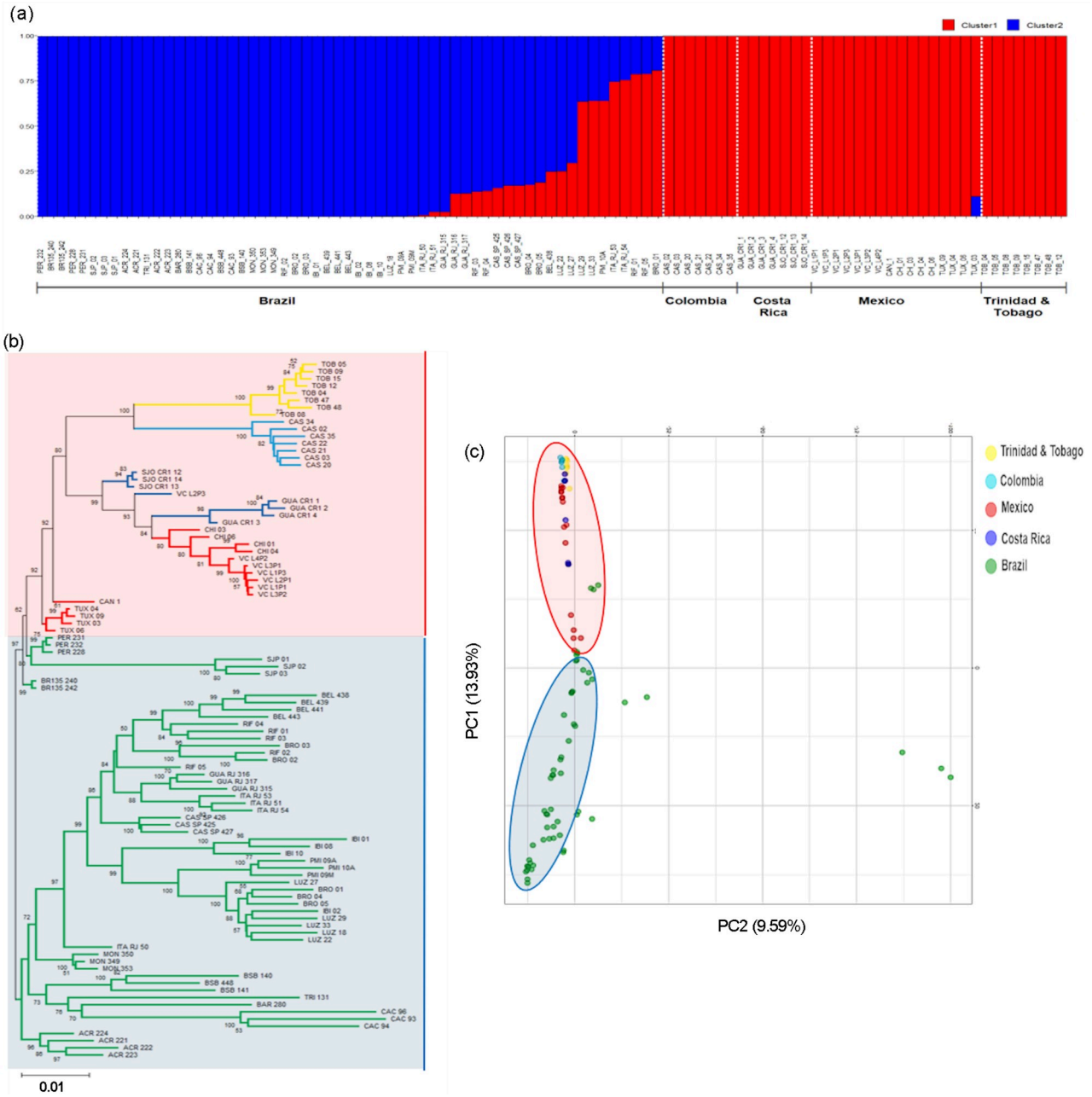

Fig 4. Population genomic structure within *A. aculeata*, based on 3256 neutral loci SNPs. a) Genomic structure from Bayesian analyses ($K = 2$). The y-axis is the population membership, and the x-axis is the sample. Each bar represents an individual and each color is inferred membership in each of the cluster; b) Neighbor-Joining dendrogram based on Nei´s genetic distance. Bootstrap support of nodes is shown. Groups: Northern genetic group (Blue); southern genetic group [41] and c) Scatterplot of the principal components analysis (PCoA) showing the dispersion of samples across the first two principal components.

https://doi.org/10.1371/journal.pone.0241025.g004

that five samples collected in the Brazilian State of Maranhão (PER and SJP) were more closely related with the 'North' group, as evident by the cluster 1 considering the assignment probability of 0.75 in the Structure software analysis (Fig 4a). This result was also corroborated by the NJ and PCoA (Fig 4b and 4c).

Table 3. Genetic diversity parameter estimates for *A. aculeata* calculated from 3259 neutral loci of SNPs.

Country	Na	I	H_O	H_E	Ar	PAr	f
Trinidad &Tobago	1.001	0.013	0.008	0.008	1.09	0.01	0.038
Colombia	1.009	0.018	0.012	0.012	1.09	0.01	-0.014
Mexico	0.994	0.009	0.004	0.005	1.09	0.01	0.147
Costa Rica	0.981	0.012	0.009	0.008	1.11	0.01	-0.139
Brazil	1.441	0.135	0.043	0.081	1.44	0.33	0.377
Brazilian State							
Pará	1.090	0.059	0.046	0.039	1.11	0.01	-0.168
Mato Grosso	1.090	0.091	0.062	0.061	1.23	0.04	-0.023
Góias	0.957	0.041	0.031	0.028	1.26	0.01	-0.068
Distrito Federal	0.870	0.053	0.033	0.035	1.49	0.01	0.015
Minas Gerais	1.190	0.089	0.043	0.058	1.09	0.01	0.198
Rio de Janeiro	1.024	0.037	0.024	0.024	1.14	0	-0.006
São Paulo	1.200	0.091	0.044	0.059	1.10	0.01	0.179
Maranhão	1.000	0.042	0.039	0.029	1.17	0.02	-0.350

Mean of different alleles (Na), Shannon's Index (I), Observed (H_O) and Expected (H_E) Heterozygosity, allelic richness (Ar), private alleles richness (PAr) and Fixation index (f).

https://doi.org/10.1371/journal.pone.0241025.t003

Genomic diversity of *A. aculeata*

Concerning the genomic diversity within *A. aculeata*, the greatest diversity was found in Brazil (H_E = 0.081) and the lowest diversity in Mexico (H_E = 0.005). Likewise, the allelic richness values were similar for all populations in the 'North' samples, varying from 1.09 to 1.11. However, the greatest allelic richness for the species was registered in Brazil (Ar = 1.44) (Table 3).

Considering the two genetic groups identified in *A. aculeata*, the Southern group had higher estimates of population size (*Ne* = 3.2) than the Northern group (*Ne* = 0.6) (S3 Table).

Due to the vast territory and the greater number of *A. aculeata* samples from Brazil, genetic diversity analyses were conducted for groups of samples according to Brazilian states. Greater diversity (H_E) was found in the states of Mato Grosso [26], São Paulo (SP), and Minas Gerais (MG), with values of 0.061, 0.059, and 0.058, respectively. In terms of allelic richness (Ar), the most accentuated values were located in the central-west region of the country, in Distrito Federal (DF), Goiás (GO), and Mato Grosso [26], with values of 1.49, 1.26, and 1.23, respectively.

Discussion

To our knowledge, this is the first study using GBS for identifying genome-wide SNPs and their application for inferring the genetic diversity and population structure in Acrocomia species and within *A. aculeata*. Sampling was broad in terms of the occurrence of Acrocomia species and comprehensively captured the genomic diversity and structure of the species.

A. aculeata

At the genus level, the distinction of *A. aculeata* as an independent genetic group or taxon was supported through the results obtained with the Bayesian analyses (Fig 2), and by the PCoA and the NJ tree (Fig 3a and 3b). A notable finding was the identification of a clear substructure within *A. aculeata*, showing two genetic groups, corresponding to a north–south split in which the samples from Brazil (Southern group, blue cluster in Fig 3) were separated from those of Central and North America (Northern group, red Cluster in Fig 3). This result was evident in

the Bayesian analysis performed at the genus level (Fig 2) as well as with only samples of *A*. *aculeata* (Fig 4a). The substructure identified in *A*. *aculeata* has not been previously reported and can be attributed to the greater number of samples included in this study, which covered a wide geographic occurrence of the species in the American continent. The presence of two genetic groups may be the result of reproductive isolation due to the Amazon Rainforest acting as a geographical barrier (S3 Fig) that prevented gene flow between them and with an independent evolution. Another hypothesis is that these two gene pools support the existence of more than one species, as reported in a previous taxonomic classification in Central and North America Countries [73].

Another interesting result observed was that individuals from the population of Maranhão presented as an admixture between the Northern and Southern groups of *A*. *aculeata* (Fig 4). The origin of the genus Acrocomia is uncertain. However, in the case of *A*. *aculeata*, based on the dates of archeological records of human use, the most accepted hypothesis suggests that the species originated in northern Brazil (in the region of Santarém, State of Pará) approximately 11,200 MY, and was later dispersed by humans to Central America [74]. According to our results, the admixture observed in the populations of Maranhão (neighboring to Pará State) (Fig 4a) may support this hypothesis, suggesting a common geographical origin of the to two genetic groups in the northeast region of Brazil. In agreement with the *A*. *aculeata* dispersion routes from South to Central and North America [74], the lower values of genetic diversity for the species found in the Northern group may have resulted from a founder effect, since all population of this cluster presented lower values of genetic diversity than those observed in the populations of the southern cluster (Brazil) (Table 3).

Bayesian analysis identified individuals of *A*. *aculeata* with a degree of genetic admixture with *A*. *totai* (Cluster 2, in Fig 2) and *A*. *intumescens* (Cluster 6, in Fig 2), suggesting gene flow between species. As *A*. *aculeata* is dispersed mainly by cattle [75,76], the agricultural expansion and livestock may have favored the dispersion of the species to areas where *A*. *totai* and *A*. *intumescens* occur, creating opportunities for hybridization due to secondary contact. There have been no reports of interspecific hybridization in the Acrocomia genus. However, a recent study using microsatellite markers also detected connectivity between populations of *A*. *aculeata* and *A*. *totai* in Brazil [12].

A. *aculeata* displays the greatest geographical distribution of the genus [3,4,22,23]. As expected for a species with a wide distribution that has adapted to diverse environmental conditions, the genetic diversity of *A*. *aculeata* was higher when compared to other species (Table 3). At the intraspecific level, the highest genetic diversity for the species was found in Brazil, especially in the States of Minas Gerais and São Paulo (Table 3). Although it is not possible to make direct comparisons due to the different types of molecular markers used, previous studies also identified a high genetic diversity for *A*. *aculeata* in the States of Minas Gerais and São Paulo [20,36,40].

An unexpected result was the low genetic diversity of *A*. *aculeata* in Mexico, where the species is also distributed in an extensive geographical area, from the north to the south of the country (Fig 1). These results could reflect the use and exploration of the species in that country and other Central American countries, where adult plants are harvested as the raw material for a fermented drink called "taverna" [77,78]. This kind of exploration is one of the main factors driving the reduction size or elimination of the natural populations, which affects the reproductive capacity of the species and its natural regeneration [78] and might also been reducing the genetic diversity and the effective population size.

A. *aculeata* is strongly associated with humans [74,75]. Even though it is considered an incipiently domesticated species, it has a wide range of uses in different countries of the Americas [13,14,79]. Therefore, patterns of genetic diversity and structure can also be the result of

different states of domestication, with different intensities of selection in each region, as also reported for other species, such as beans [80], tomato [81], and cacao [82].

A. totai

A. totai is the second most geographically dispersed species in the genus. It has been documented in eastern Bolivia, Paraguay, Central-west Brazil to northern Argentina [4,23]. The taxonomic distinction of the species has been demonstrated based on morphological data and geographic distribution [4], leaf anatomy [1], and fruit biometry [83]. However, *A. totai* is commonly regarded as *A. aculeata* due to the pronounced morphological similarity of both species, and because both have fruits with similar biometric and color characteristics [83]. Our results were congruent with the current taxonomic classification of the species. Almost all samples initially considered as *A. totai* (94%) belonged to cluster 2 with a high ancestry coefficient (> 0.75), according to Structure analysis (Fig 2), and corroborated with PCoA and NJ analysis (Fig 3a and 3b). Our results agreed with those of Lima et al. [12], that documented the clear genetic differentiation between *A. aculeata* and *A. totai* (treated as ecotypes) using microsatellite markers. Although not treated as distinct species, but considering the geographical distribution of both, several studies using molecular and morphological markers also reinforced the classification of *A. totai* as a distinct taxon. Lanes et al. [36] used microsatellite markers to demonstrate the marked genetic differences of *A. aculeata* between individuals from the Pantanal region, State of Mato Grosso do Sul, Brazil, and other regions of the country. Similarly, Silva et al. [39] (27) analyzed the variation in the internal transcribed spacer (ITS) region and identified four haplotypes. Two were shared by genotypes from São Paulo and Minas Gerais, and one was exclusive to genotypes collected in Mato Grosso do Sul. The morphological characteristics of *A. aculeata* include larger fruits (3.5 and 5.0 cm) and a pulp oil content that can reach approximately 78% (27, 68–70) while the fruits of *A. totai* are smaller (2.5 and 3.5 cm) with a pulp oil content between 26% and 33% [83–85] (68, 71, 72).

In Brazil, *A. totai* is considered to be restricted to the State of Mato Grosso do Sul [4,86,87]. An interesting finding of our study was that samples from Xambrê, Paraná (XAM) and a sample from Palmas, Tocantins (PAL_182), considered as *A. aculeata* based on Lorenzi et al., [4] taxonomic classification, were attributed to cluster 2 of *A. totai* by the Bayesian analysis (Fig 2), by PCoA, and by NJ (Fig 3a and 3b). Although the occurrence of *A. totai* in these states has not been proven, our results are consistent with the information reported on the Flora do Brazil 2020 website [23], indicating the possible occurrence of *A. totai* in these states.

Although the genetic structure and separation of *A. aculeata* from *A. totai* was evident based on the cluster analyses, the genetic differentiation (F_{ST}) between species was 0.083, which was the lowest value (Table 1). These result was consistent with the value obtained using microsatellite markers ($F_{CT} = 0.07$) by Lima et al., [12]. The findings may reflect the retention of ancestral polymorphisms, the hybridization or gene flow between species in convergent areas [12] or could be evidence of an ongoing speciation process [36].

Based on the H_E and Ar values, *A. totai* was the species with the highest level of genetic diversity (Table 2). Our results are comparable to those found in a recent study using microsatellite markers [12], in which the genetic diversity of *A. totai* was greater than that of *A. aculeata*. Similarly, previous studies also identified greater genetic diversity in populations from Mato Grosso do Sul than population from other location of Brazil, although the authors did not consider the populations to be *A. totai* [36,37]. The high diversity observed in *A. totai* could reflect its geographically widespread occurrence and expansion of genetic diversity promoted by the interspecific hybridization with *A. aculeata*.

The results of cluster analysis and genetic differentiation corroborated the classification of *A*. *totai* as an independent taxon based on morphological [4], anatomical [1], and molecular markers [12]. This taxonomic separation seems to be more appropriate than that proposed for Henderson et al. [3], which considered all tree-sized Acrocomias as a single taxonomic group called *A*. *aculeata*.

A. intumescens

Contrary to the actual taxonomic classification [4–6], our analyses did not show a clear genetic separation of *A*. *intumescens* (Figs 2, 3a and 3b). All the samples of *A*. *intumescens* were assigned to cluster 6, however presented high levels of admixture with *A*. *aculeata* (cluster 5, Fig 2). *A*. *intumescens* also showed a moderate genetic differentiation with *A*. *aculeata* (F_{ST} = 0.128, Table 1), reinforcing the close genetic relationship among both species as described by Vianna et al. [1] based on leaf anatomy. Morphologically, *A*. *intumescens* is distinguished mainly by the swelling of the stipe [4]. However, botanical characters suggested to delimit Acrocomia species have revealed an overlapping in size of fruits [83] and for oil content in the mesocarp, ranging from 37 to 78% in *A*. *aculeata* [85,88] and from 34 to 41% in *A*. *intumescens* [85,89].

A phylogenetic study by Meerow et al. [90], estimated the divergence of *A*. *intumescens* and *A*. *aculeata* 5 MA ago. The genetic structure we observed may reflect the maintenance of ancestral polymorphism, possibly as a result of the recent divergence of these species with insufficient time for the appearance of reproductive isolation mechanisms, allowing the interspecific hybridization. *A*. *intumescens* is endemic to northeast Brazil and has a restricted distribution [4,23]. Species with a restricted geographical distribution tend to have lower genetic diversity than species with a wide geographical distribution [91,92]. Consistent with this trend, *A*. *intumescens* showed lower values of heterozygosity and allelic richness than the wide geographical distribution species *(A*. *aculeata* and *A*. *totai*) (Table 2). However, the genetic diversity found in *A*. *intumesces* was comparable to that observed in other plant species associated with restricted geographic distribution [93–95].

A. crispa

A. *crispa* is an insular species with a distribution restricted to Cuba. A clear separation and a strong genetic divergence compared to the other species, as evidenced in the cluster analysis (Figs 2, 3a and 3b) and by the high values of F_{ST} (Table 1). These expectations were understood if considered that the gene flow through pollen or seed dispersal between island populations and continental populations is limited such that a strong genetic structure and a high degree of differentiation between them is expected, as reported for several species [96,97]. Our results are congruent with those reported for other tree species, which also showed high levels of genetic differentiation between island populations compared to continental populations and lower levels of genetic diversity on the islands than on the continent [60,98–100]. *A*. *crispa* displayed low values of genetic diversity (H_E = 0.020) compared with other Acrocomia species, although these values are expected for endemic island species. However, interestingly, *A*. *crispa* presented the greatest allele richness (2.29) and allele richness of private alleles (0.17) (Table 2). Based on chlorosplastic and nuclear genes, the time of divergence estimated for *A*. *crispa* as 16 Mya, while *A*. *aculeata* and *A*. *intumescens* diverged 5 Mya [90]. This more ancient divergence associated with geographic isolation may support the allelic richness and the greater number of private alleles found in *A*. *crispa*, as well as the strong genetic differentiation of from other Acrocomia species. This hypothesis has also been posited for other endemic species of islands that have congeners on the continent [101,102].

There is no detailed information about the morphological characteristics of *A*. *crispa*. However, some morphological differences have been described, such as the presence of swelling in the median region of the stipe as the most discriminating botanical characteristic [73], the smaller fruits, which varies from 1 to 3 cm [3], than that described in *A*. *totai* (2.5 to 3.5 cm), *A*. *intumescens* (3.0 to 4.0 cm) and *A*. *aculeata* (3.5 to 5.0 cm) [83] and also differences in pollen morphology with trichotomocolpated pollen in *A*. *aculeata* and monocolpous pollen in *A*. *crispa* (named *Gastrococos crispa* by the authors) [103].

A. *crispa*, previously designated to the genus Gastrococos by Moore [104], was recently allocated to the genus Acrocomia, mainly due to the sequencing of the nuclear *prk* gene [105]. Although most phylogenetic studies support the relationship between *A*. *aculeata* and *A*. *crispa* as sisters in a single monophyletic group [105–110], other phylogenetic [90] and cladistic studies [111] shown that they are sister species in paraphyletic groups. However, these phylogenetic studies were conducted at higher taxonomic levels (families, subfamilies, and tribes), with the inclusion of few species of Acrocomia. Therefore, they have limited ability to accurately reveal phylogenetic relationships of Acrocomia species.

The morphological characteristics of the species, the divergence time and our results of genetic differentiation, diversity, and structure may collectively support an independent taxonomic status of *A*. *crispa* within the genus Acrocomia. Therefore, we suggest a revision of the taxonomy for the species and considering the low sample size, our results about genetic diversity in the species should be taken with caution and additional studies including more populations and more samples per population are needed.

A. media

In contrast to the evidence of genetic divergence for *A*. *crispa*, the recognition of *A*. *media* as an independent taxonomic unit was not supported by our study. As *A*. *media* is also an island species, it would be expected to have a strong genetic structure when compared to other Acrocomia species with a continental distribution. Contrary to this assumption, all samples considered as *A*. *media* were assigned to the northern group of *A*. *aculeata*, as evidenced by three cluster analyses (Figs 2, 3a and 3b). In addition, the F_{ST} values (Table 1) also indicated low genetic differentiation of *A*. *media* compared to *A*. *aculeata*.

The estimates of genetic diversity observed in *A*. *media* were the lowest compared to other species (H_E = 0.005 and Ar = 1.08), but were consistent with several studies of population genetics in plants, which predicted that island populations have reduced levels of genetic diversity compared to continental populations [96,112]. The low genetic diversity observed in *A*. *media* can be attributed to the founder effect associated with the establishment of populations with only a few individuals [112,113] or to genetic drift due to stochastic events inherent in the islands and/or fragmentation during its formation [114]. Nevertheless, studies with greater representativeness of the species are necessary to obtain more accurate parameters of genetic diversity.

A. *media* was first described in Puerto Rico by Cook [115]. The author adopted the shortest trunk and the smallest diameter of the stipe as the differentiating characteristics of *A*. *media* from *A*. *aculeata*. However, *A*. *media* was considered synonymous with *A*. *aculeata* for a long time due to the absence of consistent botanical characteristics for differentiation. In 2013, The Plant List recognized *A*. *media* as a distinct species based on the floristic palm inventory of Proctor [116]. However, the same author mentioned that the existing information about *A*. *media* was very old and based on few individuals, suggesting an increase in the number of evaluated individuals to guarantee a more consistent morphological description of the species. The

only phylogenetic study performed with *A*. *media* included an individual from Puerto Rico, and a sample of *A*. *aculeata* from Brazil revealed that both species were closely related [105].

Based on the lack of genetic differentiation of *A*. *media*, low genetic diversity in the species, and low pairwise F_{ST} value between *A*. *media* and *A*. *aculeata*, we hypothesize that *A*. *media* is synonymous with *A*. *aculeata*. Thus, a recent introduction in Puerto Rico was not sufficient to characterize the reproductive isolation needed for the differentiation of *A*. *aculeata*.

A. hassleri and *A. glauscescens*

The genomic data of our study did not allow the assignment of distinct taxonomic units to the species *A*. *hassleri* and *A*. *glauscescens*. Based on morphological characters, the species are clearly differentiated from the others by their small size. However, based on the results obtained from the cluster analysis, they were assigned to cluster 2, being closely related to *A*. *totai* (Figs 2, 3a and 3b). However, this result should be considered with caution, as we only used one sample of each species in the analyses, which could limit the comparison of genetic estimates and decrease the probability of detecting genetic structure, as evidenced in similar studies with a low number of samples [117,118]. Further studies with a greater number of accessions are needed to increase the species representation, and to establish reliable genetic relationships between *A*. *hassleri* and *A*. *glauscescens* and other Acrocomia species.

Conclusions

Our study is the first to offer evidence of the efficiency of NGS through the application of the GBS protocol in Acrocomia. The data may constitute a reference for the application of this protocol in the genus. Even without a reference genome, we successfully identified a large number of SNPs for several species, revealing potentially valuable markers for future studies in the genus Acrocomia. The SNPs yielded unprecedented results of the genetic relationships between Acrocomia species as well as at the population level for *A*. *aculeata*. In general, our results were partially congruent with the taxonomy of the genus, supporting the current separation of some species. The genomic structure revealed the formation of well-defined genetic groups and confirmed the distinction of *A*. *aculeata*, *A*. *totai*, *A*. *intumescens*, and *A*. *crispa*, with the latter showing a strong genetic differentiation as well as the absence of genetic distinction of *A*. *media*. We recommend a review of the current taxonomic classification of *A*. *crispa* and *A*. *media*. In addition, SNPs also allowed the identification of gene flow patterns and/or hybridization between species.

In the case of *A*. *aculeata*, the data provide an overview of the genomic diversity and structure from sampling over a wide area of occurrence. The genomic data showed the existence of two large gene pools in the species at the continental level (north and south), with greater genomic diversity in the latter populations. The results from this study will serve as a reference for current and future studies on genetic diversity, taxonomy, evolution, ecology, and phylogeny of the genus Acrocomia, and will support genetic breeding, conservation, and management activities for *A*. *aculeata*.

Supporting information

S1 Fig. Delta (Δ) K values for different numbers of populations assumed (K) in the STRUCTURE analysis, estimated based on Evanno method for all Acrocomia species.
(PNG)

S2 Fig. Delta (Δ) K values for different numbers of populations assumed (K) in the STRUCTURE analysis, estimated based on Evanno method for *A. aculeata*.
(PNG)

S3 Fig. Geographic barriers to gene flow identified using Monmnier's algorithm. The blue lines indicate the position of the barriers.
(TIF)

S1 Table. Geographical location and origin of the Acrocomia species samples.
(XLSX)

S2 Table. Outlier SNP loci (putatively under selection) for Acrocomia species and within *A. aculeata* identified by PCAdapt, Fsthet and LEA packages.
(XLSX)

S3 Table. Estimated effective population size (*Ne*) for the cluster identifies by Bayesian analysis.
(XLSX)

S1 File. SNP genotype information in variant calling format (vcf) for 172 samples of Acrocomia species.
(VCF)

S2 File. Additional information of methods used for outlier SNP identification.
(DOCX)

Acknowledgments

The authors would like to thank Víctor Díaz Fuentes from the Instituto Nacional of Investigaciones Agrícolas y Pecuarias (INIFAP), Mexico and Anna Neubauer from the company Green Integrated Energies from Costa Rica, for their support during sample collection and Diego Wassner from the Facultad de Agronomía of Universidad de Buenos Aires for kindly providing *A. totai* samples from natural population of Argentina.

Author Contributions

Conceptualization: Brenda Gabriela Díaz, Maria Imaculada Zucchi, Carlos Augusto Colombo.

Data curation: Alessandro Alves-Pereira.

Formal analysis: Brenda Gabriela Díaz, Alessandro Alves-Pereira, Caléo Panhoca de Almeida.

Funding acquisition: Carlos Augusto Colombo.

Investigation: Brenda Gabriela Díaz, Alessandro Alves-Pereira, Aline Costa Lima Moraes.

Methodology: Brenda Gabriela Díaz, Alessandro Alves-Pereira, Aline Costa Lima Moraes.

Project administration: Maria Imaculada Zucchi, Carlos Augusto Colombo.

Resources: Brenda Gabriela Díaz, Suelen Alves Vianna, Joaquim Azevedo-Filho, Carlos Augusto Colombo.

Supervision: Maria Imaculada Zucchi, Carlos Augusto Colombo.

Writing – original draft: Brenda Gabriela Díaz, Maria Imaculada Zucchi.

Writing – review & editing: Brenda Gabriela Díaz, Maria Imaculada Zucchi, Carlos Augusto Colombo.

References

1. Vianna SA, Carmelo-Guerreiro SM, Noblick LR, Colombo CA. Leaf anatomy of Acrocomia (Arecaceae): an additional contribution to the taxonomic resolution of a genus with great economic potential. Plant Syst Evol. 2017; 303:233–48.
2. Martius CFPV. Historia Naturalis Palmarum. Leipzig, TC Weigel. 1824:285–6.
3. Henderson A, Galeano G, Bernal R. Field guide to the palms of the Americas. New Jersey, USA: Princeton University Press; 1995.
4. Lorenzi H, Noblick L, Kahn F. Flora Brasileira-Arecaceae (Palmeiras) Nova Odessa, SP: Instituto Plantarum. 2010.
5. The Plant list. Version 1.1. 2013.
6. Palmweb. Acrocomia Mart. Hist. Nat. Palm. n. d [http://www.palmweb.org/.
7. Stanton DW, Frandsen P, Waples RK, Heller R, Russo I-RM, Orozco-terWengel PA, et al. More grist for the mill? Species delimitation in the genomic era and its implications for conservation. Conservation Genetics. 2019; 20(1):101–13.
8. Camargo A. Species delimitation: a decade after the renaissance. The species problem-ongoing issues: IntechOpen; 2013.
9. Petit RJ, Excoffier L. Gene flow and species delimitation. Trends in Ecology & evolution. 2009; 24(7):386–93. https://doi.org/10.1016/j.tree.2009.02.011 PMID: 19409650
10. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 2003; 4(12):981–94. https://doi.org/10.1038/nrg1226 PMID: 14631358
11. Kirk H, Freeland JR. Applications and implications of neutral versus non-neutral markers in molecular ecology. International journal of molecular sciences. 2011; 12(6):3966–88. https://doi.org/10.3390/ijms12063966 PMID: 21747718
12. de Lima NE, Meerow AW, Manfrin MH. Genetic structure of two Acrocomia ecotypes (Arecaceae) across Brazilian savannas and seasonally dry forests. Tree Genet Genomes. 2020; 16(4):1–12.
13. Lorenzi GMAC. *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. Arecaceae: bases para o extrativismo sustentável. Curitiba, PR: Universidade Federal do Paraná; 2006.
14. Ramos MIL, Ramos Filho MM, Hiane PA, Braga Neto JA, Siqueira EMA. Qualidade nutricional da polpa de bocaiúva *Acrocomia aculeata* (Jacq.) Lodd. Food Sci Technol 2008; 28:90–4.
15. Colombo CA, Berton LHC, Diaz BG, Ferrari RA. Macauba: a promising tropical palm for the production of vegetable oil. OCL. 2018; 25(1):D108.
16. Coimbra MC, Jorge N. Characterization of the pulp and kernel oils from Syagrus oleracea, Syagrus romanzoffiana, and *Acrocomia aculeata*. J Food Sci. 2011; 76(8):C1156–C61. https://doi.org/10.1111/j.1750-3841.2011.02358.x PMID: 22417579
17. da Silva César A, de Azedias Almeida F, de Souza RP, Silva GC, Atabani A. The prospects of using *Acrocomia aculeata* (macaúba) a non-edible biodiesel feedstock in Brazil. Renew Sust Energ Rev. 2015; 49:1213–20.
18. Abreu AG, Priolli RHG, Azevedo-Filho JA, Nucci SM, Zucchi MI, Coelho RM, et al. The genetic structure and mating system of *Acrocomia aculeata* (Arecaceae). Genet Mol Biol. 2012; 35(1):116–21.
19. Lanes ÉC, Motoike SY, Kuki KN, Resende MD, Caixeta ET. Mating system and genetic composition of the macaw palm (*Acrocomia aculeata*): implications for breeding and genetic conservation programs. J Hered. 2016; 107(6):527–36. https://doi.org/10.1093/jhered/esw038 PMID: 27288529
20. Coelho NHP, Tambarussi EV, Aguiar BI, Roque RH, Portela RM, Braga RC, et al. 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. 2018; 19(4):879–91.
21. Abreu IS, Carvalho CR, Carvalho GMA, Motoike SY. First karyotype, DNA C-value and AT/GC base composition of macaw palm (*Acrocomia aculeata*, Arecaceae)–a promising plant for biodiesel production. Aust J Bot. 2011; 59(2):149–55.
22. Dransfield J, Uhl NW, Lange CBA, Baker WJ, Harley MM, Lewis CE. Genera Palmarum: the evolution and classification of palms: Kew Publishing; 2008.

23. Vianna SA, Campos_Rocha A. Acrocomia in Flora do Brasil 2020 em construção: Jardim Botânico do Rio de Janeiro; 2020 [<http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB15662].

24. Vianna S, Colombo C, editors. Distribuição geográfica de Acrocomia aculeata (jacq,) lodd ex mart. (arecaceae) em sua região de ocorrência. I Congresso brasileiro de macaúba; 2013.

25. Costa CJ, Marchi ECS. Germinação de sementes de palmeiras com potencial para produção de agroenergia. Embrapa Cerrados-Documentos (INFOTECA-E). 2008.

26. Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor M, Ogden R, Limborg M, et al. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular ecology resources. 2011; 11:123–36. https://doi.org/10.1111/j.1755-0998.2010.02943.x PMID: 21429169

27. Fay MF, Gargiulo R, Viruel J. The present and future for population genetics, species boundaries, biogeography and conservation. Oxford University Press UK; 2019.

28. Brumfield RT, Beerli P, Nickerson DA, Edwards SV. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology & Evolution. 2003; 18(5):249–56.

29. Vinson C, Mangaravite E, Sebbenn A, Lander T. Using molecular markers to investigate genetic diversity, mating system and gene flow of Neotropical trees. Braz J Bot. 2018; 41(2):481–96.

30. Boutet G, Carvalho SA, Falque M, Peterlongo P, Lhuillier E, Bouchez O, et al. SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. BMC genom. 2016; 17(1):1–14. https://doi.org/10.1186/s12864-016-2447-2 PMID: 26892170

31. Davey JW, Blaxter ML. RADSeq: next-generation population genetics. Brief Funct Genomics. 2010; 9(5–6):416–23. https://doi.org/10.1093/bfgp/elq031 PMID: 21266344

32. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS genet. 2010; 6(2): e1000862. https://doi.org/10.1371/journal.pgen.1000862 PMID: 20195501

33. Morgil H, Gercek YC, Tulum I. Single nucleotide polymorphisms (SNPs) in plant genetics and breeding. The Recent Topics in Genetic Polymorphisms: IntechOpen; 2020.

34. Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S. SNP markers and their impact on plant breeding. International journal of plant genomics. 2012; 2012. https://doi.org/10.1155/2012/728398 PMID: 23316221

35. Morin PA, Luikart G, Wayne RK. SNPs in ecology, evolution and conservation. Trends in ecology & evolution. 2004; 19(4):208–16.

36. Lanes ÉC, Motoike SY, Kuki KN, Nick C, Freitas RD. Molecular characterization and population structure of the macaw palm, *Acrocomia aculeata* (Arecaceae), ex situ germplasm collection using microsatellites markers. J Hered. 2015; 106(1):102–12. https://doi.org/10.1093/jhered/esu073 PMID: 25425677

37. Mengistu FG, Motoike SY, Cruz CD. Molecular characterization and genetic diversity of the macaw palm ex situ germplasm collection revealed by microsatellite markers. Diversity. 2016; 8(4):20.

38. Nucci SM. Desenvolvimento, caracterização e análise da utilidade de marcadores microssatélites em genética de população de macaúba.: Instituto Agronômico de Campinas; 2007.

39. Silva LCC, Lemos RdC, Carvalho CGI, Good-God PIV, Oliveira LOd, Costa MD-BL, et al. Genetic diversity and structure of macaw palm: implications for genetic variability sampling. Rev Arvore. 2017; 41(5).

40. Oliveira D, Melo Júnior A, Brandão M, Rodrigues L, Menezes E, Ferreira P. Genetic diversity in populations of *Acrocomia aculeata* (Arecaceae) in the northern region of Minas Gerais, Brazil. Genet Mol Res. 2012; 11(1):531–8. https://doi.org/10.4238/2012.March.8.1 PMID: 22535388

41. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nature Rev Genet. 2010; 11(10):697–709. https://doi.org/10.1038/nrg2844 PMID: 20847747

42. Velazquez MAM. Caracterización fitoquímica, molecular y evaluación de respuestas del cultivo in vitro de coyol (Acrocomia mexicana Kraw ex Marti)/por Antonio Magdiel Velazquez Mendez: Universidad Autónoma Chapingo; 2013.

43. Doyle JJ, Doyle JL. Isolation ofplant DNA from fresh tissue. Focus. 1990; 12(13):39–40.

44. Poland JA, Brown PJ, Sorrells ME, Jannink J-L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS one. 2012; 7(2): e32253. https://doi.org/10.1371/journal.pone.0032253 PMID: 22389690

45. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes genom genet. 2011; 1(3):171–82. https://doi.org/10.1534/g3.111.000240 PMID: 22384329

46. Luu K, Bazin E, Blum MG. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Ecol Resour. 2017; 17(1):67–77. https://doi.org/10.1111/1755-0998.12592 PMID: 27601374

47. Core TR. R: A language and environment for statistical computing. Vienna, Austria; 2018.

48. Flanagan SP, Jones AG. Constraints on the F ST–heterozygosity outlier approach. J Hered. 2017; 108(5):561–73. https://doi.org/10.1093/jhered/esx048 PMID: 28486592

49. Wright S. Isolation by distance. Genetics. 1943; 28(2):114. PMID: 17247074

50. Wright S. Evolution in Mendelian populations. Genetics. 1931; 16(2):97. PMID: 17246615

51. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol 2013; 30(7):1687–99. https://doi.org/10.1093/molbev/mst063 PMID: 23543094

52. Frichot E, François O. LEA: An R package for landscape and ecological association studies. Methods Ecol Evol. 2015; 6(8):925–9.

53. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol. 2005; 25:965–1978.

54. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155(2):945–59. PMID: 10835412

55. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 2005; 14(8):2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x PMID: 15969739

56. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012; 4(2):359–61.

57. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007; 23 (14):1801–6. https://doi.org/10.1093/bioinformatics/btm233 PMID: 17485429

58. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources. 2015; 15(5):1179–91. https://doi.org/10.1111/1755-0998.12387 PMID: 25684545

59. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 1978; 89(3):583–90. PMID: 17248844

60. Chen D, Zhang X, Kang H, Sun X, Yin S, Du H, et al. Phylogeography of Quercus variabilis based on chloroplast DNA sequence in East Asia: multiple glacial refugia and mainland-migrated island populations. PLoS One. 2012; 7(10):e47268. https://doi.org/10.1371/journal.pone.0047268 PMID: 23115642

61. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in genetics. 2015; 6:208. https://doi.org/10.3389/fgene.2015.00208 PMID: 26113860

62. Dray S, Dufour AB, Chessel D. The ade4 package-II: Two-table and K-table methods. R news. 2007; 7(2):47–52.

63. Villanueva RAM, Chen ZJ. ggplot2: Elegant graphics for data analysis. Taylor & Francis; 2019.

64. Monmonier MS. Maximum-difference barriers: An alternative numerical regionalization method. Geographical analysis. 1973; 5(3):245–61.

65. Jombart T, Ahmed I. adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011; 27(21):3070–1. https://doi.org/10.1093/bioinformatics/btr521 PMID: 21926124

66. Gabriel KR, Sokal RR. A new statistical approach to geographic variation analysis. Systematic zoology. 1969; 18(3):259–78.

67. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in ecology and evolution. 2013; 4(8):782–8.

68. Paquette SR, Paquette MSR. Package 'PopGenKit'. 2011.

69. Kalinowski ST. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular ecology notes. 2005; 5(1):187–9.

70. Do C, Waples RS, Peel D, Macbeth G, Tillett BJ, Ovenden JR. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular ecology resources. 2014; 14(1):209–14. https://doi.org/10.1111/1755-0998.12157 PMID: 23992227

71. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012; 28(2):298–9. https://doi.org/10.1093/bioinformatics/btr642 PMID: 22110245

72. Waples RS, Do C. LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular ecology resources. 2008; 8(4):753–6. https://doi.org/10.1111/j.1755-0998.2007.02061.x PMID: 21585883

73. Bailey LH. Acrocomia preliminary paper. Gentes Herbarum–Occasional Papers on the Kinds Plants; 1941.

74. Morcote-Rios G, Bernal R. Remains of palms (Palmae) at archaeological sites in the New World: a review. Bot Rev. 2001; 67(3):309–50.

75. Lentz DL. *Acrocomia mexicana*: Palm of the ancient Mesoamericans. J Ethnobiol. 1990; 10(2):183–94.

76. Scariot A. Seed dispersal and predation of the palm *Acrocomia aculeata*. Principes. 1998; 42:5–8.

77. Balick MJ. Production of coyol wine from *Acrocomia mexicana* (Arecaceae) in Honduras. Econ Bot. 1990; 44(1):84–93.

78. H. D-FV, Amín R-CP, Beau M-VB, Moises A-B. Caracterización morfológica del coyol Acrocomia aculeata (Jacq.), y su potencial productivo de aceite vegetal. In: Instituto Nacional de Investigaciones Forestales AyPCdIRPSCERI, editor. Folleto Técnico Tuxtla Chico, Chiapas, México2020. p. 54.

79. Ramírez Hernández BC, Zañudo Hernández J, García de Alba Verduzco JE, Délano Frier JP, Pimienta Barrios E, García Martínez MÁ. Importancia agroecológica del coyul (*Acrocomia mexicana* Karw. ex Mart.). Estudios sociales (Hermosillo, Son). 2013; 21(41):95–113.

80. Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, et al. Linkage disequilibrium and population structure in wild and domesticated populations of Phaseolus vulgaris L. Evol Appl. 2009; 2(4):504–22. https://doi.org/10.1111/j.1752-4571.2009.00082.x PMID: 25567895

81. Blanca J, Cañizares J, Cordero L, Pascual L, Diez MJ, Nuez F. Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. PloS one. 2012; 7(10):e48198. https://doi.org/10.1371/journal.pone.0048198 PMID: 23118951

82. Cornejo OE, Yee M-C, Dominguez V, Andrews M, Sockell A, Strandberg E, et al. Population genomic analyses of the chocolate tree, Theobroma cacao L., provide insights into its domestication process. Communications biology. 2018; 1(1):1–12. https://doi.org/10.1038/s42003-018-0168-6 PMID: 30345393

83. Vianna SA, Berton LHC, Pott A, Guerreiro SMC, Colombo CA. Biometric characterization of fruits and morphoanatomy of the mesocarp of Acrocomia species (Arecaceae). Int J Biol. 2017; 9(3):78–92.

84. Berton L. Avaliação de populações naturais, estimativas de parâmetros genéticos e seleção de genótipos elite de macaúba (Acrocomia aculeata). Campinas, SP Brazil: Instituto Agronômico de Campinas-IAC. 2013.

85. Conceição L, Junqueira NTV, Licurgo FdS, Antoniassi R, Wilhelm AE, BRAGA MF, editors. Teor de óleo em frutos de diferentes espécies de macaubeira (Acrocomia spp.). Embrapa Cerrados-Artigo em anais de congresso (ALICE); 2012: In: CONGRESSO BRASILEIRO DE FRUTICULTURA, 22., 2012, Bento Gonçalves. Anais.

86. Markley KS. Mbocayá or Paraguay Cocopalm—an important source of oil. Econ Bot. 1956; 10(1):3–32.

87. Rodríguez MF, Aschero CA. *Acrocomia chunta* (Arecaceae) raw material for cord making in the Argentinean Puna. J Archaeol Sci. 2005; 32(10):1534–42.

88. Berton L. Avaliação de populações naturais, estimativas de parâmetros genéticos e seleção de genótipos elite de macaúba (*Acrocomia aculeata*). Campinas, SP. Brazil: Instituto Agronômico de Campinas—IAC; 2013.

89. Bora PS, Rocha R. Macaiba palm: fatty and amino acids composition of fruits macaíba: composición de aminoácidos y ácidos grasos de frutos macaíba: composición de aminoácidos e ácidos graxos de froitos. CYTA-J Food. 2004; 4(3):158–62.

90. Meerow AW, Noblick L, Salas-Leiva DE, Sanchez V, Francisco-Ortega J, Jestrow B, et al. Phylogeny and historical biogeography of the cocosoid palms (A recaceae, A recoideae, C ocoseae) inferred from sequences of six WRKY gene family loci. Cladistics. 2015; 31(5):509–34.

91. Cole CT. Genetic variation in rare and common plants. Annu Rev Ecol Evol Syst. 2003; 34(1):213–37.

92. Hamrick J, Godt M. Allozyme diversity in plant species. In 'Plant population genetics, breeding and genetic resources'.(Eds Brown AHD, Clegg MT, Kahler AL, Weir BS) pp. 43–63. Sinauer Associates: Sunderland, MA; 1989.

93. Ellis J, Pashley C, BURKE J, McCauley D. High genetic diversity in a rare and endangered sunflower as compared to a common congener. Mol Ecol 2006; 15(9):2345–55. https://doi.org/10.1111/j.1365-294X.2006.02937.x PMID: 16842410

94. Furches MS, Small RL, Furches A. Genetic diversity in three endangered pitcher plant species (Sarracenia; Sarraceniaceae) is lower than widespread congeners. Am J Bot. 2013; 100(10):2092–101. https://doi.org/10.3732/ajb.1300037 PMID: 24088341

95. Levy E, Byrne M, Coates D, Macdonald B, McArthur S, Van Leeuwen S. Contrasting influences of geographic range and distribution of populations on patterns of genetic diversity in two sympatric Pilbara Acacias. PLoS One. 2016; 11(10):e0163995. https://doi.org/10.1371/journal.pone.0163995 PMID: 27768703

96. Franks SJ. Genetics, evolution, and conservation of island plants. J Plant Biol. 2010; 53(1):1–9.

97. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F ST. Nat Rev Genet. 2009; 10(9):639–50. https://doi.org/10.1038/nrg2611 PMID: 19687804

98. Qi X-S, Yuan N, Comes HP, Sakaguchi S, Qiu Y-X. A strong 'filter'effect of the East China Sea land bridge for East Asia's temperate plant species: inferences from molecular phylogeography and ecological niche modelling of Platycrater arguta (Hydrangeaceae). BMC Evol Biol 2014; 14(1):41. https://doi.org/10.1186/1471-2148-14-41 PMID: 24593236

99. Qiu YX, Sun Y, Zhang XP, Lee J, Fu CX, Comes HP. Molecular phylogeography of East Asian Kirengeshoma (Hydrangeaceae) in relation to Quaternary climate change and landbridge configurations. New Phytol. 2009; 183(2):480–95. https://doi.org/10.1111/j.1469-8137.2009.02876.x PMID: 19496955

100. Sakaguchi S, Qiu YX, Liu YH, Qi XS, Kim SH, Han J, et al. Climate oscillation during the Quaternary associated with landscape heterogeneity promoted allopatric lineage divergence of a temperate tree Kalopanax septemlobus (Araliaceae) in East Asia. Mol Ecol. 2012; 21(15):3823–38. https://doi.org/10.1111/j.1365-294X.2012.05652.x PMID: 22646502

101. Mitsui Y, Chen S-T, Zhou Z-K, Peng C-I, Deng Y-F, Setoguchi H. Phylogeny and biogeography of the genus Ainsliaea (Asteraceae) in the Sino-Japanese region based on nuclear rDNA and plastid DNA sequence data. Ann Bot. 2008; 101(1):111–24. https://doi.org/10.1093/aob/mcm267 PMID: 17981878

102. Nakamura K, Denda T, Kokubugata G, Huang C-J, Peng C-I, Yokota M. Phylogeny and biogeography of the Viola iwagawae-tashiroi species complex (Violaceae, section Plagiostigma) endemic to the Ryukyu Archipelago, Japan. Plant Syst Evol. 2015; 301(1):337–51.

103. Machado SR. Variaciones en la morfología polínica de Arecaceae en Cuba: abertura tricotomosulcada y estratificación de la exina. Revista Jard Bot Nac Univ Habana. 2003:71–9.

104. Moore HE. Gastrococos crispa (Kunth). Principes. 11.1968.

105. Gunn BF. The phylogeny of the Cocoeae (Arecaceae) with emphasis on Cocos nucifera. Ann Missouri Bot Gard. 2004:505–22.

106. Hahn WJ. A phylogenetic analysis of the Arecoid line of palms based on plastid DNA sequence data. Mol Phylogenet Evol. 2002; 23(2):189–204. https://doi.org/10.1016/S1055-7903(02)00022-2 PMID: 12069550

107. Asmussen CB, Dransfield J, Deickmann V, Barfod AS, Pintaud J-C, Baker WJ. A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. Bot J Linn Soc. 2006; 151(1):15–38.

108. Couvreur TL, Hahn W, Granville J-Jd, Pham J-L, Ludena B, Pintaud J-C. Phylogenetic relationships of the cultivated Neotropical palm Bactris gasipaes (Arecaceae) with its wild relatives inferred from chloroplast and nuclear DNA polymorphisms. Syst Bot. 2007; 32(3):519–30.

109. Eiserhardt WL, Pintaud J-C, Asmussen-Lange C, Hahn WJ, Bernal R, Balslev H, et al. Phylogeny and divergence times of Bactridinae (Arecaceae, Palmae) based on plastid and nuclear DNA sequences. Taxon. 2011; 60(2):485–98.

110. Roncal J, Kahn F, Millan B, Couvreur TL, Pintaud J-C. Cenozoic colonization and diversification patterns of tropical American palms: evidence from Astrocaryum (Arecaceae). Bot J Linn Soc. 2013; 171 (1):120–39.

111. Noblick LR, Hahn WJ, Griffith MP. Structural cladistic study of Cocoseae, subtribe Attaleinae (Arecaceae): Evaluating taxonomic limits in Attaleinae and the neotropical genus Syagrus. Brittonia. 2013; 65(2):232–61.

112. Whittaker RJ, Fernández-Palacios JM. Island biogeography: ecology, evolution, and conservation: Oxford University Press; 2007.

113. Templeton AR. The reality and importance of founder speciation in evolution. BioEssays. 2008; 30 (5):470–9. https://doi.org/10.1002/bies.20745 PMID: 18404703

114. Stuessy TF, Takayama K, López-Sepúlveda P, Crawford DJ. Interpretation of patterns of genetic variation in endemic plant species of oceanic islands. Bot J Linn Soc. 2014; 174(3):276–88. https://doi.org/10.1111/boj.12088 PMID: 26074627

115. Cook OF. A synopsis of the palms of Puerto Rico. The Botanical Club. 281901. p. 565–7.

116. Proctor GR. Arecaceae (Palmae). P Acevedo-Rodríguez, M T Strong. Monocots and Gymnosperms of Puerto Rico and the Virgin Islands.: Smithsonian Institution: Contributions of the United States National Herbarium; 2005.

117. Cavers S, Degen B, Caron H, Lemes MR, Margis R, Salgueiro F, et al. Optimal sampling strategy for estimation of spatial genetic structure in tree populations. Heredity. 2005; 95(4):281–9. https://doi.org/10.1038/sj.hdy.6800709 PMID: 16030529

118. Morin PA, Martien KK, Taylor BL. Assessing statistical power of SNPs for population structure and conservation studies. Mol Ecol Resour. 2009; 9(1):66–73. https://doi.org/10.1111/j.1755-0998.2008.02392.x PMID: 21564568