SCISPACE
formerly Typeset

Open access • Posted Content • DOI:10.1101/2020.10.08.331140

Whole-genome SNP analysis elucidates the genetic population structure and diversity of Acrocomia species — Source link

Brenda Gabriela Diaz, Maria Imaculada Zucchi, Maria Imaculada Zucchi, Alessandro Alves-Pereira ...+5 more authors

Institutions: State University of Campinas, IAC

Published on: 08 Oct 2020 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Genetic diversity, Acrocomia, Genetic structure and Taxonomic rank

Related papers:

- Genetic diversity and population structure of Croton urucurana Baill. (Euphorbiaceae) in Central Brazil by ISSR markers
- Genetic structure in peripheral Western European populations of the endangered species Cochlearia pyrenaica (Brassicaceae)
- Phylogeny and genetic structure in the genus Secale.
- Phylogenetic relationships of Mediterranean Hedysarea species assessed by AFLP markers
- Genetic diversity and population structure of Iranian Ephedra major

Share this paper:

View more about this paper here: https://typeset.io/papers/whole-genome-snp-analysis-elucidates-the-genetic-population-2y7wjpcqzq

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Whole-genome SNP analysis elucidates the genetic population structure and diversity of Acrocomia species

Brenda G. Díaz[1], Maria I. Zucchi[1,2], Alessandro. Alves-Pereira[1], Caléo P. de Almeida[2], Aline C. L. Moraes[1], Suelen A. Vianna[2], Joaquim A. Azevedo-Filho[2], Carlos A 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,

* Corresponding author

E-mail: carlos.colombo@sp.gov.br (CAC)

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 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 seven species of Acrocomia including 117 samples of *A. aculeata* covering a wide geographical area of occurrence, 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.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 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 in 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

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 [7] as well as for medicinal uses [7, 8]. Of these three species, *A. aculeata* is distinguished by its high productive capacity and oil quality [9]. The oil production of 4,000 oil L/ha/year estimated in Brazil far surpasses soybeans (400 L/ha) [9] and equals the oil palm, which is considered the oilseed with the highest oil yield per area, with an oil production volume of up to 6,000 L/ha [10, 11].

A. aculeata is an arborescent heliophile and monoecious. This species produces unisexual flowers in the same inflorescence [3, 4]. It has a mixed reproductive system, with a preference for allogamy [12]. It is a diploid species (2n = 30), with a genome size of 2.8 Gbp [13]. *A. aculeata* has a wide geographic distribution, occurring naturally from northern Mexico and the Antilles to southern Brazil [3, 4, 14, 15]. 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, 16] and has adapted to sandy soils and regions with low water availability [17]. Besides being a perennial species, it is beneficial for soil management and conservation since its useful life can exceed 50 years. Colombo et al. [9] 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. *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 [18]. More recently, next-generation sequencing (NGS) has facilitated the identification of single nucleotide polymorphisms (SNPs), which have emerged as the most extensively used genotyping markers due to their abundance and distribution in the genome. The use of SNPs has considerably expanded knowledge of the genetic diversity of genomes of various plant species [19] at low cost and without the need for reference genomes [20-22]. However, SNPs have not been used as markers 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 [12, 23-26]. Other approaches include the use of internal transcribed ribosomal 18S-26S spacer (ITS region) [27] and random amplification of polymorphic DNA (RAPD) markers [28]. However, most studies have focused on *A. aculeata* [12, 23, 25, 26, 29]. Only one study has analyzed the genetic diversity of *A. totai* (Lima et al., 2020).

Considering the wide distribution of *A. aculeata* in the Americas, all the studies carried out using molecular markers have revealed a limited panorama of species genetic diversity because they considered a very small geographic sampling, with genotypes obtained mainly from the states of São Paulo and Minas Gerais in Brazil (Abreu et al., 2012; Lanes et al., 2015; Mengistu 2016; Oliveira et al., 2012; Coelho et al., 2018). Only a single study has evaluated the genetic diversity of

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

natural populations of *A. aculeata* (termed *A. mexicana)* from another country besides Brazil, that being Mexico [30].

Faced with the difficulties of taxonomic resolution of the genus Acrocomia, our study aimed to apply a population genomic approach to elucidate the genetic diversity and genetic relationships of species through genomic polymorphism data. Recognizing the increasing importance of *A. aculeata* and the lack of genetic reference data in the different countries where it grows, we analyzed the genetic diversity and structure of natural populations of *A. aculeata*, considering its wide occurrence in the American continent.

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*. 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).

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 Lorenzi et al., [4]. Circles represent geographical location and origin of samples in this study. Image sources: *A.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

aculeata, A. totai, A. hassleri, A. glaucescens. A. emensis (B. G. Díaz); *A. intumescens*; *A. media* and *A. crispa* (S. A. Vianna.)

The total genomic DNA was extracted from leaf material using the Doyle & Doyle [31] 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 GBS (Genotyping by Sequencing) technique according to the protocol described by Poland et al. [32], with modifications. We digested 7 µl of the genomic DNA [30ng.µl-1] from each sample at 37 ° C for 12 hrs 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 Quantification kit (KAPA Biosystems). We prepared two libraries of 96 samples each, which were sequenced using

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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* [33]. We used the *process_radtags* module to demultiplex the samples and to remove the low-quality readings. As there is no reference genome for Acrocomia, we aligned the sequences and organized the loci using the *ustacks* module with the following parameters: the minimum sequencing depth (-m) ≥ 3, the maximum distance between *stacks* (-M) = 2; and the maximum distance between primary and secondary sequences (-N) = 2. Subsequently, a locus catalog was built using the *cstacks* module*,* allowing a maximum of 2 differences between stacks (-n) 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).

Identification of putatively neutral and under selection loci

We identified neutral SNPs and loci putatively under selection (outliers). To reduce the possibility of identifying false positives, we applied three approaches to identify outlier loci. For the first approach, we used the method based on Principal Component Analysis (PCA) from the *pcadapt* package [34], on the R platform [35]. The *pcadapt* method assumes that SNPs excessively related to the population structure are candidates to be under adaptive selection. In this approach, no a priori information about the number of populations was introduced. Initially, we carried out the principal component analysis (PCA) to define the structure of the data set, adopting the Mahalanobis distance from the z-scores in the first k-components of each locus to identify the most

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

187 related loci to the population structure. In the second approach, we used the *fsthet* package [36]
188 based on Wright's F_{ST} fixation index [37] to identify the loci with deviation from the expected
189 relationship between F_{ST} and heterozygosity (H_E), using the island migration model [38].

190 The third approach we adopted to test the association of environmental variables with the
191 genetic variation of SNP markers was the LFMM (Latent Factor Mixed Models) [39], using the LEA
192 package (Landscape Genomics and Ecological Association Test) [40] on R. platform [35]. We used
193 nineteen bioclimatic variables related to precipitation and temperature, in addition to the minimum,
194 average and maximum values of wind speed, vapor pressure, and solar radiation, obtained from
195 the WorldClim database [41]. We performed the analyzes with the following variables (correlation ≤
196 0.8): average annual temperature, average daytime variation, isothermality, average temperature
197 of the wettest four-month period, annual temperature variation, annual precipitation, precipitation in
198 the driest month, precipitation seasonality, radiation maximum solar radiation, minimum solar
199 radiation, and average wind speed. For the lfmm function, five replicates were performed with
200 200,000 MCMC interactions after 50,000 burn-ins. For the association tests, the genetic structure
201 presented between the individuals was considered with the SNMF analysis [39], determining the
202 most likely number of genetic groups for the different data sets, using 100,000 MCMC interactions,
203 and 10 repetitions☐ for the number of groups (K) varying between 1 to 15. The LFMM analysis
204 considered K = 8 (species) and 6 (*A. aculeata* Americas). The associations with environmental
205 variables were identified for the loci with corrected p-values, considering FDR = 0.1 of
206 environmental variable association to detect SNPs, from the LEA package [40], using the sparse
207 non-negative matrix factorization function (snmf) [42]. We carried out the identification of
208 environmental variables by the principal component analysis, adopting 19 bioclimatic variables
209 from the WorldClim database [41], and selecting the variables that showed the highest correlation.
210 For the snmf function, the most likely number of populations for the different data sets was
211 determined using 100,000 interactions, and 10 repetitions☐ for K = 1-15.

212 The identification of SNPs hypothetically under selection (outliers) was performed for the
213 following groups: 1) In the genus Acrocomia, considering the species as groups, and 2) within *A.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

aculeata, considering as groups the samples' countries of origin. We considered as loci putatively under selection those shared between the three identification methods (*fsthe*t, *pcadapt* and LFMM) (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 [43], considering only neutral SNPs (3227). We also used the same software to access 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 [44] implemented in the program Structure Harvester [45]. The participation coefficient for each access was given by the alignment of five repetitions of the best K through the CLUMPP method [46] by the software CLUMPAK [47].
To visualize the genetic relationships among Acrocomia species and within the *A. acuelata*, we obtained the Nei genetic distance [48] between the individuals of each data set, and the Neighbor-Joining (NJ) hierarchical classification method with 20000 bootstrap repetitions, using the poppr package[49] on R [35].

In addition, the Principal Component Analysis (PCoA) was also carried out through the ADE 4 package [50]to explore the genetic structure of the different groups using only neutral SNPs, and was visualized graphically by the ggplot2 package [51].

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*

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and *A. glaucescens* as they contain only one individual for each species), and 2) *A. aculeata.* Population estimates of allelic richness, percentage of total alleles by locus, observed heterozygosity, expected heterozygosity, and inbreeding coefficient were calculated using the diveRsity [52], poppr [49], and the PopGenKit packages [53] on R platform [35]. 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 [54].

Results

In population genetics, neutral loci are genomic regions that are influenced by mutational dynamics and demographic effects, and not by selection. However, loci under selection (i.e., outliers) generally behave differently and therefore reveal "extreme" patterns of variation [55, 56]. Since most population genetic inferences are based on neutral loci, the loci under selection can greatly influence the estimates of genetic parameters. In this sense, it is important to identify and remove the outlier loci from the analysis, with the aim to infer more reliable parameters of population genetic diversity and structure.

Based on pcadapt, fsthet, and LEA, we identified 42 outlier loci for all samples or taxonomic groups for the genus Acrocomia, and 10 outlier loci for the taxonomic group formed by samples of *A. aculeata*. 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 3227 and 3259 neutral loci, respectively.

Genomic structure of Acrocomia spp.

Structure 2.3.4 software [43] was initially used to access the genomic structures of 172 samples of Acrocomia species based on 3227 neutral SNPs. ΔK had a maximum value of K = 7 (S1 Fig), indicating the existence of seven genetic groups (Fig 2). Samples with an attribution probability score > 0.75 and < 0.75 were assigned to the “pure group” and “admixture group”,

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

268 respectively. Based on the classification of Lorenzi [4] and the geographic distribution of the
269 species, we observed a substructure of samples considered to be *A. aculeata*. Two well-defined
270 subgroups (clusters 1 and 3) strongly associated with the geographical origin of the samples were
271 evident. Cluster 1 (Fig 2) was composed of 38 samples of *A. aculeata* from Central and North
272 America (Costa Rica, Trinidad and Tobago, Puerto Rico, and Mexico) and Colombia. Cluster 2 (Fig
273 2) comprised 39 samples of *A. totai* and five samples considered as *A. aculeata*. Of the latter, four
274 were collected in the state of Parana, southeastern Brazil, (XAM, PR) and one in state of
275 Tocantins, northern Brazil (PAL). The samples from Campo Grande (CGR) showed low mixture
276 levels with clusters 1 and 5 of *A. aculeata.* Cluster 3 (Fig 2) consisted of 39 samples from Brazil.
277 The majority (n = 34) of these samples were from the southeast region of the country, with five
278 from the north region (BEL population). Cluster 4 was exclusively *A. crispa* samples, with a 100%
279 probability of assignment to the cluster.
280
281 **Fig 2. Genomic structure of 172 samples from Acrocomia species based on 3227 neutral**
282 **SNPs loci.**
283 The y-axis is the population membership, and the x-axis is the sample. Each vertical bar
284 represents a sample and color represent separate clusters (k =7)
285
286 Based on assignment probabilities ≤ 0.75, some samples were assigned to an admixture
287 group. Twenty samples of *A. aculeata* from the central-west, north, and northeast regions of Brazil,
288 and all samples of *A. intumescens* displayed a similar genomic composition, with a median level of
289 assignment (≥ 0.50). A genetic admixture of *A. aculeata* samples in cluster 5 (Fig 2) with samples
290 mainly from clusters 1 and 3 was evident. *A. intumescens* samples presented a mixture of clusters
291 5 and 6, with cluster 6 being practically exclusive to the species. Individuals from Cáceres, MT
292 (CAC), and Braúna, São Paulo (SP) (BRA), with a greater assignment to cluster 2, also showed a
293 significant degree of admixture with clusters 3 and 5.
294 The NJ and PCoA analyses (Figs 3a and 3b) performed with all the samples showed strong
295 agreement with the results of the Bayesian analysis performed using Structure software. However,

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

296 the NJ tree showed higher resolution in group/cluster recovery than the PCoA. In both analyses, *A.*
297 *crispa* was clearly separated from the rest of the Acrocomia species. In addition, there is a clear
298 genomic differentiation between *A. aculeata* and *A. totai.* Similar to the results obtained using the
299 Structure software, the NJ analysis also recovered the substructure within *A. aculeata,* separating
300 the Brazilian samples from those from other countries (Fig 3a). This separation did not result from
301 the PCoA (Fig 3b). In agreement with the results obtained using the Structure software, both PCoA
302 and NJ grouped *A. media* and *A. intumescens* samples into the cluster formed mainly by *A.*
303 *aculeata,* with *A. hassleri* and *A. glaucescens* grouped into the *A. totai* cluster. The results of NJ
304 and PCoA also agreed concerning the allocation of samples from Xambré, PR (XAM) originally
305 considered as *A. aculeata* in the cluster of *A. totai*. Samples from Braúna, SP (BRA) and Cáceres,
306 MT (CAC), which were identified as an admixture by the Structure software, occupied an
307 intermediate position between the clusters formed mainly by *A. aculeata* and *A. totai* in the PCoA.
308
309 **Fig 3. Neighbor-joining (NJ) tree and principal components analysis (PCoA) of Acrocomia**
310 **species.**
311 a) Scatterplot of the principal components analysis (PCoA) showing the dispersion of samples
312 across the first two principal components and b) Neighbor-Joining dendrogram based on Nei´s
313 genetic distance. Bootstrap support of nodes is shown.
314
315 Based on the Structure software results (Fig 2) and NJ and PCoA data (Fig 3a and 3b), the
316 samples from Xambrê, PR (XAM) previously considered *A. aculeata* were treated as *A. totai*
317 species for further analysis of differentiation and genomic diversity. The F_{ST} values enabled a
318 moderate genetic differentiation between species, with an average value of 0.469. The F_{ST} values
319 between species (Table 1) ranged from 0.083 (*A. aculeata* vs. *A. totai)* to 0.946 (*A. media* vs. *A.*
320 *crispa*). In agreement with the genomic structure analysis findings, all comparisons between *A.*
321 *crispa* and the other species showed higher values of F_{ST}, demonstrating a greater degree of
322 genetic differentiation of *A. crispa* with the other species.
323

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1. Pairwise FST estimates among five species of Acrocomia

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

Genomic diversity between species

The number of polymorphic loci 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 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 relatively high levels of inbreeding in Acrocomia species, with the exception of *A. media*, which presented negative values (Table 2).

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

Species	Na	Ne	I	H_O	H_E	Ar	PAr	f
A. aculeata	1.601	1.142	0.157	0.031	0.093	1.20	0.03	0.479
A. totai	1.534	1.160	0.176	0.074	0.106	1.23	0.07	0.262
A. intumescens	1.053	1.027	0.037	0.011	0.025	1.10	0.02	0.483
A. média	0.993	0.984	0.007	0.006	0.005	1.08	0.01	-0.145
A. crispa	0.630	0.619	0.028	0.006	0.020	2.29	0.21	0.591

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 [44] 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).

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 (Red) and c) Scatterplot of the principal components analysis (PCoA) showing the dispersion of samples across the first two principal components

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*, which are suggested to be geographically separated by the Amazon Rainforest (Fig 4b and 4c). Thus, pronounced differentiation was observed between the individuals of *A. aculeata*. 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 (red) 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 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 hierarchical classification (Fig 4b and 4c).

Genomic diversity of *A. aculeata*

Concerning the genomic diversity within *A. aculeata* species, 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).

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

Country	Na	Ne	I	H_O	H_E	Ar	PAr	f
Trinidad &Tobago	1.001	0.983	0.013	0.008	0.008	1.09	0.01	0.038
Colombia	1.009	0.996	0.018	0.012	0.012	1.09	0.01	-0.014
Mexico	0.994	0.975	0.009	0.004	0.005	1.09	0.01	0.147
Costa Rica	0.981	0.971	0.012	0.009	0.008	1.11	0.01	-0.139

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Brazil	1.441	1.124	0.135	0.043	0.081	1.44	0.33	0.377
Brazilian State								
Pará	1.090	1.044	0.059	0.046	0.039	1.11	0.01	-0.168
Mato Grosso	1.090	1.031	0.091	0.062	0.061	1.23	0.04	-0.023
Góias	0.957	0.935	0.041	0.031	0.028	1.26	0.01	-0.068
Distrito Federal	0.870	0.836	0.053	0.033	0.035	1.49	0.01	0.015
Minas Gerais	1.190	1.094	0.089	0.043	0.058	1.09	0.01	0.198
Rio de Janeiro	1.024	0.987	0.037	0.024	0.024	1.14	0	-0.006
São Paulo	1.200	1.089	0.091	0.044	0.059	1.10	0.01	0.179
Maranhão	1.000	0.979	0.042	0.039	0.029	1.17	0.02	-0.350

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

Due to the vast territory and the greater number of *A. aculeata* samples from Brazil, genetic diversity analyses were conducted considering the Brazilian States as populations. Greater diversity (H_E) was found in the states of Mato Grosso (MT), 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 (MT), 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and the NJ tree (Figs 3a and 3b). A notable finding was the identification of an accentuated substructure within *A. aculeata,* showing two genetic groups, corresponding to a north– south split in which the samples from Brazil (Northern group, blue cluster in Fig 3) were separated from those of Central and North America (Southern 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 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 [57].

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 [58]. 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 [58], the low 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 [59, 60], the agricultural expansion

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 [61].

A. aculeata displays the greatest geographical distribution of the genus [3, 4, 14, 15]. As expected for a species with a wide distribution that has adapted to diverse environmental conditions, the genetic diversity of *A. aculeata* was high 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 [23, 28, 29].

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 (Table 4). 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” [62, 63]. 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 [63] and might also been reducing the genetic diversity. *A. aculeata* is strongly associated with humans [58, 59]. Even though it is considered an incipiently domesticated species, it has a wide range of uses in different countries of the Americas [7, 8, 64]. 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 [65], tomato [66], and cacao [67].

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A. totai

A. totai was the second most geographically dispersed species in the genus. It has been documented in eastern Bolivia, Paraguay, Central-west Brazil to northern Argentina [4, 15]. The taxonomic distinction of the species has been demonstrated based on morphological data and geographic distribution [4], leaf anatomy [1], and fruit biometry [68]. 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 [68]. 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 assignment probability (> 0.75), according to Structure analysis (Fig 2), and corroborated with PCoA and NJ analysis (Figs 3a and 3b). Our results agreed with those of Lima et al. [61], 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. (23) 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. (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% (68, 71, 72.

In Brazil, *A. totai* is considered to be restricted to the State of Mato Grosso do Sul [4, 69, 70]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

494 PCoA, and by NJ (Figs 3a and 3b). Although the occurrence of *A. totai* in these states has not
495 been proven, our results are consistent with the information reported on the Flora do Brazil 2020
496 website [15], indicating the possible occurrence of *A. totai* in these states.

497 Although the genetic structure and separation of *A. aculeata* from *A. totai* was evident
498 based on the cluster analyses, the genetic differentiation ($F_{ST)}$ between species was 0.083, which
499 was the lowest value (Table 1). These result was consistent with the value obtained using
500 microsatellite markers (F_{CT} = 0.07) by Lima et al., [61]. The findings may reflect the retention of
501 ancestral polymorphisms, the hybridization or gene flow between species in convergent areas [61]
502 or could be evidence of an ongoing speciation process [23].

503 Based on the H_E and Ar values, *A. totai* was the species with the highest level of genetic
504 diversity (Table 2). Our results are comparable to those found in a recent study using microsatellite
505 markers [61], in which the genetic diversity of *A. totai* was greater than that of *A. aculeata*. Similar,
506 previous studies also identified greater genetic diversity in populations from Mato Grosso do Sul
507 than population from other location of Brazil, although the authors did not consider the populations
508 to be *A. totai* [23, 25]. The high diversity observed in *A. totai* could reflect its geographically
509 widespread occurrence and expansion of genetic diversity promoted by the interspecific
510 hybridization with *A. aculeata.*

511 The results of cluster analysis and genetic differentiation corroborated the classification of
512 *A. totai* as an independent taxon based on morphological [4], anatomical [1], and molecular
513 markers [61]. This taxonomic separation seems to be more appropriate than that proposed for
514 Henderson et al. [3], which considered all tree-sized Acrocomias as a single taxonomic group
515 called *A. aculeata*.

516

517 *A. intumescens*

518 Contrary to the actual taxonomic classification [4-6], our analyses did not show a clear
519 genetic separation of *A. intumescens* (Figs 2, 3a, and 3b). All the samples of *A. intumescens* were

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 [68] and for oil content in the mesocarp, ranging from 37 to 78% in *A. aculeata* [71, 72] and from 34 to 41% in *A. intumescens* [71, 73].

A phylogenetic study by Meerow et al. [74], estimate 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, 15]. Species with a restricted geographical distribution tend to have reduced genetic diversity than species with a wide geographical distribution [75, 76]. 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 [77-79].

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 (Fig 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 [80, 81]. Our results are congruent with those reported for other tree species, which also showed high levels of genetic

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

differentiation between island populations compared to continental populations and lower levels of genetic diversity on the islands than on the continent [82-85]. *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 [74]. 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 [86, 87].

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 Bailey [57], the smaller fruits, 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) [68] and also differences in pollen morphology with trichotomocolpated pollen in *A. aculeata* and monocolpous pollen in *A. crispa* (named *Gastrococos crispa* by the authors) [88].

A. crispa, previously designated to the genus Gastrococos by Moore [89], was recently allocated to the genus Acrocomia, mainly due to the sequencing of the nuclear *prk* gene [90]. Although most phylogenetic studies that analyzed support the relationship between *A. aculeata* and *A. crispa* as sisters in a single monophyletic group [90-95], other phylogenetic [74] and cladistic studies [96] 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

573 The morphological characteristics of the species, the divergence time and our results of
574 genetic differentiation, diversity, and structure may collectively support an independent taxonomic
575 status of *A. crispa* within the genus Acrocomia. Therefore, we suggest a revision of the taxonomy
576 for the species.

577

578 *A. media*

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

586 The patterns of genetic diversity observed in *A. media* were the lowest compared to other
587 species (H_E= 0.005 and Ar = 1.08), but were consistent with several studies of population genetics
588 in plants, which predicted that island populations have reduced levels of genetic diversity
589 compared to continental populations [80, 97]. The low genetic diversity observed in *A. media* can
590 be attributed to the founder effect associated with the establishment of populations with only a few
591 individuals [97, 98] or to genetic drift due to stochastic events inherent in the islands and/or
592 fragmentation during its formation [99]

593 *A. media* was first described in Puerto Rico by Cook [100]. The author adopted the shortest
594 trunk and the smallest diameter of the stipe as the differentiating characteristics of *A. media* from
595 *A. aculeata*. However, *A. media* was considered synonymous with *A. aculeata* for a long time due
596 to the absence of consistent botanical characteristics for differentiation. In 2013, The Plant List
597 recognized *A. media* as a distinct species based on the floristic palm inventory of Proctor [101].
598 However, the same author mentioned that the existing information about *A. media* was very old
599 and based on few individuals, suggesting an increase in the number of evaluated individuals to

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 [90].

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* (Fig. 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 [102, 103]. 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

626 Acrocomia species as well as at the population level for *A. aculeata.* In general, our results were
627 partially congruent with the taxonomy of the genus, supporting the current separation of some
628 species. The genomic structure revealed the formation of well-defined genetic groups and
629 confirmed the distinction of *A. aculeata, A. totai*, *A. intumescens*, and *A. crispa*, with the latter
630 showing a strong genetic differentiation as well as the absence of genetic distinction of *A. media*.
631 We recommend a review of the current taxonomic classification of *A. crispa* and *A. media*. In
632 addition, SNPs also allowed the identification of gene flow patterns and/or hybridization between
633 species.

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

641

642 References

643 1. Vianna SA, Carmelo-Guerreiro SM, Noblick LR, Colombo CA. Leaf anatomy of Acrocomia
644 (Arecaceae): an additional contribution to the taxonomic resolution of a genus with great economic
645 potential. Plant Syst Evol. 2017;303:233-248.
646 2. Martius CFPV. Historia Naturalis Palmarum. Leipzig, TC Weigel. 1824:285-286.
647 3. Henderson A, Galeano G, Bernal R. Field guide to the palms of the Americas. New Jersey,
648 USA: Princeton University Press; 1995.
649 4. Lorenzi H, Noblick L, Kahn F. Flora Brasileira-Arecaceae (Palmeiras) Nova Odessa, SP:
650 Instituto Plantarum. 2010.
651 5. The Plant list Version 1.1 2013 [Available from: http://www.theplantlist.org/
652 6. Palmweb. Acrocomia Mart. Hist. Nat. Palm. n. d [Available from: http://www.palmweb.org/.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

653 7. Lorenzi GMAC. *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. Arecaceae: bases para o
654 extrativismo sustentável. Curitiba, PR: Universidade Federal do Paraná; 2006.
655 8. Ramos MIL, Ramos Filho MM, Hiane PA, Braga Neto JA, Siqueira EMA. Qualidade
656 nutricional da polpa de bocaiúva *Acrocomia aculeata* (Jacq.) Lodd. Food Sci Technol 2008;28:90-
657 94.
658 9. Colombo CA, Berton LHC, Diaz BG, Ferrari RA. Macauba: a promising tropical palm for the
659 production of vegetable oil. OCL. 2018;25(1):D108.
660 10. Coimbra MC, Jorge N. Characterization of the pulp and kernel oils from Syagrus oleracea,
661 Syagrus romanzoffiana, and *Acrocomia aculeata*. J Food Sci. 2011;76(8):C1156-C1161.
662 11. da Silva César A, de Azedias Almeida F, de Souza RP, Silva GC, Atabani A. The prospects
663 of using *Acrocomia aculeata* (macaúba) a non-edible biodiesel feedstock in Brazil. Renew Sust
664 Energ Rev. 2015;49:1213-1220.
665 12. Abreu AG, Priolli RHG, Azevedo-Filho JA, Nucci SM, Zucchi MI, Coelho RM, et al. The
666 genetic structure and mating system of *Acrocomia aculeata* (Arecaceae). Genet Mol Biol.
667 2012;35(1):116-121.
668 13. Abreu IS, Carvalho CR, Carvalho GMA, Motoike SY. First karyotype, DNA C-value and
669 AT/GC base composition of macaw palm (*Acrocomia aculeata*, Arecaceae)–a promising plant for
670 biodiesel production. Aust J Bot. 2011;59(2):149-155.
671 14. Dransfield J, Uhl NW, Lange CBA, Baker WJ, Harley MM, Lewis CE. Genera Palmarum: the
672 evolution and classification of palms: Kew Publishing; 2008.
673 15. Vianna SA, Campos_Rocha A. *Acrocomia in* Flora do Brasil 2020 em construção: Jardim
674 Botânico do Rio de Janeiro; 2020 [Available from:
675 <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB15662.
676 16. Vianna S, Colombo C, editors. Distribuição geográfica de acrocomia aculeata (jacq,) lodd ex
677 mart.(arecaceae) em sua região de ocorrência. I Congresso brasileiro de macaúba; 2013.
678 17. Costa CJ, Marchi ECS. Germinação de sementes de palmeiras com potencial para produção
679 de agroenergia. Embrapa Cerrados-Documentos (INFOTECA-E). 2008.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

680 18. Vinson C, Mangaravite E, Sebbenn A, Lander T. Using molecular markers to investigate
681 genetic diversity, mating system and gene flow of Neotropical trees. Braz J Bot. 2018;41(2):481-496.
682 19. Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. The Plant
683 Genome. 2012;5(3):92-102.
684 20. Boutet G, Carvalho SA, Falque M, Peterlongo P, Lhuillier E, Bouchez O, et al. SNP discovery
685 and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea
686 RIL population. BMC genom. 2016;17(1):1-14.
687 21. Davey JW, Blaxter ML. RADSeq: next-generation population genetics. Brief Funct Genomics.
688 2010;9(5-6):416-423.
689 22. Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population
690 genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS genet.
691 2010;6(2):e1000862.
692 23. Lanes ÉC, Motoike SY, Kuki KN, Nick C, Freitas RD. Molecular characterization and
693 population structure of the macaw palm, *Acrocomia aculeata* (Arecaceae), ex situ germplasm
694 collection using microsatellites markers. J Hered. 2015;106(1):102-112.
695 24. Lanes ÉC, Motoike SY, Kuki KN, Resende MD, Caixeta ET. Mating system and genetic
696 composition of the macaw palm (*Acrocomia aculeata*): implications for breeding and genetic
697 conservation programs. J Hered. 2016;107(6):527-536.
698 25. Mengistu FG, Motoike SY, Cruz CD. Molecular characterization and genetic diversity of the
699 macaw palm ex situ germplasm collection revealed by microsatellite markers. Diversity.
700 2016;8(4):20.
701 26. Nucci SM. Desenvolvimento, caracterização e análise da utilidade de marcadores
702 microssatélites em genética de população de macaúba. : Instituto Agronômico de Campinas; 2007.
703 27. Silva LCC, Lemos RdC, Carvalho CGI, Good-God PIV, Oliveira LOd, Costa MD-BL, et al.
704 Genetic diversity and structure of macaw palm: implications for genetic variability sampling. Rev
705 Arvore. 2017;41(5).

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

706 28. Oliveira D, Melo Júnior A, Brandão M, Rodrigues L, Menezes E, Ferreira P. Genetic diversity
707 in populations of Acrocomia aculeata (Arecaceae) in the northern region of Minas Gerais, Brazil.
708 Genet Mol Res. 2012;11(1):531-538.
709 29. Coelho NHP, Tambarussi EV, Aguiar BI, Roque RH, Portela RM, Braga RC, et al.
710 Understanding genetic diversity, spatial genetic structure, and mating system through microsatellite
711 markers for the conservation and sustainable use of *Acrocomia aculeata* (Jacq.) Lodd. Ex Mart.
712 Conserv Genet. 2018;19(4):879-891.
713 30. Velazquez MAM. Caracterización fitoquímica, molecular y evaluación de respuestas del
714 cultivo in vitro de coyol (Acrocomia mexicana Kraw ex Marti)/por Antonio Magdiel Velazquez
715 Mendez: Universidad Autónoma Chapingo; 2013.
716 31. Doyle JJ, Doyle JL. Isolation ofplant DNA from fresh tissue. Focus. 1990;12(13):39-40.
717 32. Poland JA, Brown PJ, Sorrells ME, Jannink J-L. Development of high-density genetic maps
718 for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS one.
719 2012;7(2):e32253.
720 33. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and
721 genotyping loci de novo from short-read sequences. G3: Genes genom genet. 2011;1(3):171-182.
722 34. Luu K, Bazin E, Blum MG. pcadapt: an R package to perform genome scans for selection
723 based on principal component analysis. Mol Ecol Resour. 2017;17(1):67-77.
724 35. R Core Team. R: A language and environment for statistical computing. Vienna, Austria;
725 2018.
726 36. Flanagan SP, Jones AG. Constraints on the F ST–heterozygosity outlier approach. J Hered.
727 2017;108(5):561-573.
728 37. Wright S. Isolation by distance. Genetics. 1943;28(2):114.
729 38. Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97.
730 39. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and
731 environmental gradients using latent factor mixed models. Mol Biol Evol 2013;30(7):1687-1699.
732 40. Frichot E, François O. LEA: An R package for landscape and ecological association studies.
733 Methods Ecol Evol. 2015;6(8):925-929.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

734 41. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated
735 climate surfaces for global land areas. Int J Climatol. 2005;25:965–1978.
736 42. Martins H, Caye K, Luu K, Blum MG, François O. Identifying outlier loci in admixed and in
737 continuous populations using ancestral population differentiation statistics. Molecular ecology.
738 2016;25(20):5029-5042.
739 43. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus
740 genotype data. Genetics. 2000;155(2):945-959.
741 44. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the
742 software STRUCTURE: a simulation study. Mol Ecol 2005;14(8):2611-2620.
743 45. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing
744 STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359-
745 361.
746 46. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for
747 dealing with label switching and multimodality in analysis of population structure. Bioinformatics.
748 2007;23(14):1801-1806.
749 47. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for
750 identifying clustering modes and packaging population structure inferences across K. Molecular
751 ecology resources. 2015;15(5):1179-1191.
752 48. Nei M. Estimation of average heterozygosity and genetic distance from a small number of
753 individuals. Genetics. 1978;89(3):583-590.
754 49. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population
755 genetic data with emphasis on clonality. Frontiers in genetics. 2015;6:208.
756 50. Dray S, Dufour AB, Chessel D. The ade4 package-II: Two-table and K-table methods. R
757 news. 2007;7(2):47-52.
758 51. Villanueva RAM, Chen ZJ. ggplot2: Elegant graphics for data analysis. Taylor & Francis;
759 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

760 52. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. diveRsity: An R package for
761 the estimation and exploration of population genetics parameters and their associated errors.
762 Methods in ecology and evolution. 2013;4(8):782-788.
763 53. Paquette SR, Paquette MSR. Package 'PopGenKit'. 2011.
764 54. Kalinowski ST. hp-rare 1.0: a computer program for performing rarefaction on measures of
765 allelic richness. Molecular ecology notes. 2005;5(1):187-189.
766 55. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of
767 population genomics: from genotyping to genome typing. Nat Rev Genet. 2003;4(12):981-994.
768 56. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics.
769 Nature Rev Genet. 2010;11(10):697-709.
770 57. Bailey LH. *Acrocomia* preliminary paper. Gentes Herbarum – Occasional Papers on the
771 Kinds Plants; 1941.
772 58. Morcote-Rios G, Bernal R. Remains of palms (Palmae) at archaeological sites in the New
773 World: a review. Bot Rev. 2001;67(3):309-350.
774 59. Lentz DL. *Acrocomia mexicana*: Palm of the ancient Mesoamericans. J Ethnobiol.
775 1990;10(2):183-194.
776 60. Scariot A. Seed dispersal and predation of the palm *Acrocomia aculeata*. Principes.
777 1998;42:5-8.
778 61. de Lima NE, Meerow AW, Manfrin MH. Genetic structure of two Acrocomia ecotypes
779 (Arecaceae) across Brazilian savannas and seasonally dry forests. Tree Genet Genomes.
780 2020;16(4):1-12.
781 62. Balick MJ. Production of coyol wine from Acrocomia mexicana (Arecaceae) in Honduras.
782 Econ Bot. 1990;44(1):84-93.
783 63. H. D-FV, Amín R-CP, Beau M-VB, Moises A-B. Caracterización morfológica del coyol
784 Acrocomia aculeata (Jacq.), y su potencial productivo de aceite vegetal. In: Instituto Nacional de
785 Investigaciones Forestales AyPCdIRPSCERI, editor. Folleto Técnico Tuxtla Chico, Chiapas,
786 México2020. p. 54.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

787 64. Ramírez Hernández BC, Zañudo Hernández J, García de Alba Verduzco JE, Délano Frier
788 JP, Pimienta Barrios E, García Martínez MÁ. Importancia agroecológica del coyul (*Acrocomia*
789 *mexicana* Karw. ex Mart.). Estudios sociales (Hermosillo, Son). 2013;21(41):95-113.
790 65. Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, et al. Linkage disequilibrium and
791 population structure in wild and domesticated populations of Phaseolus vulgaris L. Evol Appl.
792 2009;2(4):504-522.
793 66. Blanca J, Cañizares J, Cordero L, Pascual L, Diez MJ, Nuez F. Variation revealed by SNP
794 genotyping and morphology provides insight into the origin of the tomato. PloS one.
795 2012;7(10):e48198.
796 67. Cornejo OE, Yee M-C, Dominguez V, Andrews M, Sockell A, Strandberg E, et al. Population
797 genomic analyses of the chocolate tree, Theobroma cacao L., provide insights into its domestication
798 process. Communications biology. 2018;1(1):1-12.
799 68. Vianna SA, Berton LHC, Pott A, Guerreiro SMC, Colombo CA. Biometric characterization of
800 fruits and morphoanatomy of the mesocarp of Acrocomia species (Arecaceae). Int J Biol.
801 2017;9(3):78-92.
802 69. Markley KS. Mbocayá or Paraguay Cocopalm—an important source of oil. Econ Bot.
803 1956;10(1):3-32.
804 70. Rodríguez MF, Aschero CA. Acrocomia chunta (Arecaceae) raw material for cord making in
805 the Argentinean Puna. J Archaeol Sci. 2005;32(10):1534-1542.
806 71. Conceição L, JUNQUEIRA NTV, Licurgo FdS, ANTONIASSI R, WILHELM AE, BRAGA MF,
807 editors. Teor de óleo em frutos de diferentes espécies de macaubeira (Acrocomia spp.). Embrapa
808 Cerrados-Artigo em anais de congresso (ALICE); 2012: In: CONGRESSO BRASILEIRO DE
809 FRUTICULTURA, 22., 2012, Bento Gonçalves. Anais ….
810 72. Berton L. Avaliação de populações naturais, estimativas de parâmetros genéticos e seleção
811 de genótipos elite de macaúba (Acrocomia aculeata). Campinas, SP. Brazil: Instituto Agronômico
812 de Campinas - IAC; 2013.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

813 73. Bora PS, Rocha R. Macaiba palm: fatty and amino acids composition of fruits macaíba:
814 composición de aminoácidos y ácidos grasos de frutos macaíba: composición de aminoácidos e
815 ácidos graxos de froitos. CYTA-J Food. 2004;4(3):158-162.
816 74. Meerow AW, Noblick L, Salas-Leiva DE, Sanchez V, Francisco-Ortega J, Jestrow B, et al.
817 Phylogeny and historical biogeography of the cocosoid palms (A recaceae, A recoideae, C ocoseae)
818 inferred from sequences of six WRKY gene family loci. Cladistics. 2015;31(5):509-534.
819 75. Cole CT. Genetic variation in rare and common plants. Annu Rev Ecol Evol Syst.
820 2003;34(1):213-237.
821 76. Hamrick J, Godt M. Allozyme diversity in plant species. In 'Plant population genetics,
822 breeding and genetic resources'.(Eds AHD Brown, MT Clegg, AL Kahler, BS Weir) pp. 43–63.
823 Sinauer Associates: Sunderland, MA; 1989.
824 77. Ellis J, Pashley C, BURKE* J, McCauley D. High genetic diversity in a rare and endangered
825 sunflower as compared to a common congener. Mol Ecol 2006;15(9):2345-2355.
826 78. Furches MS, Small RL, Furches A. Genetic diversity in three endangered pitcher plant
827 species (Sarracenia; Sarraceniaceae) is lower than widespread congeners. Am J Bot.
828 2013;100(10):2092-2101.
829 79. Levy E, Byrne M, Coates D, Macdonald B, McArthur S, Van Leeuwen S. Contrasting
830 influences of geographic range and distribution of populations on patterns of genetic diversity in two
831 sympatric Pilbara Acacias. PLoS One. 2016;11(10):e0163995.
832 80. Franks SJ. Genetics, evolution, and conservation of island plants. J Plant Biol. 2010;53(1):1-
833 9.
834 81. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining,
835 estimating and interpreting F ST. Nat Rev Genet. 2009;10(9):639-650.
836 82. Chen D, Zhang X, Kang H, Sun X, Yin S, Du H, et al. Phylogeography of Quercus variabilis
837 based on chloroplast DNA sequence in East Asia: multiple glacial refugia and mainland-migrated
838 island populations. PLoS One. 2012;7(10):e47268.
839 83. Qi X-S, Yuan N, Comes HP, Sakaguchi S, Qiu Y-X. A strong 'filter'effect of the East China
840 Sea land bridge for East Asia's temperate plant species: inferences from molecular phylogeography

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

841 and ecological niche modelling of Platycrater arguta (Hydrangeaceae). BMC Evol Biol
842 2014;14(1):41.
843 84. Qiu YX, Sun Y, Zhang XP, Lee J, Fu CX, Comes HP. Molecular phylogeography of East
844 Asian Kirengeshoma (Hydrangeaceae) in relation to Quaternary climate change and landbridge
845 configurations. New Phytol. 2009;183(2):480-495.
846 85. Sakaguchi S, QIU YX, LIU YH, QI XS, KIM SH, Han J, et al. Climate oscillation during the
847 Quaternary associated with landscape heterogeneity promoted allopatric lineage divergence of a
848 temperate tree Kalopanax septemlobus (Araliaceae) in East Asia. Mol Ecol. 2012;21(15):3823-3838.
849 86. Mitsui Y, Chen S-T, Zhou Z-K, Peng C-I, Deng Y-F, Setoguchi H. Phylogeny and
850 biogeography of the genus Ainsliaea (Asteraceae) in the Sino-Japanese region based on nuclear
851 rDNA and plastid DNA sequence data. Ann Bot. 2008;101(1):111-124.
852 87. Nakamura K, Denda T, Kokubugata G, Huang C-J, Peng C-I, Yokota M. Phylogeny and
853 biogeography of the Viola iwagawae-tashiroi species complex (Violaceae, section Plagiostigma)
854 endemic to the Ryukyu Archipelago, Japan. Plant Syst Evol. 2015;301(1):337-351.
855 88. Machado SR. Variaciones en la morfología polínica de Arecaceae en Cuba: abertura
856 tricotomosulcada y estratificación de la exina. Revista Jard Bot Nac Univ Habana. 2003:71-79.
857 89. Moore HE. *Gastrococos crispa* (Kunth). Principes. 11.1968.
858 90. Gunn BF. The phylogeny of the Cocoeae (Arecaceae) with emphasis on Cocos nucifera. Ann
859 Missouri Bot Gard. 2004:505-522.
860 91. Hahn WJ. A phylogenetic analysis of the Arecoid line of palms based on plastid DNA
861 sequence data. Mol Phylogenet Evol. 2002;23(2):189-204.
862 92. Asmussen CB, Dransfield J, Deickmann V, Barfod AS, Pintaud J-C, Baker WJ. A new
863 subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. Bot J
864 Linn Soc. 2006;151(1):15-38.
865 93. Couvreur TL, Hahn W, Granville J-Jd, Pham J-L, Ludena B, Pintaud J-C. Phylogenetic
866 relationships of the cultivated Neotropical palm Bactris gasipaes (Arecaceae) with its wild relatives
867 inferred from chloroplast and nuclear DNA polymorphisms. Syst Bot. 2007;32(3):519-530.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

868 94. Eiserhardt WL, Pintaud J-C, Asmussen-Lange C, Hahn WJ, Bernal R, Balslev H, et al.
869 Phylogeny and divergence times of Bactridinae (Arecaceae, Palmae) based on plastid and nuclear
870 DNA sequences. Taxon. 2011;60(2):485-498.
871 95. Roncal J, Kahn F, Millan B, Couvreur TL, Pintaud J-C. Cenozoic colonization and
872 diversification patterns of tropical American palms: evidence from Astrocaryum (Arecaceae). Bot J
873 Linn Soc. 2013;171(1):120-139.
874 96. Noblick LR, Hahn WJ, Griffith MP. Structural cladistic study of Cocoseae, subtribe Attaleinae
875 (Arecaceae): Evaluating taxonomic limits in Attaleinae and the neotropical genus Syagrus. Brittonia.
876 2013;65(2):232-261.
877 97. Whittaker RJ, Fernández-Palacios JM. Island biogeography: ecology, evolution, and
878 conservation: Oxford University Press; 2007.
879 98. Templeton AR. The reality and importance of founder speciation in evolution. BioEssays.
880 2008;30(5):470-479.
881 99. Stuessy TF, Takayama K, López-Sepúlveda P, Crawford DJ. Interpretation of patterns of
882 genetic variation in endemic plant species of oceanic islands. Bot J Linn Soc. 2014;174(3):276-288.
883 100. Cook OF. A synopsis of the palms of Puerto Rico. . The Botanical Club. 281901. p. 565-567.
884 101. Proctor GR. Arecaceae (Palmae). P Acevedo-Rodríguez, M T Strong. Monocots and
885 Gymnosperms of Puerto Rico and the Virgin Islands.: Smithsonian Institution: Contributions of the
886 United States National Herbarium; 2005.
887 102. Cavers S, Degen B, Caron H, Lemes MR, Margis R, Salgueiro F, et al. Optimal sampling
888 strategy for estimation of spatial genetic structure in tree populations. Heredity. 2005;95(4):281-289.
889 103. Morin PA, Martien KK, Taylor BL. Assessing statistical power of SNPs for population structure
890 and conservation studies. Mol Ecol Resour. 2009;9(1):66-73.

891

892 Supporting information

893 **S1 Table. Geographical location and origin of the Acrocomia species samples.**

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

894 **S2 Table. SNP loci outliers (putatively under selection) for Acrocomia species and within *A.***

895 ***aculeata* identified by PCAdapt, Fsthet and LEA packages.**

896 **S1 Fig. Delta (Δ) K values for different numbers of populations assumed (K) in the**

897 **STRUCTURE analysis, estimated based on Evanno method for all Acrocomia species.**

898 **S2 Fig. Delta (Δ) K values for different numbers of populations assumed (K) in the**

899 **STRUCTURE analysis, estimated based on Evanno method for *A. aculeata.***

900 **S1 File. SNP genotype information in variant calling format (vcf) for 172 samples of**

901 **Acrocomia species.**

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

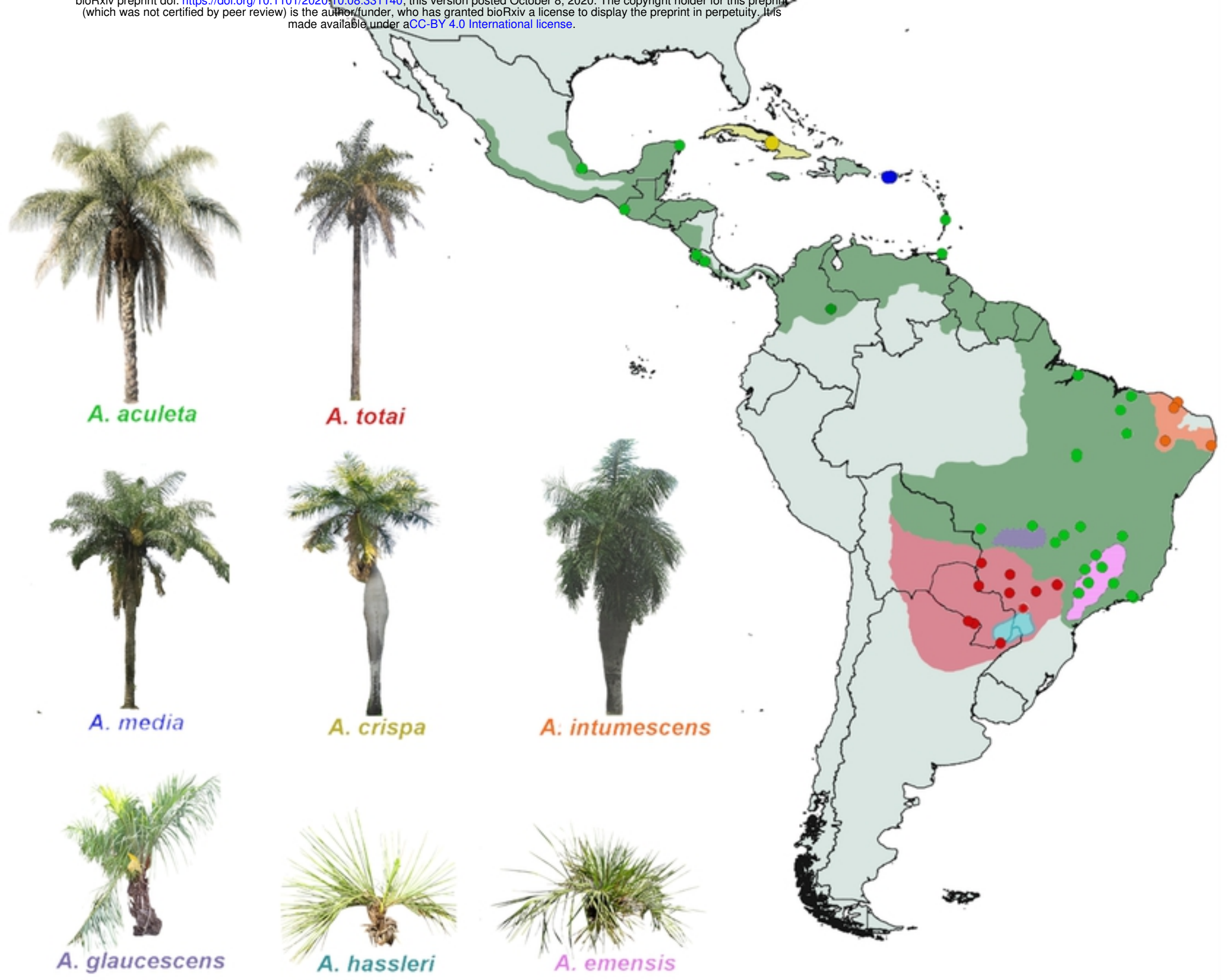


Fig1.species distribution

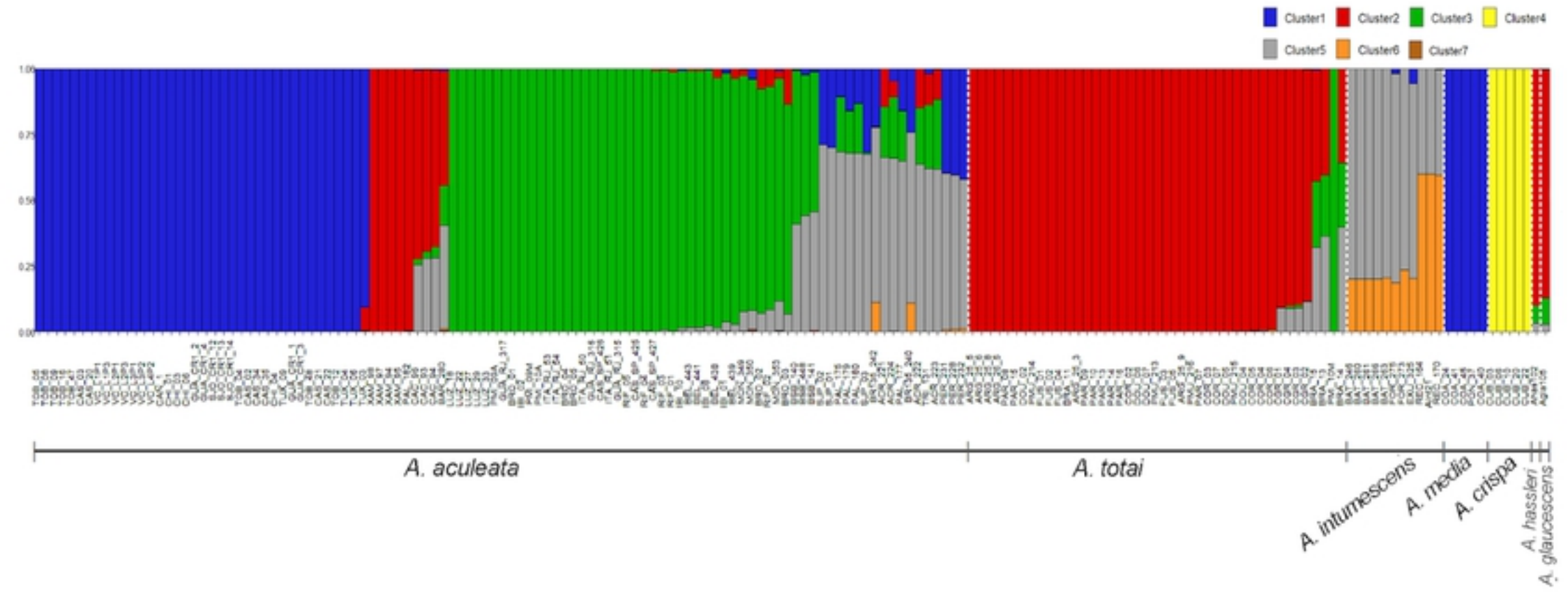


Fig2.species structure

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.08.331140; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

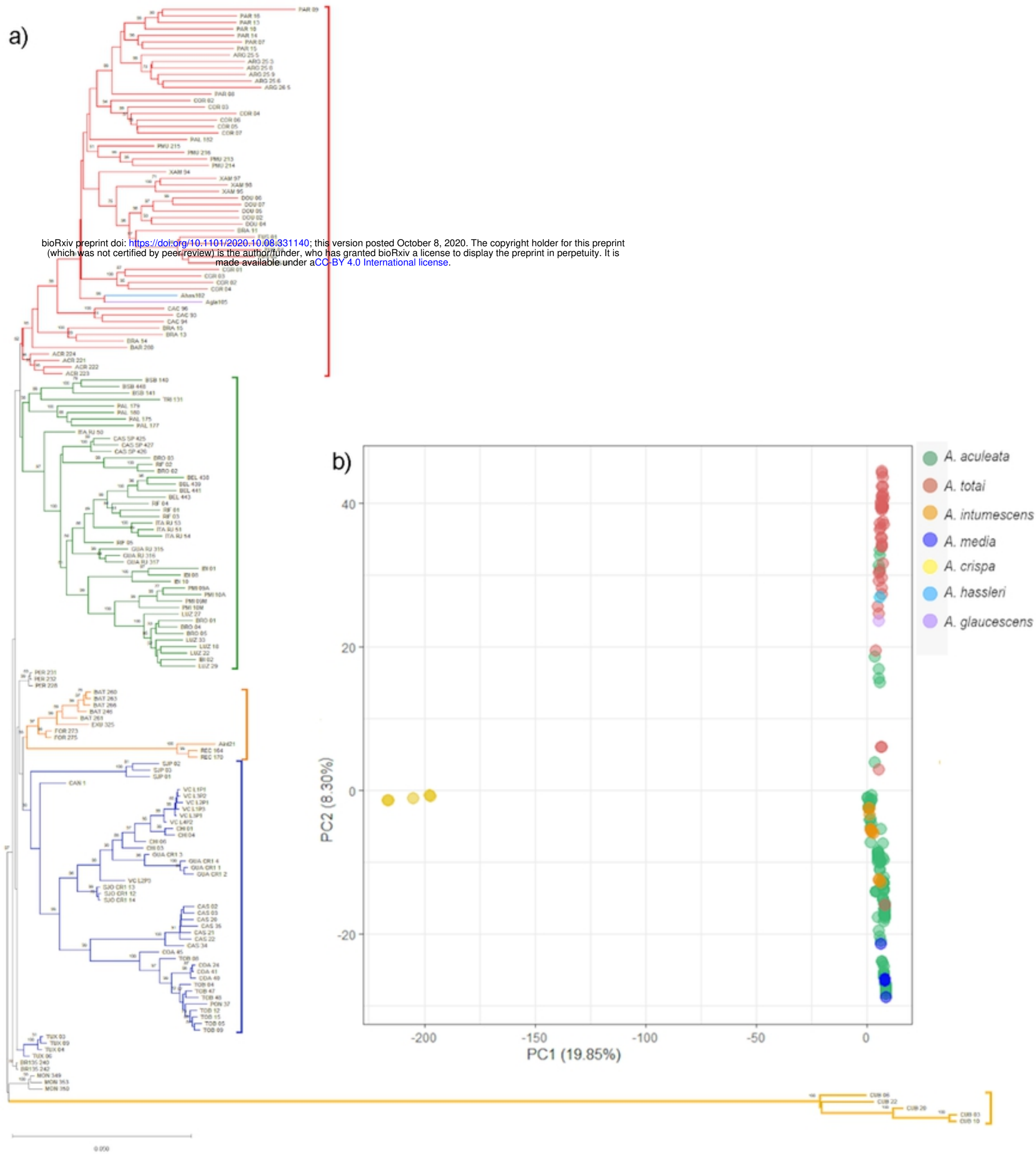
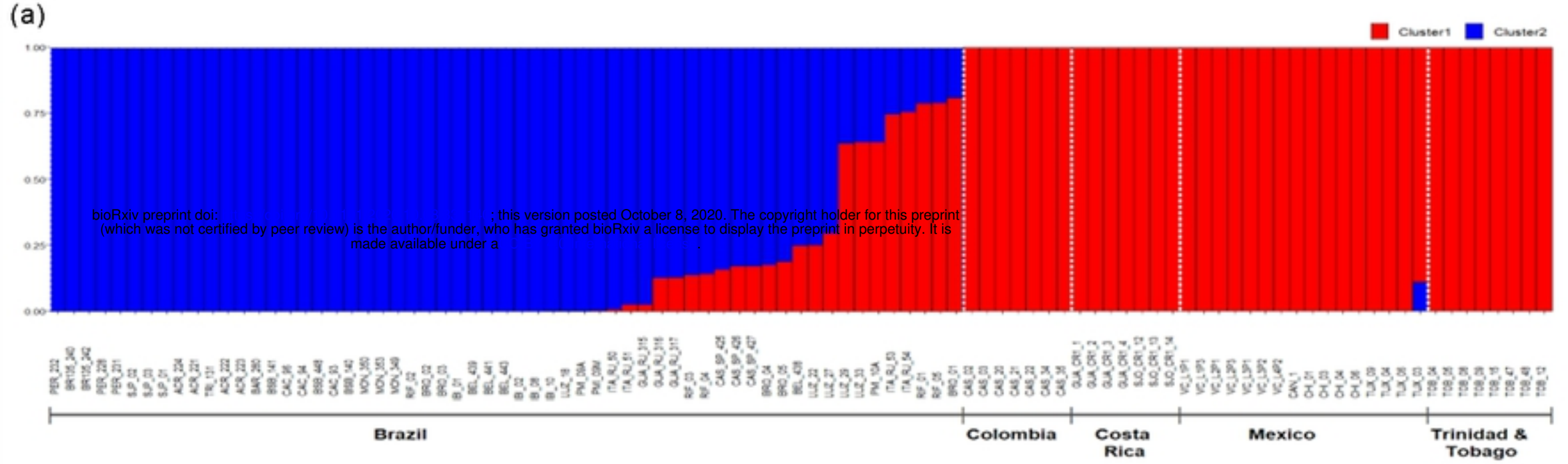


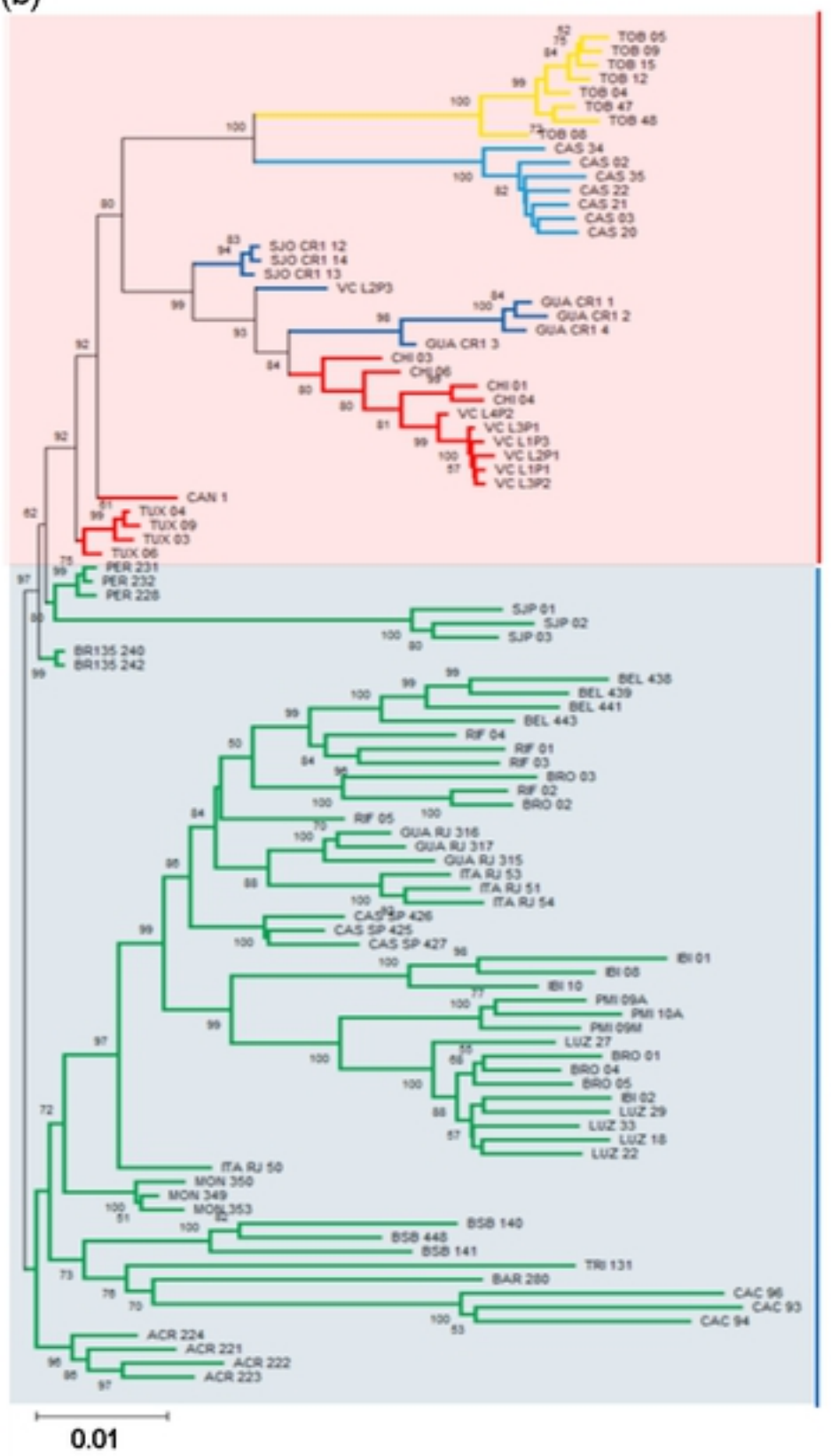
Fig3.species NJ_PCoA

bioRxiv preprint doi: ; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a .

(a)



(b)



(c)

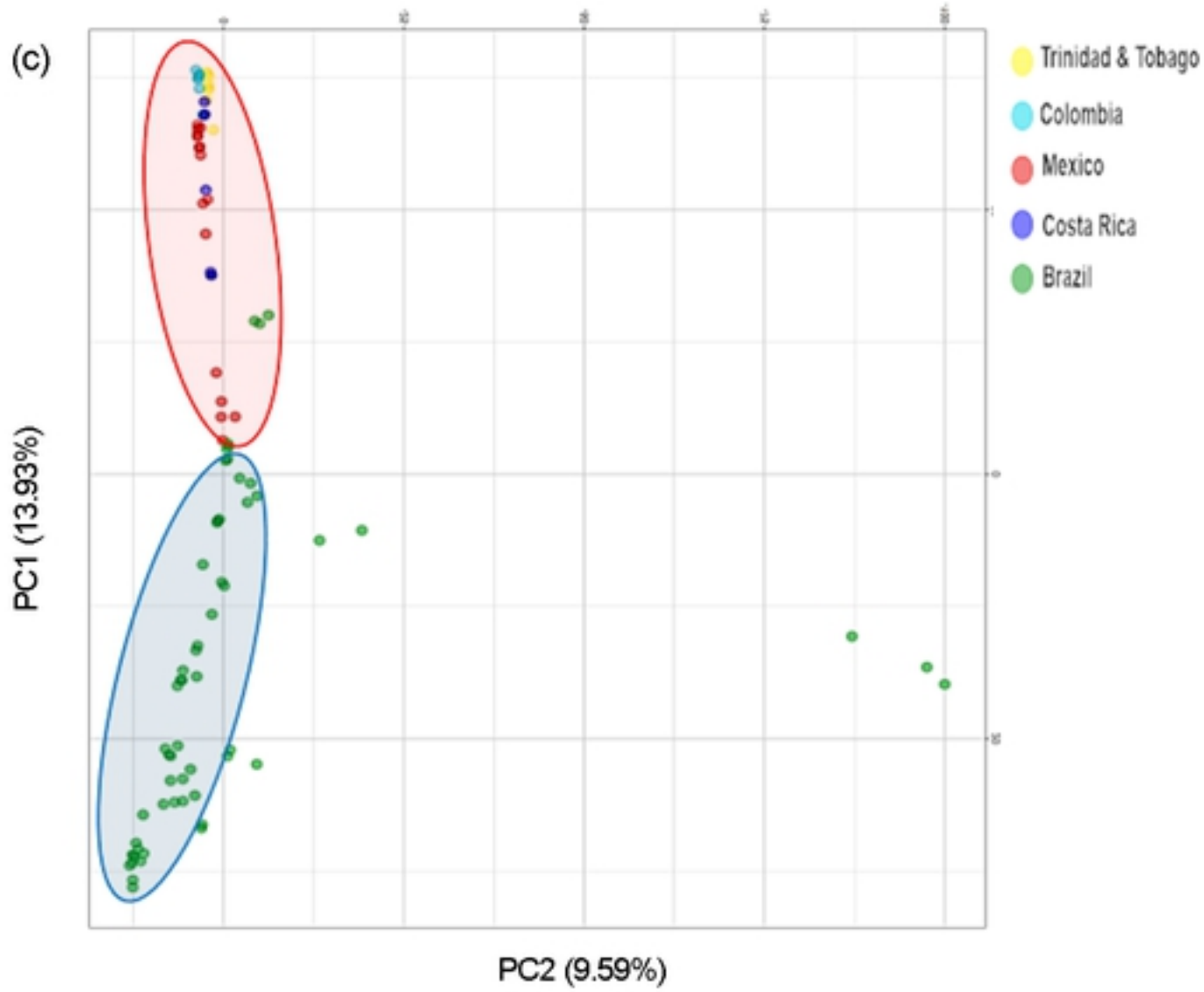


Fig4.A_aculeata strcu_NJ_PCoA