diversity

MDPI

Article

Molecular Characterization and Genetic Diversity of the Macaw Palm Ex Situ Germplasm Collection Revealed by Microsatellite Markers

Fekadu G. Mengistu [1,*], Sérgio Y. Motoike [2] and Cosme D. Cruz [3]

1 Kulumsa Agricultural Research Center (KARC), Ethiopian Institute of Agricultural Research (EIAR), P.O.Box 489, Asella, Ethiopia
2 Departamento de Fitotecnia, Universidade Federal de Viçosa, Av. P.H. Rolfs, Campus, Viçosa, MG 36570-000, Brazil; motoike@ufv.br
3 Departamento de Biologia Geral, Universidade Federal de Viçosa, Av. P.H. Rolfs, Campus, Viçosa, MG 36570-000, Brazil; cdcruz@ufv.br
* Correspondence: fgebretensay@yahoo.com; Tel.: +251-0968-23-55-27; Fax: +251-022-331-1508

Academic Editor: Mario A. Pagnotta
Received: 29 June 2016; Accepted: 9 October 2016; Published: 13 October 2016

Abstract: Macaw palm (*Acrocomia aculeata*) is native to tropical forests in South America and highly abundant in Brazil. It is cited as a highly productive oleaginous palm tree presenting high potential for biodiesel production. The aim of this work was to characterize and study the genetic diversity of *A. aculeata* ex situ collections from different geographical states in Brazil using microsatellite (Simple Sequence Repeats, SSR) markers. A total of 192 accessions from 10 provenances were analyzed with 10 SSR, and variations were detected in allelic diversity, polymorphism, and heterozygosity in the collections. Three major groups of accessions were formed using PCoA—principal coordinate analysis, UPGMA—unweighted pair-group method with arithmetic mean, and Tocher. The Mantel test revealed a weak correlation (r = 0.07) between genetic and geographic distances among the provenances reaffirming the result of the grouping. Reduced average heterozygosity (Ho < 50%) per locus (or provenance) confirmed the predominance of endogamy (or inbreeding) in the germplasm collections as evidenced by positive inbreeding coefficient (F > 0) per locus (or per provenance). AMOVA—Analysis of Molecular Variance revealed higher (48.2%) genetic variation within population than among populations (36.5%). SSR are useful molecular markers in characterizing *A. aculeata* germplasm and could facilitate the process of identifying, grouping, and selecting genotypes. Present results could be used to formulate appropriate conservation strategies in the genebank.

Keywords: *Acrocomia aculeata*; biodiesel; domestication; genebank; genetic diversity; SSRs

1. Introduction

Macaw palm (*Acrocomia aculeata* (Jacq.) (Lodd. ex Mart.))–*Arecaceae* (2n = 2x = 30) is commonly known as macaúba in Brazil [1]. This arborescent, spiny and single-stemmed palm is monoecious and self-compatible, and entomophily and anemophily forms of pollinations are reported [2]. It bears a mixed reproductive system, with a predominance of outcrossing [3,4]. The combination of the two pollination strategies with flexible reproductive systems suggests that *A. aculeata* can be highly successful in the colonization of new areas, as evidenced by the ample distribution of the species in the Neotropics. It is a very resilient palm and has abundant distribution in Brazil mainly in the regional States of Ceará, Minas Gerais, Mato Grosso, Mato Grosso do Sul, and São Paulo [2].

A. aculeata is little known globally, however, in recent years, it has raised interest due to its potential for social and economic use as an oil producer, considering that it is cited as one of the most important new sources of oil for biofuel [5,6]. It produces fruits yielding up to 25 t/ha corresponding

to about 4000 kg of oil. The solid waste is converted to charcoal and nutritious cakes that can be used to generate energy and feed livestock as well [7,8]. The biochemical properties of the oil are proved to be suitable for the cosmetic industry and for biodiesel production [9–11]. Moreover, this palm has environmental benefits as it can be fostered in impoverished soils and drought prevailing areas, which is a desirable trait for plants in order to rehabilitate degraded pastures or for agroforestry practices [12]. Hence, *A. aculeata* can be a suitable option for production of biodiesel among the common food-based oleaginous plants such as soya bean, sunflower, and oil palms [13].

It is not commercially cultivated like the domesticated *Arecaceae* palms such as *Elaeis guineensis*, *Cocus nucifera*, and *Enterpe oleraceae*, which are important elements in the Brazilian savannah, and present great genetic diversity in natural populations [5,14,15]. The palm's genetic diversity suffers by predatory extractivism, unsustainable land use, and climate change [16,17]. Hence, genetic resource conservation and its sustainable use have paramount importance for future genetic improvement. A central point in its sustainable conservation is the knowledge of the genetic diversity present in genebank collection and potential exploitation of the genetic materials by breeding programs. The germplasm characterization and species genetic diversity could be effectively integrated by molecular analyses.

Therefore, we characterized and studied the genetic diversity of the macaw palm germplasm collections in a genebank using microsatellites (Simple Sequence Repeats—SSRs) [18,19]. SSRs are well known molecular markers for their potentially high information content and versatility as molecular tools in germplasm characterization [20,21]. They are often co-dominant, highly reproducible, frequent in most eukaryotes and are quite useful in various aspects of molecular genetic studies [22,23]. Another aim was to study the distribution of the genetic diversity and in particular if a correlation exists between the genetic and the geographic distances and if distinct genetic groupings are formed among populations. These results will be useful in future conservation activities.

2. Experimental Section

2.1. Plant Material and DNA Isolation

Leaf samples from 192 *A. aculeata* germplasm accessions were obtained from the ex situ plant collection, macaúba Active Genebank, situated in Araponga (S2040 01, W423115), State of Minas Gerais, Brazil. The accessions were originated from seeds collected in six regional states of the country (Figure 1) and germinated using a pre-germination protocol as described in patent INPI 014070005335 [24]. The accessions represent 10 provenances having a total of 41 populations coded as BGP and 3–5 individuals were considered per population (Table 1).

Figure 1. Map shows the six geographical states in Brazil, where the original plant materials were obtained. MG = Minas Gerais; SP = São Paulo; MS = Mato Grosso do Sul; PA = Pará; PE = Pernambuco; PB = Paraiba. Araponga is a city in MG State, where the genebank is situated in which the experimental plant materials were collected.

Table 1. List of 41 *Acrocomia aculeata* populations, number of individuals, states (or provenances), and GPS coordinates.

No.	Population	Number of Individuals	State/Provenance *	Coordinates **		No.	Population	Number of Individuals	State/Provenance *	Coordinates **	
				Latitude	Longitude					Latitude	Longitude
1	BGP99	5	PA	S 06 03 58.0	W 49 33 39.0	22	BGP11	5	EMG	S 19 14 01.2	W 43 03 28.4
2	BGP82	5	PE	S 07 14 23.0	W 36 46 55.0	23	BGP9	5	EMG	S 19 33 12.0	W 46 51 10.1
3	BGP124	4	PB	S 08 48 49.0	W 36 57 14.0	24	BGP78	5	EMG	S 18 51 25.6	W 46 52 55.2
4	BGP51	5	SP	S 21 32 04.6	W 48 44 24.7	25	BGP37	5	EMG	S 18 40 51.3	W 46 33 41.4
5	BGP34	5	SP	S 22 25 10.8	W 50 34 43.1	26	BGP33	5	EMG	S 19 19 40.3	W 46 38 11.5
6	BGP47	5	SP	S 22 29 14.2	W 50 46 16.2	27	BGP21	4	WMG	S 19 31 15.9	W 46 31 42.2
7	BGP20	5	NMG	S 16 39 52.7	W 43 53 58.9	28	BGP2	5	WMG	S 20 39 20.4	W 43 18 45.2
8	BGP27	5	NMG	S 16 21 20.7	W 44 25 30.5	29	BGP76	5	WMG	S 19 41 51.4	W 43 11 27.7
9	BGP22	4	NMG	S 17 25 54.0	W 45 08 59.5	30	BGP25	5	WMG	S 17 06 54.6	W 43 49 16.4
10	BGP16	5	NMG	S 16 26 07.6	W 44 00 50.5	31	BGP64	5	WMG	S 16 44 12.7	W 43 51 54.9
11	BGP49	5	NMG	S 20 38 58.0	W 44 01 15.5	32	BGP105	3	MS	S 20 29 52.5	W 55 18 39.3
12	BGP10	5	NMG	S 21 03 12.9	W 44 16 28.2	33	BGP102	4	MS	S 20 30 38.6	W 55 37 59.7
13	BGP68	5	SMG	S 21 11 27.6	W 44 19 29.7	34	BGP104	4	MS	S 20 27 55.9	W 55 46 41.7
14	BGP3	5	SMG	S 21 09 52.2	W 44 08 49.5	35	BGP117	3	MS	S 20 27 56.5	W 55 46 38.2
15	BGP51	5	SMG	S 21 17 20.5	W 44 49 12.6	36	BGP118	4	MS	S 20 50 22.3	W 55 54 53.3
16	BGP5	5	SMG	S 19 05 02.0	W 44 39 13.9	37	BGP112	5	MS	S 20 50 16.5	W 55 54 51.8
17	BGP14	5	SMG	S 19 56 29.0	W 44 36 12.0	38	BGP106	3	MS	S 21 28 42.3	W 56 10 03.6
18	BGP18	5	CMG	S 19 52 34.0	W 43 52 20.5	39	BGP92	5	MS	S 21 28 45.7	W 56 10 06.6
19	BGP24	5	CMG	S 19 53 20.2	W 43 41 11.5	40	BGP103	5	MS	S 21 42 04.8	W 57 50 39.0
20	BGP1	5	CMG	S 20 17 42.6	W 43 42 30.9	41	BGP119	4	MS	S 21 42 06.0	W 57 50 32.4
21	BGP52	5	CMG	S 20 50 13.1	W 42 54 27.3						

* State (or provenances) including: PA = Pará, PE = Pernambuco, PB=Paraiba, SP = São Paulo, NMG = North Minas Gerais, SMG = South Minas Gerais, CMG = Central Minas Gerais, EMG = East Minas Gerais, WMG = West Minas Gerais, and MS = Mato Grosso do Sul. ** Coordinates are in degrees, minutes, and seconds for both the latitude (S = South) and longitude (W = West).

Genomic DNA was isolated from leaflets following the CTAB (Cetyl Tri-methyl Ammonium Bromide) method [25] with modifications as described in [26]. DNA samples were quantified with a Multiscan^{TM}GO Microplate Spectrophotometer using absorbance at 260/280 nm. The integrity of the DNA samples was confirmed with 2% agarose gel electrophoresis and the working concentration was adjusted to 30 ng·μL^{-1}.

2.2. Condition of Polymerase Chain Reaction (PCR) and Electrophoresis

PCR was performed according to Nucci et al. [5], except for a lower primers concentration (0.15 μM each) and higher $MgCl_2$ concentration (4 mM for primers Aac04 and Aac12). The amplification cycles were also programmed according to Nucci et al. [5] in a thermal cycler (Applied Biosystem®Verti®cycler). Five of the primers (Aacu07, Aacu10, Aacu12, Aacu26, and Aacu30) were obtained from Nucci et al. [5] and three (Aacu38, Aacu45, and Aacu74) identified from Nucci [27]. The other two markers (Aac04 and Aac12) were obtained from sets of SSR makers originally developed for *Astrocaryum aculeatum* and selected for *A. aculeata* [28] (Table 2). PCR products were denatured in a bromophenol blue dye solution at 95 °C for 5 min in the thermal cycler just before running in 6% polyacrylamide gel electrophoresis in 1xTBE (Tris-Borate-EDTA, Sigma-Aldrich Corporation, St. Louis, MO, USA) buffer solution at 60 W for 1 h and 40 min.

Table 2. Primer pairs of 10 SSR (Simple Sequence Repeats) markers used in the study along with average values obtained for different parameters per locus.

Locus	Forward and Reverse Primer Sequence (5′–3′)	Allele Size (bp)	*A*	H_o	H_e	F	PIC	T_m (°C)	Source *
Aacu07	F: ATCGAAGGCCCTCCAATACT R: AAATAAGGGGACCCTCCAA	153–177	6	0.43	0.48	0.10	0.43	56	a
Aacu10	F: TGCCACATAGAGTGCTTGCT R: CTACCACATCCCCGTGAGTT	168–186	8	0.58	0.69	0.16	0.65	56	a
Aacu12	F: GAATGTGCGTGCTCAAAATG R: AATGCCAAGTGACCAAGTCC	190–202	11	0.57	0.71	0.20	0.67	56	a
Aacu26	F: ACTTGCAGCCCCATATTCAG R: CAGGAACAGAGGCAAGTTC	273–316	9	0.41	0.63	0.35	0.56	56	a
Aacu30	F: TGTGGAAGAAACAGGTCCC R: TCGCCTTGAGAAATTATGGC	148–158	6	0.39	0.43	0.09	0.38	56	a
Aacu38	F: TTCTCAGTTTCGTGCGTGAG R: GGGAGGCATGAGGAATACAA	316–346	6	0.13	0.64	0.80	0.58	56	b
Aacu45	F: CAGACTACCAGGCTTCCAGC R: TCATCATCGCAGCTTGACTC	260–284	5	0.30	0.38	0.21	0.34	56	b
Aacu74	F: TACTGTTGTGCCAAGTCCCA R: GAGCACAAGGGGGATATCAA	278–313	9	0.26	0.45	0.42	0.42	56	b
Aac04	F: GCATTGTCATCTGCAACCAC R: GCAGGGGCCATAAGTCATAA	258–306	8	0.61	0.72	0.15	0.68	60	c
Aac12	F: GCTCTGTAATCTCGGCTTCCT R: TCCAGTTCAAGCTCTCTCAGC	229–247	4	0.06	0.31	0.81	0.27	60	c
Mean		-	7.2	0.37	0.54	0.33	0.50	-	-

A = number of alleles per locus; H_o = observed heterozygosity; H_e = expected heterozygosity; F = inbreeding coefficient; PIC = polymorphic information content; T_m = primer annealing temperature. * Sources of SSR markers: a [5]; b [27]; c [28].

2.3. Polyacrylamide Gel Staining

After electrophoresis, the PCR products were visualized in polyacralamide gels stained with silver nitrate ($AgNO_3$) according to Brito et al. [29]. The gels were immersed and agitated in several coloring steps in different solutions at different concentrations and durations until all allelic bands were totally visible for evaluation. Finally, the stained gels were allowed to dry out in the air and scanned for documentation and DNA fragments were scored as co-dominant alleles for data analyses.

2.4. Data Analyses

Allelic diversity, heterozygosity and polymorphism level of the SSR markers and for each provenance, were estimated from the co-dominant data. Allelic diversity was estimated by quantifying number of alleles per locus (*A*) and total number of alleles per provenance (N_t) [30]. Average observed ($\overline{H}_o$) and expected ($\overline{H}_e$) heterozygosity per provenance was calculated by Equations (1) and (2), respectively. H_o and H_e (Equation (3)) are observed and expected heterozygosity per locus, respectively, in which *a* is number of loci and *pi* is frequency of the i^{th} allele at j^{th} locus [31]. Inbreeding coefficient (F) per locus (or provenance) was estimated from H_o and H_e to determine the level of inbreeding (Equation (4)) [32]. Polymorphic information content (PIC) per locus was calculated by Equation (5) [33], where *pi* and *pj* are frequencies of the i^{th} allele at j^{th} locus. Percentage of polymorphic loci (*P*) per provenance was estimated based on a criterion [34], at allelic frequency of less than 0.95 per locus. The criterion was designated herein as *P*95.

Principal coordinate analysis (PCoA) was performed for graphical dispersion of the accessions on bi-dimensional axes. PCoA was done from the genetic distance between pairs of accessions [35]. Nei's genetic distance between pairs of provenances was computed with Equation (6), where *I* is Nei's genetic identity estimated by Equation (7), in which p_{ijx} and p_{ijy} are frequencies of the i^{th} allele at j^{th} locus of provenance *x* and *y* respectively and *L* stands for the number of loci [36]. A dendogram was constructed from the genetic distance matrix between pairs of provenances using UPGMA—unweighted pair-group method with arithmetic mean. Tocher was also used to form homogenous groups of provenances from Nei's genetic distance matrix.

A Mantel test was applied using the Pearson correlation to test the hypothesis of relationships between genetic and geographic distances among *A. aculeata* accessions obtained from different regional states in Brazil. Analysis of Molecular Variance (AMOVA) was done to estimate the amount of genetic variation among and within the populations/or provenances. Φ-Statistics (Equations (8)–(10)) were computed to test the null hypothesis ($\hat{\sigma}_a^2 = \hat{\sigma}_b^2 = \hat{\sigma}_c^2 = 0$), where $\hat{\sigma}_a^2$, $\hat{\sigma}_b^2$ and $\hat{\sigma}_c^2$ are genetic variations among provenances, among populations, and among individuals, respectively [37,38]. The computed Φ-values were compared against values obtained under 1000 permutations for significance tests. Data analyses were performed using GENES [39] and GenAlex [35] statistical software programs.

$$[\overline{H}_o = \frac{1}{L}\sum_{j=1}^{L} \mathrm{H_{o(}}_{j)}] \quad (1)$$

$$[\overline{H}_e = \frac{1}{L}\sum (1 - \sum_{j=1}^{L} pi^2)] \quad (2)$$

$$[\mathrm{H_e} = 1 - \sum_{j=1}^{a} p_i^2] \quad (3)$$

$$[\mathrm{F} = 1 - \frac{\mathrm{H_o}}{\mathrm{H_e}}] \quad (4)$$

$$[PIC = 1 - \sum_{j=1}^{a} p_i^2 - \sum_{i,j=1}^{a}\sum_{(i\#j)}^{a} p_i^2 p_j^2] \quad (5)$$

$$[Nei\_D = -\ln(I)] \quad (6)$$

$$I = -\ln[\frac{\sum_{j=1}^{L}\sum_{k=1}^{a_j} p_{ijx}p_{ijy}}{\sqrt{\sum_{j=1}^{L}\sum_{k=1}^{a_j} p_{ijx}^2 \sum_{j=1}^{L}\sum_{k=1}^{a_j} p_{ijy}^2}}] \quad (7)$$

$$[\Phi_{CT} = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_T^2}] \quad (8)$$

$$[\Phi_{ST} = \frac{\hat{\sigma}_a^2 + \hat{\sigma}_b^2}{\hat{\sigma}_T^2}] \quad (9)$$

$$[\Phi_{SC} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_c^2}] \quad (10)$$

3. Results and Discussion

3.1. SSR Allelic Polymorphism, Heterozygosity and Informativeness

A total of 72 alleles were detected in the analysis of 192 *A. aculeata* accessions using ten SSR markers. A range of 4 (Aac12)–11 (Aacu12) alleles per locus were obtained with average of 7.2 alleles per locus (Table 2). In other study, using five of the SSRs (Aacu07, Aacu10, Aacu12, Aacu26, and Aacu30), a total of 30 alleles with average of 6 alleles per locus were reported from 43 accessions of *A. aculeata* from São Paulo and Minas Gerais populations [5]. In the present study, those five markers detected 40 alleles with average of 8 alleles per locus (Table 2). Hence, the average number of alleles obtained per locus was higher here than in Nucci et al. [5]. The wider coverage of the genetic materials analyzed from the six geographical states in this study led to yield more alleles per locus and consequently resulted in higher allelic diversity.

H_o ranged from 0.06 (Aac12) to 0.61(Aac04) with average of 0.37 per locus; H_e from 0.31(Aac12) to 0.72 (Aac04) with average of 0.54 per locus; while PIC varied from 0.27 (Aac12) to 0.68 (Aac04) with average of 0.50 per locus (Table 2). According to the criteria set by Bostein et al. [33], the SSR markers used in this study were informative and polymorphic to characterize the germplasm accessions (or populations) in *A. aculeata*. SSR markers are classified as informative when PIC > 0.50, reasonably informative (0.25 < PIC < 0.50) or less informative (PIC < 0.25). The number of alleles, H_o, H_e, and PIC obtained in this study was nearly similar to that of Nucci et al. [5] (average H_o = 0.27; H_e = 0.57; and PIC = 0.54), who first characterized SSR markers for *A. aculeata*. This could explain that the allelic frequencies of the loci have not changed significantly over generations. Although we could not trace precisely the germplasm analyzed by Nucci et al. [5], we may speculate that some similar accessions might be analyzed in the present study, probably from Minas Gerais and São Paulo. Besides, since macaw palm is still in the wild, under certain modes of random matting systems governing the rule of Hardy Weinberg Equilibrium in the absence of selection, mutation, and migration [32], allelic frequency in a given population could remain unchanged over generations. However, the low proportion of the average H_o (<50%) could be caused by involvement of inbreeding or crossing between genetically related individuals. This was explained by a positive average inbreeding coefficient (F > 0) obtained per locus and per provenance confirming the presence of heterozygote deficiency in all the provenances studied (Table 2, Table 3). Although *A. aculeata* has a mixed mating system [2,4,15], its monoecious inflorescences could favor selfing or crossing between genetically related individuals that could reduce the proportion of heterozygotes in its progenies. According to Hartl and Clark [32], a positive inbreeding coefficient indicates predominance of inbreeding and values close to zero are expected under random mating, and negative values indicate an excess of heterozygote due to negative assortative (disassortative) mating or selection.

Table 3. Average estimates of genetic diversity parameters for ten *Acrocomia aculeata* provenances based on ten polymorphic SSR markers.

Provenances	N_t	P_a	N_a	$\overline{H}_o$	$\overline{H}_e$	*F*	P_{95}
PA	32	0.44	3.20	0.46	0.54	0.15	90
PE	29	0.40	2.90	0.34	0.45	0.24	90
PB	28	0.38	2.80	0.35	0.49	0.29	100
SP	39	0.53	3.90	0.32	0.56	0.42	90
NMG	52	0.71	5.20	0.32	0.62	0.48	90
SMG	39	0.53	3.90	0.36	0.52	0.32	80
CMG	42	0.58	4.20	0.29	0.56	0.48	90
EMG	43	0.59	4.30	0.30	0.51	0.40	90
WMG	55	0.75	5.10	0.28	0.63	0.55	100
MS	57	0.78	5.70	0.46	0.59	0.22	100
Mean	42	-	4.12	0.35	0.55	0.36	92

N_t = total number of alleles per provenance; P_a = proportion of alleles per provenance; N_a = average number of alleles per provenance; $\overline{H}_o$ = average observed heterozygosity per provenance; $\overline{H}_e$ = average expected heterozygosity per provenance; F = inbreeding coefficient; *P* (%) = percentage of polymorphic loci with a criterion (P_{95}) mentioned in materials and methods. Provenances included: PA = Pará, PE = Pernambuco, PB = Paraiba, SP = São Paulo, NMG = North Minas Gerais, SMG = South Minas Gerais, CMG = Central Minas Gerais, EMG = East Minas Gerais, WMG = West Minas Gerais, and MS = Mato Grosso do Sul.

3.2. Genetic Diversity

Based on the ten SSR markers, N_t detected per provenance varied from 28 (PB) to 57 (MS), hence, N_a ranged from 2.8 to 5.7 respectively (Table 3). Consequently, the highest proportion of alleles (78%) was obtained in MS and the least (38%) in PB. The variation in the proportion of alleles (or allelic diversity) among the provenances attested the presence of genetic diversity of *A. aculeata* in the genebank. Besides, high average *P* (92%) was obtained per provenance with a range of 80%–100% based on the criterion proposed by Cole [34] (Table 3). The polymorphism level reported here was much higher than that obtained by Oliveira et al. [14] using Random Amplified Polymorphic DNA (RAPD) markers (*P* = 79%), analyzing *Acrocomia aculeata* from natural populations. This could be explained by the higher polymorphic level of SSRs compared with RAPD markers. Hence, the results attested the SSRs are highly polymorphic molecular markers to study genetic variability in *A. aculeata*. Moreover, SSRs are co-dominant while RAPD are dominant.

Genetically different groups of accessions were formed with different methods of grouping. Three distinct groups were formed using PCoA on the first two coordinates, explaining 31.5% and 20% of the total variability, respectively (Figure 2).

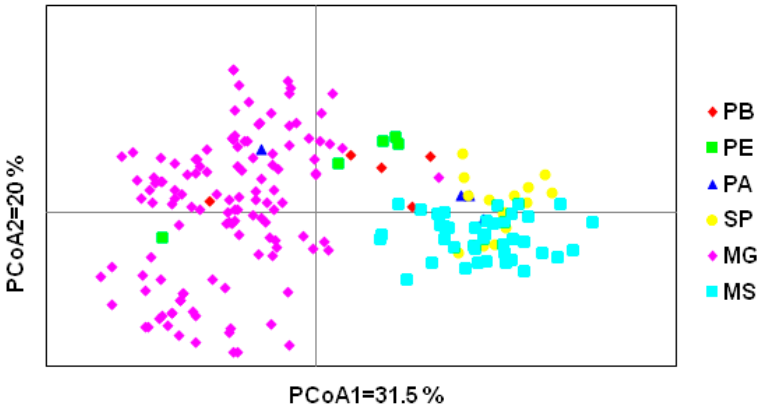

Figure 2. Graphical dispersion of 192 individuals using Principal Coordinate Analysis (PCoA) showing grouping of the accessions into different distinct groups. Provenances include PA = Pará, PE = Pernambuco; PB = Paraiba; SP = São Paulo; MG = Minas Gerais (containing five provenances: NMG, SMG, CMG, EMG, and WMG shown as one big group); MS = Mato Grosso do Sul.

Collections from MG, represented the largest group in the study composed of five provenances (NMG, SMG, CMG, EMG, and WMG) (Table 1), clearly separated from the rest of the groups. Likewise, accessions from MS and SP formed the second group, while the third group composed of accessions from the regional States of PB and PE (Figure 2). The PCoA depicted the complete distinctness of MG collections, the genetic similarity between accessions from PA, SP, and MS and the genetic relatedness between PB and PE collections. Hence, formation of the different groups reiterated the presence of diverse genetic variability among the germplasm collections in the genebank.

Establishment of the three major groups using PCoA was consistent with the other two methods used in our analyses (UPGMA, Figure 3 and Tocher, Table 4). Using UPGMA, at 70% of dissimilarity, three hierarchical groups of *A. aculeata* provenances were established. Similar to the PCoA method, with UPGMA, MG provenances formed the first group and PA, PE, and PB established the second group, representing collections from the northern part of the country, while SP and MS formed the third distinct group, reaffirming the genetic relatedness between collections from the two neighboring geographical states (Figure 3). However, with the optimization method (Tocher), one additional group was formed due to the separation of PA provenance from the third group and formed the fourth independent group (sub-group) (Table 4). This is more likely because, unlike UPGMA, Tocher considers more similar groups (using least genetic distances) in each stage of group formation to establish new homogenous groups based on their genetic similarities [30]. Hence, there is a possibility to establish an additional group (sub-group) with the method of Tocher at the last stage of grouping. This could be confirmed from the dendogram (Figure 3) that, at a low percentage of dissimilarity (50%–55%), using local criterion, PA remained as an independent and separate group, reaffirming its relative genetic distance from the two neighboring provenances (PB and PE). However, the least genetic distance (Table 4, $D_{3,4}$ = 0.50), between Groups 3 (PB and PE) and 4 (PA), elucidated the genetic relatedness among collections from the three geographical States (PA, PE, and PB), as evidenced by the PCoA (Figure 2) and UPGMA (Figure 3), which represented collections from the northern part of Brazil.

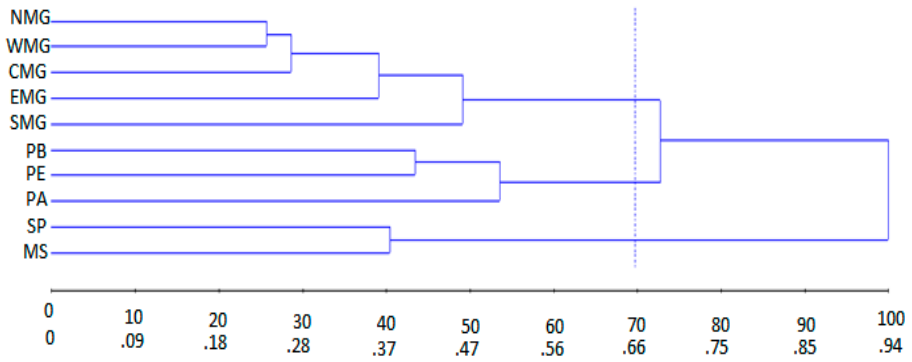

Figure 3. UPGMA dendogram of ten *Acrocomia aculeata* provenances constructed from genetic distance [36]. Provenances: PA = Pará; PE = Pernambuco; PB = Paraiba; SP = São Paulo; NMG = North Minas Gerais; SMG = South Minas Gerais; CMG = Central Minas Gerais; EMG = East Minas Gerais; WMG = West Minas Gerais; MS = Mato Grosso do Sul. The first row numbers from 0–100 are percentages of dissimilarity; and the second row numbers from 0–0.94 are levels of fusion (average genetic distance) corresponding to percentage of dissimilarity.

Although the distinctness of the three major groups was confirmed by different methods, our results in relation to the genetic relatedness among the groups showed inconsistency with the hypothesis that collections from closer geographical regions are genetically more similar than distant ones and vice versa. This inconsistency primarily came from the results of the complete dissimilarity between the first (MG) and the second group (SP and MS), composed of provenances from neighboring geographical states (Figure 1). Secondly, at a higher percentage of dissimilarity (above 70%), using

UPGMA, collections from MG State were genetically closer to that of the distant States of PB, PE, and PA than its neighbor States (SP and MS) (Figure 3). This scenario was also explained by the mean inter-group genetic distances using Tocher (Table 4). High average genetic distance ($D_{1,2}$ = 1.05) was obtained between the first (MG) and the second group (MS and SP) than between the first and the third group (PE and PB, $D_{1,3}$ = 0.64) and the first and the fourth group (PA, $D_{1,4}$ = 0.77). Moreover, this result was confirmed with a Mantel test showing a weak correlation (r = 0.07) between genetic and geographic distances among the provenances studied (Figure 4). Similar results were also reported in other species, such as Italian red clover [40], globe artichoke [41], and Italian emmer wheat [42].

Table 4. Grouping of ten *Acrocomia aculeata* provenances using the method of Tocher. With Tocher, homogeneous groups are formed, as it uses least genetic distance at each stage of group formation. Hence, mean intra-group distance is always less than mean inter-group distance.

Group	Provenances	Mean Intra-Group Distance	Mean Inter-Group Distance
1	NMG WMG CMG EMG SMG	0.37	$D_{1,2}$ = 1.05; $D_{1,3}$ = 0.64; $D_{1,4}$ = 0.77
2	SP MS	0.38	$D_{2,3}$ = 0.80; $D_{2,4}$ = 0.67
3	PB PE	0.41	$D_{3,4}$ = 0.50
4	PA	-	-

Provences: PA = Pará; PE = Pernambuco; PB = Paraiba; SP = São Paulo; NMG = North Minas Gerais; SMG = South Minas Gerais; CMG = Central Minas Gerais; EMG = East Minas Gerais; WMG = West Minas Gerais; MS = Mato Grosso do Sul. *D* is the mean genetic distance between each pair of the groups.

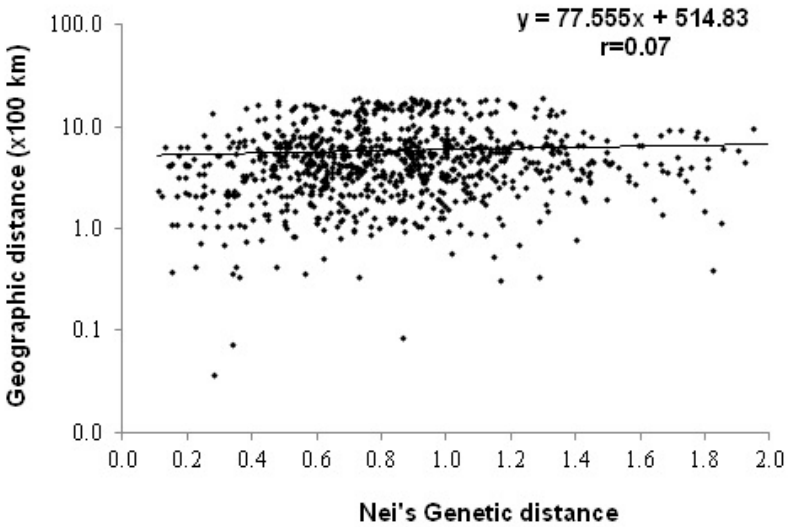

Figure 4. Mantel test revealing weak correlation between genetic and geographical distances in *Acrocomia aculeata* accessions collected from different geographical states in Brazil.

3.3. Analysis of Molecular Variance (AMOVA)

From AMOVA, more genetic variation within populations (48.2%) than among populations (36.5%) was obtained (Table 5). These variations were statistically significant ($p < 0.01$) in that the Φ-Statistics values obtained from estimated variances were higher than those values obtained under 1000 permutations. Previous studies in *A. aculeata* reported similar results, showing higher genetic variability within population than among populations [14,15]. The higher genetic diversity within population is a result of the mixed mating system in *A. aculeata* and the involvement of metapopulation structure in natural populations [15]. Metapopulation structure of species is caused by fragmentation of lands and creates spatially separated populations which interact at some level. However, this could favor genetic drift and restricted gene flow, which cause decrease in genetic diversity within a population especially in cross-pollinated species [43]. However, in our case, the mixed mating nature

of *A. aculeata* kept the genetic variability higher within population than among population. Similar works also reported higher genetic variation within populations in some palms and different tree species, such as Canarian endemic palm tree *Phoenix canariensis* [43], *Populus tremuloides* Michx [44], *Digitalis minor* [45], *Piper hispidinervum* [46], and *Trichilia pallida* [47].

Table 5. AMOVA for 178 *Acrocomia aculeata* accessions based on ten polymorphic loci.

Source of Variance	df	Variance	%	Φ-Statistics	Sig.
Among provenance	6	0.1217	15.31	Φ_{CT} = 0.4307	*
Among populations/provenance	31	0.2901	36.47	Φ_{SC} = 0.1531	*
Among individuals/population	140	0.3835	48.22	Φ_{ST} = 0.5178	*

* Significant at $p < 0.01$. Φ-Statistics are compared with values obtained from 1000 permutations. AMOVA performed using 178 individuals of 38 populations from seven provenances (regions) including SP = São Paulo; NMG = North Minas Gerais; SMG = south Minas Gerais; CMG = Central Minas Gerais; EMG = East Minas Gerais; WMG = West Minas Gerais; MS = Mato Grosso do Sul. Parã, Pernambuco and Paraiba not included in the analysis since they have only one population.

4. Conclusions

SSRs markers resulted in being very useful and efficient in characterizing *A. aculeata* germplasm. The SSR markers used are polymorphic among the *A. aculeata* accessions analyzed in this study and established different groups based on their genetic distances. This would facilitate the process of identifying, grouping and selecting genotypes during pre-breeding. Since *A. aculeata* is perennial and has a long cycle of growth, the use of SSR markers will accelerate the process of selecting genotypes at early stages, which will save time and resources. Moreover, the high genetic variations within population underlines the importance of having many genotypes in the genebank. The result will also help to minimize problems of replicates of genetic materials in the genebank and maintain genetic variability for sustainable use for future breeding programs. Further studies are necessary to investigate why genetic distance among populations did not couple with geographic distance, which will help in finding out the nature of gene flow and population structure of *A. aculeata* in Brazil.

Acknowledgments: The authors would like to acknowledge Carlos Nick for his technical help in obtaining leaf samples from Macaúba Active Genebank (BAG–Macaúba respository nº: 084/2013/CGEN/MMA) situated in the experimental farm of the Universidade Federal de Viçosa in the municipality of Arapongaa, Minas Gerais, Brazil. We are also grateful to Dra. Eveline T. Caixeta, Dra. Kacilda N. Kuki, Dra. Telma, Renata D. Freitas and Éder Lanes for their unconditional assistance in Plant Biotechnology Laboratory, Universidade Federal de Viçosa. This work was financed by Petróleo Brasileiro S.A (Petrobras); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and The academy of sciences for the developing world (TWAS).

Author Contributions: Fekadu G. Mengistu initiated the project with the help of Sérgio Y. Motoike. Sérgio originally established the genebank in the municipality of Araponga, Universidade Federal de Viçosa, Brazil, where we obtained the genetic materials for this study. Fekadu did the sample collection, DNA extraction, PCR assays, data analysis, manuscript writing and language editing. Cosme D. Cruz assisted in the data analysis with important inputs especially in the diversity part.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abreu, S.I.; Carvalho, C.R.; Carvalho, G.M.A.; Motoike, S.Y. 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*, 149–155. [CrossRef]
2. Scariot, A.O.; Lleras, E.; Hay, J.D. Reproductive biology of the palm *Acrocomia aculeata* in Central Brazil. *Biotropica* **1991**, *23*, 12–22. [CrossRef]
3. Scariot, A.O.; Lleras, E.; Hay, J.D. Flowering and fruiting phenologies of the palm *Acrocomia aculeata*: Patterns and consequences. *Biotropica* **1995**, *27*, 168–173. [CrossRef]

4. Abreu, A.G.; Priolli, R.H.; Azevedo-Filho, J.A.; Nucci, S.M.; Zucchi, M.I.; Coelho, R.M.; Colombo, C.A. The genetic structure and mating system of *Acrocomia aculeata* (Arecaceae). *Genet. Mol. Biol.* **2012**, *35*, 119–121. [CrossRef] [PubMed]
5. Nucci, S.M.; Azevedo-Filho, A.; Colomo, A.C.; Priolli, G.H.R.; Coelho, M.R.; Mata, T.L.; Zucchi, I.M. Development and characterization of microsatellites markers from the macaw. *Mol. Ecol. Resour.* **2008**, *8*, 224–226. [CrossRef] [PubMed]
6. Lanes, E.C.M.; Costa, P.M.A.; Motoike, S.Y. Alternative fuels: Brazil promotes aviation biofuels. *Nature* **2014**, *511*, 31. [CrossRef] [PubMed]
7. Joshua; Tickell, K. *From the Fryer to the Fuel Tank: The Complete Guide to Using Vegetable Oil as an Alternative Fuel*, 2nd ed.; Green Teach Publishing: Sarasota, FL, USA, 1999. Available online: http://www3.zetatalk.com/docs/Biodiesel/From_Fryer_To_Fuel_Tank_1999.pdf (accessed on 19 September 2014).
8. Moura, E.F.; Motoike, S.Y.; Ventrella, M.C.; Sá Junior, A.Q.; Carvalho, M. Somatic embryogenesis in macaw palm (*Acrocomia aculeata*) from zygotic embryos. *Sci. Hortic.* **2009**, *119*, 447–454. [CrossRef]
9. Fortes, I.C.P.; Baugh, P.J. Study of analytical on-line pyrolysis of oils from Macauba fruit (*Acrocomia sclerocarpa* M.) via GC/MS. *J. Braz. Chem. Soc.* **1999**, *10*, 469–477. [CrossRef]
10. Bora, P.S.; Rocha, R.V.M. Macaíba palm: Fatty and amino acids composition of fruits. *Cienc. Tecnol. Aliment.* **2004**, *4*, 158–162. [CrossRef]
11. Hiane, P.A.; Ramos-Filho, M.M.; Ramos, M.I.L.; Macedo, M.L.R. Bocaiúva, *Acrocomia aculeata* (Jacq.) Lodd., pulp and kernel oils: Characterization and fatty acid composition. *Braz. J. Food Technol.* **2005**, *8*, 256–259.
12. Motoike, S.Y.; Kuki, K.N. The potential of macaw palm (*Acrocomia aculeata*) as source of biodiesel in Brazil. *Int. Rev. Chem. Eng.* **2009**, *1*, 632–635.
13. Teixeira, L.C. Potencialidades de oleaginosas para produção de biodiesel. In *Produção de Oleaginosas para Bioiesel*; Lacerda, V., Ed.; Informe Agropecuário: Belo Horizonte, Brazil, 2005; pp. 18–27.
14. Oliveira, D.A.; Júnior-Melo, A.F.; Brandão, M.M.; Rodrigues, L.A.; Menezes, E.V.; Ferreira, P.R.B. Genetic diversity in populations of *Acrocomia aculeata* (Arecaceae) in the northern region of Minas Gerais. *Genet. Mol. Res.* **2012**. [CrossRef] [PubMed]
15. Colombo, C.A.; Nucci, S.M.; Priolli, R.H.G.; Zucchi, M.I.; Carhalho, C.R.L.; Chrfl, B.L.; Siqueira, W.G.; Azvedo-Filho, J.A. Genetic Diversity of Macaw Palm (*Acrocomia aculeata*) by Microsatellite Markers. In Proceedings of the International Plant and Animal Genome Conference, San Diego, CA, USA, 11–16 January 2013; Available online: http://www.intlpag.org (accessed on 7 October 2013).
16. Faleiro, F.G.; Costa, A.M.; Karia, C.T.; Andrade, R.P.; Junqueira, N.T.V.; Pereira, A.V.; Pereira, E.B.; Sano, S.M. Molecular Markers and Geographic Information Systems as a Tool to Study Native Plant Species in the Brasilian Savannas, 2008. Available online: http://www.cpac.embrapa.br (accessed on 20 July 2012).
17. Ribeiro, L.M.; Souza, P.P.; Rodrigues, A.G.; Oliveira, T.G.S.; Garcia, Q.S. Overcoming dormancy in macaw palm diaspores, a tropical species with potential for use as bio-fuel. *Seed Sci. Technol.* **2011**, *39*, 303–317. [CrossRef]
18. Litt, M.; Luty, J.M. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **1989**, *44*, 397–401. [PubMed]
19. Tautz, D.; Trick, M.; Dover, G.A. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **1986**, *322*, 652–656. [CrossRef] [PubMed]
20. Ferreira, M.E. Molecular Analysis of Gene Banks for Sustainable Conservation and Increased Use of Crop Genetic Resources: The Role of Biotechnology, 2005. Available online: http://www.fao.org/biotech (accessed on 5 April 2012).
21. Mondini, L.; Noorani, A.; Pagnotta, M.A. Assessing plant genetic diversity by molecular tools. *Diversity* **2009**, *1*, 19–35. [CrossRef]
22. Mohan, M.; Nair, S.; Bhagwat, A.; Krishna, T.G.; Yano, M.; Bhatia, C.R.; Sasaki, T. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breed.* **1997**, *3*, 87–103. [CrossRef]
23. Agarwal, M.; Shivastava, N.; Padh, H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* **2008**. [CrossRef] [PubMed]
24. Motoike, S.Y.; Lopes, F.A.; Sá Júnior, A.Q.; Carvalho, M.; Oliveira, M.A.R. Processo de Germinação e Produção de Sementes Pré-Germinadas de Palmeiras do Gênero *Acrocomia*. Patente: Submetido à Lei de Patentes. Protocolo INPI: 014070005335; 2007. Available online: http://www.inpi.gov.br/ (accessed on 21 July 2014).
25. Doyle, J.J.; Doyle, J.L. Isolation of plant DNA from fresh tissue. *Focus* **1990**, *12*, 13–15.

26. Lanes, E.C.M.; Nick, C; Kuki, K.N.; Freitas, R.D.; Motoike, S.Y. Genomic DNA isolation of *Acrocomia aculeata* (*Arecaceae*) from leaf and stipe tissue samples for PCR analysis. *Genet. Mol. Res.* **2013**, *12*, 3905–3911.
27. Nucci, S.M. Development, Characterization and Analyses of the Use of Microsatellite Markers in Macaw Palm Population Genetics. Master's Thesis, University of Sâo Paulo, São Paulo, Brazil, 2007.
28. Mengistu, F.G.; Motoike, S.Y.; Caixeta, E.T.; Cruz, C.D.; Kuki, K.N. Cross-species amplification and characterization of new microsatellite markers for the macaw palm, *Acrocomia aculeata* (*Arecaceae*). *Plant Genet. Resour.* **2015**. [CrossRef]
29. Brito, G.G.; Caixeta, E.T.; Gallina, P.A.; Zambolim, E.M.; Zambolim, L.; Diola, V.; Loureiro, M.E. Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. *Euphytica* **2010**, *173*, 255–264. [CrossRef]
30. Cruz, C.D.; Ferreira, F.M.; Pessoni, L.A. *Biometria Aplicada ao Estudo da Diversidade Genética*, 1st ed.; UFV-Universidade Federal de Viçosa: Viçosa, Brazil, 2011.
31. Nei, M. Estimation of average heterozygosity and genetic distances from a small number of individuals. *Genetics* **1978**, *89*, 583–590. [PubMed]
32. Hartl, D.L.; Clark, A.G. *Principles of Population Genetics*, 3rd ed.; Sinauer Associates: Sunderland, MA, USA, 1997.
33. Bostein, D.; Whiter, L.; Skolnick, M.; Davis, R.W. Construction of a Genetic Linkage Map in Man using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331.
34. Cole, C.T. Genetic variation in rare and common plants. *Annu. Rev. Ecol. Syst.* **2003**, *34*, 213–237. [CrossRef]
35. Peakall, R.; Smouse, P.E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295.
36. Nei, M. Genetic distance between populations. *Am. Nat.* **1972**, *106*, 238–292. [CrossRef]
37. Excoffier, L.; Smouse, P.E.; Quattro, J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **1992**, *131*, 479–491. [PubMed]
38. Excoffier, L. Analysis of population subdivision. In *Handbook of Statistical Genetics*; Balding, D.J., Bishop, M., Cannings, C., Eds.; John Wiley Sons: London, UK, 2001; pp. 271–307.
39. Cruz, C.D. *Programa Genes: Biometria*; UFV-Universidade Federal de Viçosa: Viçosa, Brazil, 2006.
40. Pagnotta, M.A.; Annicchiarico, P.; Farina, A.; Proietti, S. Characterizing the moleular and morphophysiological diversity of Italian red clover. *Euphytica* **2011**, *179*, 393–404. [CrossRef]
41. Crinò, P.; Tavazza, R.; Rey Munoz, N.A.; Nisini, P.T.; Saccardo, F.; Ancora, G.; Pagnotta, M.A. Recovery, morphological and molecular characterization of globe artichoke 'Romanesco' landraces. *Genet. Resour. Crop. Evol.* **2008**, *55*, 823–833. [CrossRef]
42. Pagnotta, M.A.; Mondini, L.; Atallah, M.F. Morphologocal and molcular characterization of Italian emmer wheat accessions. *Euphytica* **2005**, *146*, 29–37. [CrossRef]
43. Pérez-González, M.A.; Caujapé-Castells, J.; Sosa, P.A. Allozyme variation and structure of the Canarian endemic palm tree *Phoenix canariensis* (Arecaceae): Implications for conservation. *Heredity* **2004**, *93*, 307–315. [CrossRef] [PubMed]
44. Yeh, F.C.; Chong, D.K.; Yang, R.C. RAPD variation within and among natural populations of trembling aspen (*Populus tremuloides* Michx). *J. Hered.* **1995**, *86*, 454–460. [PubMed]
45. Sales, E.; Nebauer, S.G.; Mus, M.; Segura, J. Population genetic study in the Balearic endemic plant species *Digitalis minor* (Scrophulariaceae) using RAPD markers. *Am. J. Bot.* **2001**, *88*, 1750–1759. [CrossRef] [PubMed]
46. Wadt, L.H. Estruturagenética de Populações Naturais de Pimenta Longa (*Piper hispidinervum*), Visandoseuuso e CoservaçãO. Ph.D. Thesis, Universidade de Federal de São Paulo, São Paulo, Brazil, 2001.
47. Zimback, L.; Mori, E.S.; Kagyama, P.Y.; Veiga, R.F.A.; Junior, J.R.S.M. Estrutura genética de populações de *Trichilia pallida* Swartz (Meliaceae) por marcadores RAPD. *Sci. For.* **2004**, *65*, 114–119.