CROP BREEDING AND APPLIED BIOTECHNOLOGY

ARTICLE

Analysis of the genetic structure and diversity of a Brazilian macadamia nut (*Macadamia integrifolia*) germplasm

Graciela da Rocha Sobierajski[1*], Gabriel Constantino Blain[1], Adna Cristina Barbosa Sousa[2], Letícia Jungmann Cançado[3], Guilherme Pereira[4], Anete de Souza[5] and Antonio Augusto F Garcia[6]

Crop Breeding and Applied Biotechnology
22(3): e41382231, 2022
Brazilian Society of Plant Breeding.
Printed in Brazil
http://dx.doi.org/10.1590/1984-70332022v22n3a24

Abstract: *Macadamia is a nut tree native to Australian rainforest. Due to the small wild populations and the economic interest, studies assessing genetic information are fundamental for use and conservation of this species. Therefore, the present study aimed to assess the structure and genetic diversity of 28 cultivars of macadamia originated from United States, Australia and Brazil, introduced or developed by the Agronomic Institute's breeding program, using 29 new microsatellite loci. The microsatellite loci showed high genetic diversity (3.65 alleles per locus on average). Our results also suggested the existence of no genetic structure between cultivars, regardless of their geographic origin. The modified Rogers' distances between cultivars ranged from 0.227 to 0.671. Despite the lack of information about genealogy, the cultivars from this Brazilian germplasm showed moderate genetic diversity, so they can be used as parents in future crosses.*

Keywords: *Microsatellite, nut tree, cross-amplification, molecular marker, polymorphis*

***Corresponding author:**
E-mail: graciela.rocha@sp.gov.br
ORCID: 0000-0002-7211-9268

Received: 08 February 2022
Accepted: 25 June 2022
Published: 28 July 2022

[1] Instituto Agronômico de Campinas, Avenida Barão de Itapura, 1481, 13075-630, Botafogo, Campinas, SP, Brazil
[2] Universidade Federal da Paraíba, Cidade Universitaria, s/n, 58051-900, Castelo Branco, João Pessoa, PB, Brazil
[3] Embrapa Agroenergia, Parque Estação Biológica, s/n, 70770-901, Asa Norte, Brasília, DF, Brazil
[4] Universidade Federal de Viçosa, Avenida Peter Henry Rolfs, s/n, 36570-900, Campus Universitário, Viçosa, MG, Brazil
[5] Universidade Estadual de Campinas, Cidade Universitária Zeferino Vaz, s/n, 13083-970, Barão Geraldo, Campinas, SP, Brazil
[6] Escola Superior de Agricultura "Luiz de Queiroz", USP, Avenida Pádua Dias, 11, 13418-900, Agronomia, Piracicaba, SP, Brazil

INTRODUCTION

Macadamia integrifolia Maiden & Betche (Proteaceae) is a nut tree native to Australia and occurs wildly in the southeast of Queensland and northern New South Wales States (Topp et al. 2019). Currently, only 1% of the Australian rainforest is preserved and restricted into fragments (Sobierajski 2019). This scenario emphasizes the importance of well-characterized *ex situ* germplasms (Busanello et al. 2020, Guan et al. 2020, Carena 2021). The genetic variation existing in germplasms can provide genes for resistance to pests, diseases and adaptation to climate changes (Topp et al. 2019, Carena 2021).

Despite the Australian origin of macadamia, its first breeding program was established by the Hawaii Agricultural Experiment Station (HAES), Hawaii/United States (Aradhya et al. 1998, Hardner et al. 2019, Nock et al. 2019, Topp et al. 2019). Since the Hawaiian germplasm was established from seeds of a small number of trees, the narrow genetic base may lead to limitations for cultivar development (Aradhya et al. 1998, Steiger et al. 2003). This fact highlights the importance of germplasm characterization worldwide. Such characterization may shed light on the relationship between cultivars and

wild individuals (Aradhya et al. 1998, Steiger et al. 2003).

Several genetic marker techniques have been used as part of strategies for conservation and genetic breeding of macadamia, e.g., Amplified Fragment Length Polymorphisms - AFLP (Steiger et al. 2003), and Randomly amplified DNA fingerprinting - RAF (Peace et al. 2005). The common point of these studies is the use of non-specific genetic markers, genotyping HAES and Australian cultivars, and wild species. Aradhya et al. (1998) also carried out a molecular characterization study, but they considered only four cultivars released by the Agronomic Institute of Campinas (IAC) in Brazil. Schmidt et al. (2006) and Nock et al. (2014) developed microsatellite (SSR) specific markers for macadamia. These last two studies considered no cultivar from IAC germplasm.

In Brazil, seeds of macadamia were introduced in 1931 from cultivars developed by HAES. The only macadamia breeding program in Brazil was established by the IAC in 1940 (Toledo Piza et al. 2019). Up to this date, this program released 15 cultivars with high productivity and large nut size. Unfortunately, there is a lack of information on the pedigree and genetic diversity of these cultivars. Despite this drawback, Brazil ranked 7th in the world standings with production of 6,000 tons of nut in shell in 2015, corresponding to 3.75% of world production (Toledo Piza et al. 2019). The major producing regions are situated in Southeastern Brazil. However, the agro-climatic zoning for macadamia in Brazil indicates that regions situated in central-west and south Brazil may also be regarded as fit zones for establishment of new plantations (Sobierajski 2019).

A key point to macadamia production is the development of cultivars well-adapted to local conditions. For instance, new macadamia cultivars have increased producers' profits in Australia (Topp et al. 2017, Kern et al. 2022). Therefore, it is essential to characterize the germplasm from a molecular-genetic point of view. Considering the genetic diversity as a key-factor for assessing differences among individuals of the same species (He and Li 2020, Carena 2021), the hypotheses of the present study are that the Brazilian macadamia germplasm shows genetic diversity among cultivars, which may be used in the IAC's macadamia breeding program. Aiming at generating new information on macadamia germplasm in Brazil, we developed twenty-nine *M. integrifolia* specific SSR loci and assessed the genetic diversity and structure of 28 cultivars of macadamia. Additionally, these loci were tested for cross-amplifications in *Grevillea robusta* and *G. banksii*. These two species were added to this study because they are from the same botanical family as *Macadamia* sp., are economic relevant and there is no specific SSR marker developed for *G. banksii*.

MATERIAL AND METHODS

Genomic DNA from macadamia cultivar IAC 9-20, which has high yield and nut quality, was extracted from fresh leaves using Cetyltrimethylammonium bromide (CTAB) extraction method (Inglis et al. 2018). This DNA was used to construct the enriched genomic library according to the methodology described in Billotte et al. (1999). This methodology includes the following stages: DNA digestion with the *RsaI* enzyme; Rsa21 and Rsa25 adapters linkage; amplification, purification and selection of fragments containing microsatellite regions; cloning of fragments by pGEM-T vector in *Escherichia coli* (Promega, WI, USA); and sequencing using SP6 and T7 primers promoters. Sequences from the recombinant colonies were assembled and edited using Seqman software (DNA Star 2000). The Simple Sequence Repeat Identification Tool (SSRIT) was used to identify SSRs present in all non-redundant sequences (Temnykh et al. 2001). The Primer Select software (DNA Star 2000) was used to design the primer pairs. The DNA sequences were submitted to GenBank (Table 1), from National Center for Biotechnology Information – NCBI (http://www.ncbi.nlm.nih.gov).

Twenty-eight genotypes of macadamia from the IAC germplasm and producers in Dois Córregos, São Paulo State, Brazil, were assessed with regard to their SSR polymorphisms. The DNA sampling was composed of 11 cultivars developed by the Agronomic Institute, five cultivars developed by Hawaii Agricultural Experiment Station, four cultivars developed by the Hidden Valley Plantation (Australia), and two producers' selections (Aloha and Flor Roxa – Supplemental material 1). These cultivars were chosen because of their high acceptance by the Brazilian producers or because they show features of interest for breeding program. In addition, six genotypes showing morphological divergence from those of the IAC germplasm were included to confirm their genetic origin. The cross amplification was tested in two species of silver oak (*Grevillea robusta* and *G. banksii*), collected from the University of Campinas (UNICAMP) campus, for all polymorphic loci. The DNA extraction of all samples was carried out using Cetyltrimethylammonium bromide (CTAB) method (Inglis et al. 2018).

Table 1. Description of 29 microsatellite loci developed for *Macadamia integrifolia* and cross-species amplification (+ successful, - unsuccessful), in *G. robusta* (a) and *G. banksii* (b)

SSR locus	GenBank accession no.	Repeat motif	T_a (ºC)	Primer sequences (5' - 3')	N	Expected allele size (pb)	Transferability	
							a	b
MInt001	JX137062	$(AG)_{27}$	60°	F: CAGCCCATGAAATAACAAT R: CATAATGGTGCAGTGATAGTA	29	271	+	+
MInt005	JX137064	$(TG)_9(GA)_{16}$	TD	F: CCTGGAAGAGCTGACCTAAAA R: CACAATCACGACCAGTAAACAA	30	274	+	+
MInt007	JX137065	$(TG)_6(AG)_4$	TD	F: TGGTTTATATCTCCTCCTTGTT R: ATCTCCACCCATTCACTTCT	28	107	+	+
MInt009	JX137066	$(AG)_{19}$	TD	F: ATCTGCCACTCCATTTTA R: GACTTCTGTTTCCCTTCTA	30	170	+	+
MInt011	JX137067	$(AG)_{29}$	TD	F: GCCCAAAATAGAGTCAAAGTC R: TTAAGCGGCAAATCAAGA	30	263	+	+
MInt013	JX137068	$(TC)_{23}$	TD	F: TAAGTTTGAAGACAGTGAGTGG R: TGAAAGTGTTAAGAGGCAGAA	29	97	+	+
MInt015	JX137069	$(TC)_{23}$	62°	F: AACCTGATGACACCGCCTTCTC R: CTCTCCTCCCGTTTGCTATTCA	27	168	-	-
MInt019	JX137071	$(AG)_{11}(AG)_3$	TD	F: ATTTCGGTTGGGGTTCTCC R: CCCATTTGCTCTTTCATTTCAT	30	203	+	+
MInt025	JX137074	$(TC)_8$	TD	F: ACGTGCGTTAGATGTTTG R: TAATTTTGGTTGTAGGGGTAG	30	212	+	+
MInt029	JX137076	$(GA)_{21}(AG)_3$	TD	F: CACTATTTGGGGCTCCTCAT R: ACCCATCCCTAACTCACTCTCA	29	209	+	+
MInt031	JX137077	$(GT)_{17}(GC)_3$	TD	F: ATAATCAGTAAACAACGCTCAA R: TCCCACAATTCCAGACCA	30	296	+	+
MInt033	JX137078	$(TG)_3$	TD	F: CAACTTAACTTCAACGGGTAG R: AAAATTTCATCCTCTTCACAA	28	229	-	-
MInt035	JX137079	$(CT)_3$	TD	F: GTCACCGGTCCGATTCTGT R: ATGGCTTGGCTTCCTTGTTT	30	283	+	+
MInt037	JX137080	$(CT)_4$	45°	F: AATCGGATACTTCTTTCTAA R: ACCGTATGCTTCTGTTTT	28	208	-	+
MInt039	JX137081	$(TG)_3(TG)_4$	TD	F: GCTAACTACCATACGAAACTGC R: TAAATTGAACTTGCTTGCTCTT	29	212	+	-
MInt041	JX137082	$(AG)_4$	TD	F: TCCCAAGATGACCAAAGAAG R: AACTCAAAGAAAAAGGCAAATC	30	224	+	+
MInt045	JX137084	$(TG)_3$	TD	F: TAGACCAGTAAAACATCAGG R: AAAGCTTTACCACCAACTAT	30	212	+	+
MInt047	JX137085	$(CA)_9(AT)_3$ $(TA)_3$	60°	F: ACCCTGTCACCTTTGTCA R: GTTTTCGTGTCTTCTCTTTATG	16	181	-	-
MInt051	JX137087	$(GA)_3$	TD	F: TGCTCACCTCAAATCGTAG R: TTAAAGAGTGGGTGAGAAGAGT	30	181	+	+
MInt055	JX137089	$(TGG)_3$	TD	F: TCGGAAGAAAAATGTCAC R: ACTCAAATGCTCTCAATCA	30	162	+	+
MInt059	JX137091	$(AAG)_3(GT)_{89}$ $(AG)_8(GA)_9$	TD	F: TATCAGGTAGTGTTCTTTTT R: TGTGTTTTCTTATTTATTCTA	30	241	+	+
MInt063	JX137093	$(TG)_3$	TD	F: GTAGGAACAAAGAAATGGTC R: TTACAAGTATGGAAGTGGAGT	29	264	-	+
MInt065	JX137094	$(TA)_3$	TD	F: CGGTTTTTCTTCTTTCACTCAA R: TCATGGTTTCAGGATTTCACTA	30	137	+	+
MInt079	JX137101	$(CT)_{11}(TC)_{12}$	TD	F: CAGACACTTGCCCCTACATAC R: GTCTCAAACTCTCAACCACTCA	29	212	+	+

cont. Table 1

MInt087	JX137105	$(GA)_{33}$	TD	F: CACCCTCTTCATCACCGTCGTT R: TTCCCCAGTCCATCCCAATCT	30	204	+	+
MInt089	JX137106	$(AC)_3$	TD	F: GATCTGACGCCTTACGCTGACT R: CTTTGGGGTTTCGATTGGAGAG	28	190	-	+
MInt093	JX137108	$(TC)_3(TC)_4$	TD	F: CAACCAAAGCCATCAAAGAGTG R: ACGCTAGACGGTGGGAGAAA	30	140	-	-
MInt095	JX137109	$(CA)_3(TA)_3$	TD	F: GTTCGCCATCCTTTTGACA R: CAGGGTGCTTTGAGAACATTAG	30	142	+	+
MInt099	JX137111	$(GT)_5(GA)_{13}(AG)_4$	TD	F: AAAAGGGGGAAAAAGAGATTGT R: ACTTGTCGCTATGCACTGGATT	29	264	+	+

Ta: annealing temperature; TD: *touchdown*; F: forward sequence; R: reverse sequence; N: sample.

PCR amplifications were performed in 25 µL total volume containing 10 ng of genomic DNA, 0.8 µM forward and reverse primers, 100 µM of each dNTP, 1.5 mM $MgCl_2$, 10 mM Tris–HCl, 50 mM KCl, and 0.5 U Taq DNA Polymerase (Invitrogen, CA, USA). The PTC-200 thermal cycler (MJ Research, Waltham, MA/USA) was used for PCR amplifications with touchdown cycling program (temperature ranging from 45 to 60 °C). Amplification products were genotyped using electrophoresis on 6% denaturing polyacrylamide gels, 1x TBE buffer, and visualized using silver staining 0.2% (Creste et al. 2001).

The SSR loci characterization was based on the following genetic parameters: expected heterozygosity (He), observed heterozygosity (Ho), and polymorphism information content (PIC; as described in Faria et al. 2022). These parameters were calculated using FSTAT program (Goudet 1995). The fixation index (*f*) and the tests of Hard-Weinberg Equilibrium (HWE) were performed using the Tools for Population Genetic Analysis (TFPGA; Miller 1997). The linkage disequilibrium (LD) was analyzed using the GDA software (Lewis and Zaykin 2002). The presence of null alleles was tested using the Micro-checker software (Oosterhout et al. 2004).

The genetic structure of IAC macadamia germplasm was investigated by Structure software (Pritchard et al. 2000), considering the probability of each genotype being allocated in *K* clusters. The parameters established to run the software were: 20 independent iterations for each *K* value; *K* ranging from 1 to 10; and 10,000 Markov Chain Monte Carlo (MCMC; 10,000 burn-in) steps. The selected model of ancestry was the admixture, which assumes that the genomes may have a proportion of more than one population (Grünwald et al. 2017). The model of allele frequency adopted in this study was correlated among populations. The estimate of the most-likely number of *K* subpopulations was obtained by Δ*K* (Evanno et al. 2005), from Structure Selector software (Li and Liu 2017).

The genetic distances between pairs of macadamia cultivars and *Grevillea* sp. were estimated by the modified Rogers' distances, and the unweighted pair group method, with arithmetic mean (UPGMA) as agglomeration method (Silva 2016), was used to construct the dendrogram. To verify the accuracy of these estimates, 1,000 bootstraps per locus were performed. The cut-off was based on the K-means algorithm from 'factoextra' R package (Kassambara and Mundt 2019).

The Principal Component Analysis (PCA) was applied to describe similarities among the cultivars of the macadamia germplasm and *Grevillea* sp. The modified Rogers' distance, bootstraps, dendrogram and PCA were obtained using the 'cluster' (Maechler et al. 2019), 'MASS' (Ripley et al. 2019) and 'ggplot2' (Wickham et al. 2019) packages in R-software environment (R Core Team 2019).

RESULTS AND DISCUSSION

Eighty-two clones containing microsatellite sequences were considered suitable for primer design. Thus, 51 primer pairs were designed, of which 29 were polymorphic (Table 1). Twenty-five microsatellite loci were cross-amplified in *G. robusta* and *G. banksii*. The number of macadamia's alleles per locus ranged from two to nine, with an average value of 3.65 alleles per locus (Table 2). Schmidt et al. (2006) developed 33 SSR loci for macadamia and obtained from one to 15 alleles per locus, while five loci were monomorphic. Nock et al. (2014) developed 12 SSR loci for macadamia, which showed between two and nine alleles per locus, with average values equal to 5.92 (*M. integrifolia* and hybrids) and 4.92

Table 2. Estimate of allele frequency by microsatellite locus (SSR) in macadamia (*Macadamia integrifolia*)

Locus	Allele	Allele frequency	Locus	Allele	Allele frequency	Locus	Allele	Allele frequency
MInt001	1	0.414	MInt019	1	0.183	MInt055	1	0.033
	2	0.121		2	0.150		2	0.083
	3	0.069		3	0.550		3	0.884
	4	0.396		4	0.100	MInt059	1	0.450
MInt005	1	0.450		5	0.017		2	0.550
	2	0.183	MInt025	1	0.700	MInt063	1	0.172
	3	0.050		2	0.300		2	0.672
	4	0.317	MInt029	1	0.121		3	0.086
MInt007	1	0.750		2	0.034		4	0.052
	2	0.214		3	0.017		5	0.017
	3	0.036		4	0.103	MInt065	1	0.167
MInt009	1	0.183		5	0.534		2	0.017
	2	0.017		6	0.172		3	0.250
	3	0.017		7	0.017		4	0.117
	4	0.667	MInt031	1	0.467		5	0.033
	5	0.117		2	0.583		6	0.250
MInt011	1	0.083	MInt033	1	0.679		7	0.067
	2	0.433		2	0.321		8	0.067
	3	0.017	MInt035	1	0.350		9	0.033
	4	0.167		2	0.650	MInt079	1	0.534
	5	0.200	MInt037	1	0.214		2	0.466
	6	0.033		2	0.268	MInt087	1	0.067
	7	0.017		3	0.518		2	0.666
	8	0.033	MInt039	1	0.069		3	0.267
	9	0.017		2	0.759	MInt089	1	0.018
MInt013	1	0.207		3	0.172		2	0.616
	2	0.017	MInt041	1	0.067		3	0.821
	3	0.776		2	0.050	MInt093	1	0.983
MInt015	1	0.037		3	0.816		2	0.017
	2	0.241		4	0.067	MInt095	1	0.250
	3	0.204	MInt045	1	0.933		2	0.750
	4	0.074		2	0.067	MInt099	1	0.017
	5	0.019	MInt047	1	0.875		2	0.086
	6	0.370		2	0.125		3	0.621
	7	0.056	MInt051	1	0.900		4	0.276
				2	0.100	Number of alleles		106
						Number of rare alleles		25
						Average of alleles by loci per loci		3.65

Rare alleles: frequency ≤ 0.05.

(*M. tetraphylla*) alleles. This study obtained allele frequencies ranging from 0.017 to 0.983. 106 alleles were detected, of which 25 were classified as rare alleles (*frequency* ≤0.05), representing 23.6% of total alleles. The loci MInt011 and MInt065 showed the largest number of alleles (nine), of which five and three were classified as rare, respectively. Only one locus (MInt093) showed allele frequency over 0.95, indicating a high risk of fixation.

The values for He and Ho ranged from 0.034 to 0.836 and 0.034 to 0.967, respectively (Table 3). The average values for He and Ho were 0.46 and 0.43, respectively, showing small reduction of heterozygosity when the Hard-Weinberg Equilibrium (HWE) is expected. Of the 29 loci developed in this study, only two showed significant deviations of HWE after Bonferroni correction (p<0.005): MInt011 and MInt035. Despite the selection pressure resulting from the breeding program, this result suggests that, in terms of mean of the loci, the cultivars meet the Hardy-Weinberg Equilibrium. The recent macadamia

Table 3. Microsatellite genetic diversity parameters of *Macadamia integrifolia*: number of alleles (A), expected heterozygosity (He), observed heterozygosity (Ho), p-value of chi-square test for *Hardy-Weinberg Equilibrium (HWE),* polymorphism information content (PIC) and *null allele frequency*

Locus	A	He	Ho	HWE	PIC	Null allele frequency
MInt001	4	0.663	0.690	0.102	0.58	0.000
MInt005	4	0.672	0.533	0.096	0.60	0.000
MInt007	3	0.397	0.286	1.000	0.34	0.000
MInt009	5	0.516	0.600	0.088	0.46	0.012
MInt011	9	0.747	0.867	0.001*	0.70	0.000
MInt013	3	0.361	0.448	0.069	0.30	0.000
MInt015	7	0.767	0.852	0.051	0.72	0.011
MInt019	5	0.642	0.600	0.329	0.59	0.000
MInt025	2	0.427	0.267	0.548	0.33	0.000
MInt029	7	0.669	0.552	0.042	0.62	0.000
MInt031	2	0.506	0.733	0.177	0.37	0.000
MInt033	2	0.444	0.500	0.246	0.34	0.000
MInt035	2	0.463	0.700	0.002*	0.35	0.033
MInt037	3	0.625	0.286	1.000	0.54	0.000
MInt039	3	0.397	0.276	1.000	0.35	0.000
MInt041	4	0.327	0.167	0.078	0.31	0.000
MInt045	2	0.126	0.133	0.099	0.14	0.000
MInt047	2	0.226	0.250	0.059	0.20	0.000
MInt051	2	0.183	0.133	0.162	0.16	0.000
MInt055	3	0.215	0.100	0.132	0.22	0.000
MInt059	2	0.503	0.500	1.000	0.37	0.000
MInt063	5	0.517	0.379	1.000	0.47	0.000
MInt065	9	0.836	0.967	0.053	0.80	0.000
MInt079	2	0.506	0.586	0.081	0.37	0.023
MInt087	3	0.488	0.467	0.096	0.41	0.000
MInt089	3	0.304	0.357	0.158	0.26	0.000
MInt093	2	0.034	0.034	0.251	0.03	0.000
MInt095	2	0.381	0.433	1.000	0.30	0.000
MInt099	4	0.540	0.241	0.063	0.46	0.000
Mean	3.65	0.46	0.43	-	0.40	-

P values for the HWE test, significance threshold adjusted using Bonferroni correction: * significant at $P < 0.05$;

domestication may explain this result, since the current cultivars are genetically close to their wild relatives (Topp et al. 2019). Schmidt et al. (2006) found He and Ho values ranging from 0.027 to 0.875 and 0.027 to 0.882, respectively, with 20 loci showing significant HWE deviations. Nock et al. (2014) found a mean He of 0.626 (from 0.165 to 0.8 loci) and mean Ho of 0.571 (from 0.091 to 0.909) for *M. integrifolia* and hybrids, whereas for *M. tetraphylla*, they found mean He of 0.632 (from 0.133 to 0.847) and Ho of 0.573 (from 0.143 to 0.857).

The average PIC value among loci was 0.40, with values ranging from 0.03 to 0.80 (Table 3). Schmidt et al. (2006) obtained an average value for PIC equal to 0.480 (from 0.026 to 0.848). The loci MInt009, MInt015, MInt035 and MInt079 were significant for the presence of null alleles. Null alleles may occur when mutations in the prime annealing site prevent the efficiency of at least one primer and this leads to a failure of amplification during the PCR reaction (Rico et al. 2017). The presence of null alleles may be the result of underestimation of heterozygotes (Rico et al. 2017). Schmidt et al. (2006) showed evidence of the presence of null alleles in locus MinμS53, which had no PCR products in 48% of the samples. These authors recommended that this locus, despite the evidence of null alleles, should be used as codominant marker in genome mapping if the parents were contrasting for presence/absence of mark. No significant linkage disequilibrium was detected among all loci, as indicated by the chi-squared test ($p<0.001$).

The results of the Structure software suggested the existence of no genetic structure among the cultivars. Although the Evanno et al. (2005) method showed the highest Δ*K* for *K* = 2 (Figure 1A), a more detailed analysis of the estimated average and standard deviation of the likelihood of the distinct models (*K* = 1 to *K* = 10) indicated that the model with *K* = 1 is also a likely one (absence of structure). This latter model showed the second highest likelihood and lowest standard deviation among the 20 repetitions (Figure 1B). As pointed out by Soares et al. (2020), the Evanno et al. (2005) Δ*K* method may not be able to assess the lack of genetic structure. As described below, this lack of genetic structure between the cultivars is also suggested by all other methods used in this study.

The modified Rogers' distances ranged from 0.227 (between 'HAES 741' and 'HAES 660') to 0.671 (between 'IAC Campinas B a' and the species *G. banksii*; Figure 2). The dendrogram classified the genotypes into two groups: an external formed by *Grevillea* species and another corresponding to all macadamia cultivars. One may expect that cultivars originated from distinct geographic regions would form different clusters, belonging to distinct subpopulations. However, this expectation could not be observed in Figure 2. According to the modified Rogers' distances, all macadamia cultivars formed a single cluster, apart from the *Grevillea* species. In other words, all macadamia cultivars, regardless of their geographic origin, did not show genetic divergence capable of placing them into distinct subpopulations. This result is in line with Aradhya et al. (1998), who observed that cultivars released by IAC were dispersed among Hawaiian, Californian and Australian cultivars. The results depicted in Figure 2 also indicate that the genotypes 'IAC 4-20 a' and 'IAC 4-20 b' showed a relatively high genetic divergence with respect to 'IAC 4-20' cultivar, which is from the IAC macadamia germplasm. This may suggest a nomenclature mistake. A similar feature is also observed in 'IAC 9-20' and 'IAC Campinas B' and their corresponding genotypes. Several authors have described historical (Hardner 2016), phenotypical (Alam et al. 2019) and molecular genetic (Aradhya et al. 1998, Steiger et al. 2003, Peace et al. 2005, Alam et al. 2019) similarity between 'HAES 660' and 'HAES 741'. This similarity between these two cultivars indicate that they were obtained from an autogamous full-sib progeny (Alam et al. 2019). The results of the PCA (Figure 3) agree with those depicted in Figure 2. As the cluster analysis, the PCA also indicated that *G. robusta* and *G. banksii* formed an external group and the macadamia cultivars formed a single cloud of points regardless of their geographic origin.

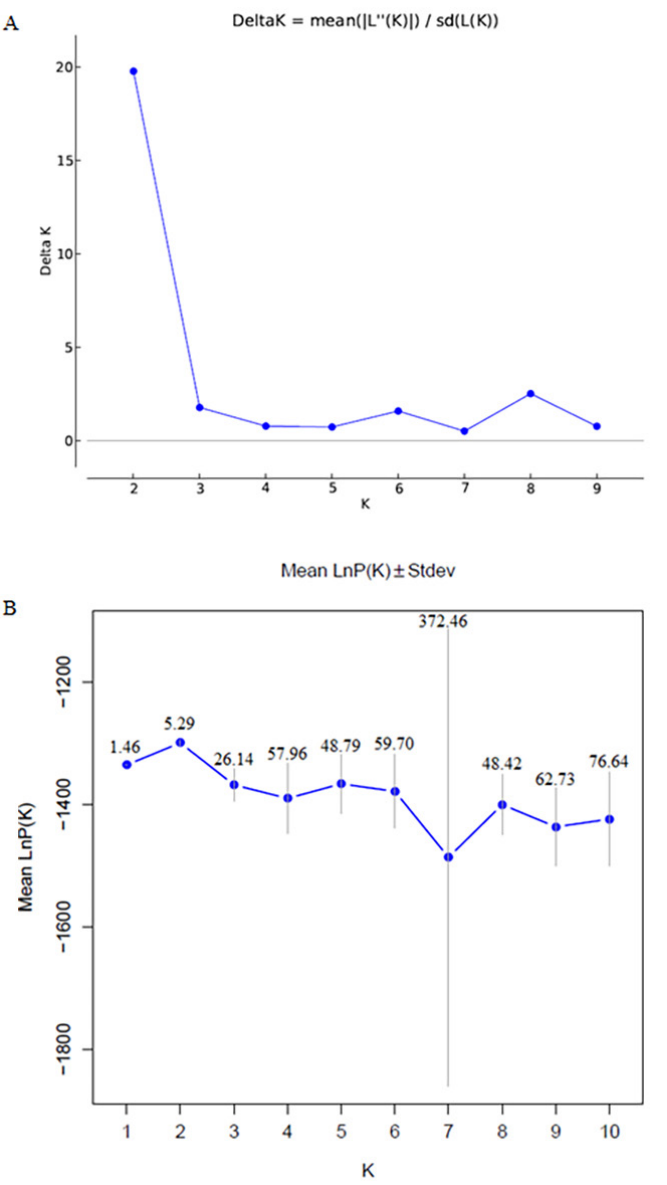

Figure 1. Genetic structure of IAC's macadamia germplasm: (A) Δ*K* values for the different number of clusters (*K*), calculated according to Evanno et al. (2005); (B) mean and standard deviation of the likelihoods of the different models obtained with *K* ranging from 1 to 10, indicating *K* = 1 as the most likely model.

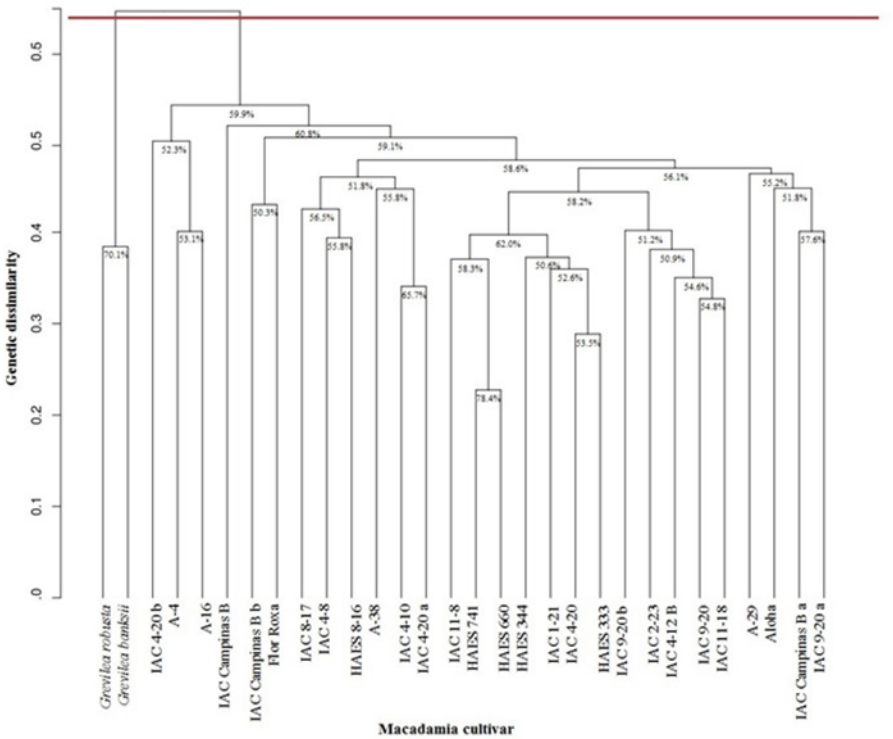

Figure 2. Dendrogram based on modified Rogers' distances showing the genetic relationship among 28 *Macadamia integrifolia* cultivars and the species *Grevillea robusta* and *G. banksii* (cut-off based on K-means algorithm; Kassambara and Mundt 2019).

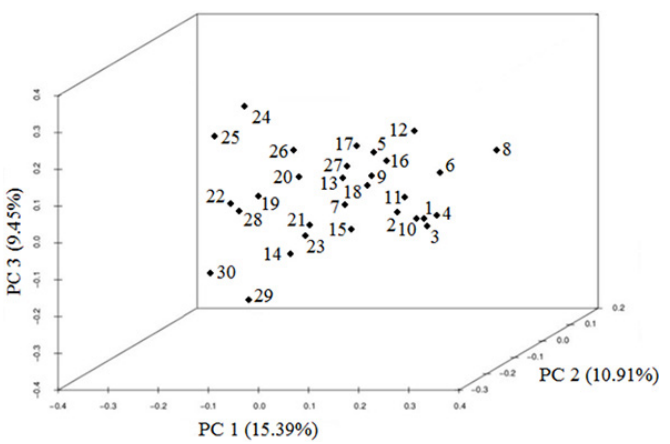

Figure 3. Principal Component Analysis with the respective proportions of variation explained by the first three principal components. Cultivars: 1 - IAC 2-23; 2 - IAC 4-20; 3 - IAC 9-20; 4 - IAC 11-18; 5 - IAC 4-10; 6 - IAC 4-8; 7 - IAC 11-8; 8 - IAC 8-17; 9 - IAC 1-21; 10 - IAC 4-12 B; 11 - IAC 4-20 a; 12 - HAES 816; 13 - HAES 344; 14 - HAES 741; 15 - HAES 660; 16 - HAES 333; 17 - IAC Campinas B; 18 - IAC Campinas B a; 19 - IAC 4-20 b; 20 - IAC 9-20 a; 21 - IAC 9-20 b; 22 - IAC Campinas B b; 23 - Flor Roxa; 24 - A-4; 25 - A-16; 26 - A-29; 27 - A- 38; 28 - Aloha; 29 - *Grevillea robusta*; 30 - *G. banksii*.

CONCLUSIONS

The Agronomic Institute's Macadamia Germplasm shows no genetic structure between cultivars. The 28 cultivars evaluated in this study were classified into single population regardless of their distinct geographic origin.

Despite the lack of information about the historical genealogy, the cultivars from the IAC Germplasm have moderate genetic diversity, so they can be used as parents for new crosses.

The microsatellite markers developed in this study show high polymorphic level among the loci, which may be suitable for genetic studies, including those addressing genetic conservation, determination of the degree of relatedness among individuals or groups of accessions, and breeding.

The cross-species amplification data suggest that microsatellites developed for *M. integrifolia* are useful for genetics studies of *G. robusta* and *G. banksii*.

ACKNOWLEDGMENTS

To São Paulo State Research Support Foundation (FAPESP, Brazil, Process 2011/06141-0) and to Agronomic Institute (project SGP 66) for the financial support. To Brazilian Agricultural Research Corporation (EMBRAPA, Brazil) and to National Council for Scientific and Technological Development (CNPq, Brazil, Process 307616/2019-3) for providing the grant for the first and the second author, respectively. Supplemental materials are available from corresponding author.

REFERENCES

Alam M, Hardner CM, Nock C, O'Connor K and Topp B (2019) Historical and molecular evidence of genetic identity of macadamia cultivars HAES741 and HAES660. **HortScience 54**: 616-620.

Aradhya MK, Yee LK, Zee FT and Manshardt RM (1998) Genetic variability in macadamia. **Genetic Resources and Crop Evolution 45**: 19-32.

Billotte N, Lagoda PJR, Risterucci AM and Baurens FC (1999) Microsatellite enriched libraries: applied methodology for the development of SSR markers in tropical crops. **Fruits 54**: 277-288.

Busanello C, Venske E, Stafen CF, Pedrolo AM, Luz VK, Pedron T, Paniz FP, Batista BL, Magalhães Júnior AM, Oliveira AC and Pegoraro C (2020) Is the genetic variability of elite rice in southern Brazil really disappearing? **Crop Breeding and Applied Biotechnology 20**: e262620214.

Carena M (2021) Germplasm enhancement and cultivar development: the need for sustainable breeding. **Crop Breeding and Applied Biotechnology 21**: e385621S4.

Creste S, Tulmann Neto A and Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter 19**: 299-306.

DNA Star (2000) Laser Gene: expert sequence analysis software. Available at <https://www.dnastar.com/>. Accessed on January 26, 2012.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Molecular Ecology 14**: 2611-2620.

Faria SV, Zuffo LT, Rezende WM, Caixeta DG, Pereira HD, Azevedo CF and DeLima RO (2022) Phenotypic and molecular characterization of a set of tropical maize inbred lines from a public breeding program in Brazil. **BMC Genomics 23**: 54-70.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-Statistics. **Journal of Heredity 86**: 485-486.

Grünwald NJ, Everhart SE, Knaus BJ and Kamvar ZN (2017) Best practices for population genetic analyses. **Phytopathology 107**: 1000-1010.

Guan C, Zhang Y, Zhang P, Chachar S, Wang R, Du X and Yang Y (2020) Germplasm conservation, molecular identity and morphological characterization of persimmon (*Diospyros kaki* Thunb.) in the NFGP of China. **Scientia Horticulturae 272**: 109490.

Hardner CM (2016) Macadamia domestication in Hawaii. **Genetic Resources and Crop Evolution 63**: 1411-1430.

Hardner CM, Silva JC, Williams E, Meyers N and McConchie C (2019) Breeding new cultivars for the Australian macadamia industry. **HortScience 54**: 621-628.

He T and Li C (2020) Harness the power of genomic selection and the potential of germplasm in crop breeding for global food security in the era with rapid climate change. **Crop Journal 8**: 688-700.

Inglis PW, Pappas MCR, Resende LV and Grattapaglia D (2018) Fast and inexpensive protocols for consistent extraction of high-quality DNA and RNA from challenging plant and fungal samples for high throughput SNP genotyping and sequencing applications. **PLoS One 13**: e0206085.

Kassambara A and Mundt F (2019) Package 'factoextra'. Available at <https://cran.r-project.org/web/packages/factoextra/factoextra.pdf>. Accessed on May 5, 2019.

Kern S, Santos B, Topp B, Cave R, Bignell G, Mulo S and Hardner C (2022) Using choice analysis of growers' preferences to prioritize breeding traits in horticultural tree crops: a macadamia case study. **Scientia Horticulturae 294**: 110766.

Lewis P and Zaykin D (2002) Genetic Data Analysis (GDA): computer program for the analysis of allelic data. Available at <http://phylogeny.uconn.edu/software/>. Accessed on January 26, 2012.

Li YL and Liu JX (2017) Structure Selector: A web-based software to select and visualize the optimal number of clusters using multiple methods. **Molecular Ecology Resources 18**: 176-177.

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K, Studer M, Roudier P, Gonzalez J, Kozlowski K, Schubert E and Murphy K (2019) Package 'cluster'. Available at <https://cran.r-project.org/web/packages/cluster/cluster.pdf>. Accessed on May 5, 2019.

Miller MP (1997) Tools for Population Genetic Analysis (TFPGA) 1.3.: a windows program for the analysis of allozyme and molecular population genetic data. Available at <http://bioweb.usu.edumpmbio>. Accessed on January 26, 2012.

Nock CJ, Elphinstone MS, Ablett G, Kawamata A, Hancock W, Hardner CM and King GJ (2014) Whole genome shotgun sequences for microsatellite discovery and application in cultivated and wild macadamia (Proteaceae). **Applications in Plant Science 2**: 1300089.

Nock CJ, Hardner CM, Montenegro JD, Termizi AAA, Hayashi S, Playford J, Edwards D and Batley J (2019) Wild origins of macadamia domestication identified through intraspecific chloroplast genome sequencing. **Frontiers in Plant Science 10**: 1-15.

Oosterhout C, Hutchinson B, Wills D and Shipley P (2004) Micro-Checker: Software for identifying and correcting genotyping errors in

microsatellite data. **Molecular Ecology Notes 4**: 535-538.

Peace CP, Allan P, Vithanage V, Turnbull CGN and Carroll BJ (2005) Genetic relationships among macadamia varieties grow in South Africa as assessed by RAF markers. **South African Journal of Plant and Soil 22**: 71-75.

Pritchard JK and Stephens Mand Donnelly P (2000) Inference of population structure using multilocus genotype data. **Genetics 155**: 945-959.

R Core Team (2019) R: A language and environment for statistical computing. Available at <http://www.r-project.org>. Accessed on May 31, 2019.

Rico C, Cuesta JA, Drake P, Macpherson E, Bernatchez L and Marie AD (2017) Null alleles are ubiquitous at microsatellite loci in the wedge clam (*Donax trunculus*). **PeerJ 5**: e3188.

Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A and Firth D (2019) 'MASS' Package. Available at <https://cran.r-project.org/web/packages/MASS/MASS.pdf>. Accessed on May 31, 2019.

Schmidt AL, Scott L and Lowe AJ (2006) Isolation and characterization of microsatellite loci from macadamia. **Molecular Ecology Notes 6**: 1060-1063.

Silva AR (2016) **Métodos de análise multivariada em R**. FEALQ, Piracicaba, 167p.

Soares SD, Bandeira LF, Ribeiro SB, Telles MPC, Silva JA, Borges CT, Coelho ASG and Novaes E (2020) Genetic diversity in populations of African mahogany (*Khaya grandioliola* C. DC.) introduced in Brazil. **Genetics and Molecular Biology 43**: e20180162.

Sobierajski GR (2019) Botânica, morfologia e exigências edafoclimáticas. In Pimentel L and Borém A (eds) **Macadâmia: do plantio à colheita**. UFV, Viçosa, p. 20-35.

Steiger DL, Moore PH, Zee F, Liu Z and Ming R (2003) Genetic relationship of macadamia cultivars and species reveled by AFLP markers. **Euphytica 132**: 269-277.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Catinhour S and McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations and genetic marker potential. **Genome Resources 11**: 1441-1452.

Toledo Piza PLB, Piza IM and Almeida Neto JTP (2019) A cultura. In Pimentel L and Borém A (eds) **Macadâmia: do plantio à colheita**. UFV, Viçosa, p. 9-19.

Topp B, Hardner C, Alam M, Akinsanmi O, O'Connor K, Mai T and Russell D (2017) Opportunities and challenges in macadamia breeding. In **International Macadamia Research Symposium**. USDA, Big Island, p. 18.

Topp BL, Nock CJ, Hardner CM, Alam M and O'Connor KM (2019) Macadamia (*Macadamia* spp.) breeding. In Al-Khayri JM, Jain S and Johnson D (eds) **Advances in plant breeding strategies: nut and beverage crops**. Springer International Publishing, Cham, p. 221-251.

Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke K and Woo K (2019) Package "ggplot2". Available at <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>. Accessed on May 31, 2019.