Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2019) 43: 437-449

doi:10.3906/tar-1807-25

Diversity of avocado (*Persea americana* Mill.) cultivars from Antioquia (Northeast Colombia) and comparison with a worldwide germplasm collection

Gloria Patricia CAÑAS-GUTIÉRREZ[1,2], Librada ALCARAZ[3], Jose Ignacio HORMAZA[1], Rafael Eduardo ARANGO-ISAZA[2,3], Clara Inés SALDAMANDO-BENJUMEA[4,*]

[1]Corporation for Biological Research (CIB), Unit of Phytosanity and Biological Control, Medellin, Colombia
[2]Corporation for Biological Research (CIB), Unit of Vegetal Biotechnology, Medellin, Colombia
[3]Experimental Station La Mayora – Spanish National Research Council (CSIC), Algarrobo-Costa, Malaga, Spain
[4]Biosciences School, Faculty of Sciences, National University of Colombia (UNALMED), Medellin, Colombia

Received: 05.07.2018 • **Accepted/Published Online:** 09.04.2019 • **Final Version:** 05.08.2019

Abstract: In this study, the genetic diversity of 90 avocado (*Persea americana* Mill) cultivars from Antioquia (Colombia) was compared to 67 germplasm collection accessions using 14 microsatellites. An average of 4.32 ± 2.0 alleles per locus was found, as observed in previous studies. The expected and observed heterozygosity ranges were 0.384–0.724 and 0.393–0.686, respectively. The Antioquian avocados were genetically structured according to an analysis of molecular variance test (fixation index (F_{ST}) = 0.054, P < 0.00001). An unweighted pair group method with arithmetic mean (UPGMA) dendrogram with F_{ST} paired data produced 2 clusters: one composed by Antioquian avocados and the other by the germplasm collection. Another UPGMA dendrogram with individual Nei–Li distances and additional STRUCTURE analysis separated the Antioquian avocados into 3 clusters (K = 3). Combining samples from the Colombian and germplasm collections produced K = 2. Genetic differentiations between the Antioquian and worldwide avocado germplasm collection could be due to crosses within the Antioquian avocados having been enhanced by insect pollinators, whereas avocados stored in the germplasm collection were constituted by cultivars with known genetic origins. Findings from this study demonstrated that criollo avocado genetics are unique in Antioquia, since the species has been naturally crossed in the field and its closest accession is from Guatemala. Nevertheless, it is important to continue genotyping this species in other locations in Colombia from Sylvester and cultivar populations of this crop to determine its origin.

Key words: Germplasm, microsatellites, *Persea americana*, population structure, STRUCTURE

1. Introduction

Avocado (*Persea americana* Mill.) is an important subtropical evergreen tree, native to Central America and Mexico, with 2n = 24 chromosomes (Ashworth and Clegg, 2003). This species is a member of the family Lauraceae, a mostly subtropical or tropical group included within the magnoliid clade of early-divergent angiosperms. Its family comprises approximately 50 genera and 2500 to 3000 species of trees and shrubs (Chanderbali et al., 2008). The first movements of avocado from its center of origin (central Mexico) occurred through the migration of big mammals (sloths and mammals of the family Gomphotheriidae) to Mesoamerica (Galindo-Tovar and Arzate-Fernández, 2010). Later on, *P. americana* was cultivated and domesticated by the first Mesoamerican cultures, the Mokaya, who might have transmitted their cultural practices to further civilizations, such as the Mayans and Olmecs (Chen et al., 2008). Its consumption, and possibly selection, took place from 4000 to 2800 BC and, by the time the Spaniards conquered Central and South America, this fruit had been cultivated from Mexico to Peru (Storey et al., 1986; Bost et al., 2013). Three horticultural races or botanical varieties of avocado have adapted to different climatic conditions and have traditionally been recognized as Mexican [*P. americana* var. *drymifolia* (Schltdl. & Cham)], Guatemalan (*P. americana* var. *guatemalensis* L.O.Williams), and West Indian (*P. americana* var. *americana* Mill.), which originated from highland Mexico, highland Guatemala, and lowland Mexico (Tierras bajas), respectively (Corona-Jacome et al., 2016). These botanical varieties are distinguishable on the basis of their morphological, physiological,

* Correspondence: cisaldam@unal.edu.co

and horticultural traits (Bergh and Lahav, 1996). Most commercial avocado cultivars are interracial hybrids developed from chance seedlings (Cañas-Guitierrez et al., 2015).

The evaluation of their morphological characters is inaccurate due to the influence of environmental factors, subjectivity, and a limited number of discriminating traits (Alcaraz and Hormaza, 2007). Consequently, avocado characterization and diversity studies have been strengthened by analyses of genetic markers including isozymes (Torres and Bergh, 1978), minisatellites (Mhameed et al., 1997), random amplified polymorphic DNA (RAPD) (Fiedler et al., 1998), restriction fragment length polymorphism (RFLP) (Davis et al., 1998), single sequence repeats (SSRs) (Ashworth and Clegg, 2003; Schnell et al., 2003; Alcaraz and Hormaza, 2007; Kwame et al., 2008; Gross-German and Viruel, 2013), amplified fragment length polymorphism (AFLP) (Douhan et al., 2011; Cañas-Gutierrez et al., 2015), and single nucleotide polymorphism (SNP) (Chen et al., 2008).

Microsatellites, also known as SSRs, are stretches of DNA that consist of tandem repeats of 1–6 bp. These molecular markers are located throughout nuclear, mitochondrial, and chloroplast genomes and represent suitable markers for studying population structure due to their codominance and high polymorphism (Freeland, 2005). Several microsatellite loci have been developed in avocado (Sharon et al., 1997; Ashworth et al., 2004; Borrone et al., 2007; Gross-German and Viruel, 2013). These markers have been used for linkage map construction (Sharon et al., 1997; Borrone et al., 2007), molecular germplasm characterization (Alcaraz and Hormaza, 2007), and diversity analysis (Schnell et al., 2003).

World avocado production reached almost 5 million tons in 2013, with 60% of this production coming from relatively few countries such as Mexico, the Dominican Republic, Colombia, Peru, and Indonesia, with Mexico being the principal worldwide producer with 30% of the total world production in 2013. In Colombia, particularly in the department of Antioquia, which has optimal climate and soil conditions for avocado growing, avocado production has increased in the last few years due to the potential interest of this crop for the export market (Cañas-Gutierrez et al., 2015).

In this work, a molecular characterization of Antioquian (Northwest Colombia) local avocado trees, named criollos (N = 90), collected in the field in different agroecological regions, was performed with 14 microsatellites developed by Sharon et al. (1997) and Ashworth et al. (2004) and tested by Alcaraz and Hormaza (2007) to determine their level of genetic diversity and compare these cultivars to other common avocado cultivars (N = 67) preserved in germplasm collections.

2. Materials and methods

2.1. Plant material and genomic DNA extraction

A sample collection was performed in the Department of Antioquia (in the north of Colombia) during 2008 and 2009. Young leaves were collected from a total of 90 criollo avocado trees located in 3 agroecological zones that were chosen according to Holdridge life-zones to present differences in climatic conditions. The sampling sites were also selected because avocado growers obtain seeds from these locations for rootstocks and grafting of different commercial varieties, particularly to produce the variety Hass. The locations included 2 populations from the western subregion Páramo (plateau), Sonsón and Abejorral; 5 populations from the western subregion, El Retiro, Marinilla, Rio Negro, La Ceja, and San Vicente; and 4 populations from the southwestern subregion, Caramanta, Santa Bárbara, Valparaiso, and Montebello (Table 1). Additionally, a total of 67 *P. americana* microsatellite genotyping data were obtained from accessions, including rootstocks, commercial varieties, and Spanish selections, maintained at the La Mayora Institute of Subtropical and Mediterranean Horticulture (IHSM, UMA-CSIC, Algarrobo-Costa, Málaga, Spain) (Alcaraz and Hormaza, 2007) (Table 2).

The total genomic DNA was obtained from avocado leaves based on the extraction method standardized by Cañas-Gutierrez et al. (2015).

2.2. Microsatellite analysis

Based on their high polymorphism, 14 SSRs were selected from those used by Alcaraz and Hormaza (2007) (Table 3). PCR amplifications were performed in a volume of 15 µL containing 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8), 0.01% Tween 20, 2 mM $MgCl_2$, 0.1 mM of each dNTP, 0.4 µM of each primer, 25 ng of genomic DNA, and 0.5 units of BioTaq DNA polymerase (Bioline, London, UK). Forward primers were labeled with WellRed fluorescent dyes at the 5' end (Proligo, France). Reactions were carried out in an I-Cycler thermocycler (Bio-Rad Laboratories, Hercules, CA, USA) using the following temperature profile: an initial step of 1 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s at 50 °C, and 1 min at 72 °C; and a final step of 5 min at 72 °C. The PCR products were analyzed using capillary electrophoresis in a CEQ 8000 capillary DNA analysis system (Beckman Coulter, Fullerton, CA, USA). Each reaction was repeated twice to minimize the run-to-run variation.

2.3. Data analysis

The allele size determination and mean number of alleles per SSR locus (N) were obtained from each sampled avocado population. The number of different alleles (Na), number of effective alleles (Ne = 1 / (sum pi^2)), information index (I = Shannon's information index = –1 × sum (pi ×

Table 1. List of avocado sampling sites in the department of Antioquia (Northeast Colombia)**.** West plateau = Oriente altiplano, West = Oriente, Southwest = Sur oriente, West paramo = Oriente páramo, West = Oriente.

ID	Municipality	Subregion	Altitude (m a.s.l.)	ID	Municipality	Subregion	Altitude (m a.s.l.)
17	Sonsón	Oriente altiplano	2237	175	Rionegro	Oriente	2168
24	Sonsón	Oriente altiplano	1907	176	Rionegro	Oriente	2172
25	Sonsón	Oriente altiplano	1892	201	Rionegro	Oriente	2150
27	Sonsón	Oriente altiplano	1915	202	Rionegro	Oriente	2150
33	Sonsón	Oriente altiplano	1833	203	Rionegro	Oriente	2166
316	Sonsón	Oriente altiplano	2205	204	Rionegro	Oriente	2158
317	Sonsón	Oriente altiplano	2207	145	La Ceja	Oriente	2080
322	Sonsón	Oriente altiplano	2200	147	La Ceja	Oriente	2066
329	Sonsón	Oriente altiplano	2152	150	La Ceja	Oriente	2151
330	Sonsón	Oriente altiplano	2146	152	La Ceja	Oriente	2137
331	Sonsón	Oriente altiplano	2136	156	La Ceja	Oriente	2178
354	Sonsón	Oriente altiplano	2340	157	La Ceja	Oriente	2164
356	Sonsón	Oriente altiplano	2341	4	Santa Barbara	Sur oriente	894
359	Sonsón	Oriente altiplano	2258	5	Santa Barbara	Sur oriente	864
360	Sonsón	Oriente altiplano	2268	107	Caramanta	Sur oriente	2086
362	Sonsón	Oriente altiplano	2273	109	Caramanta	Sur oriente	2089
374	Sonsón	Oriente altiplano	2145	117	Caramanta	Sur oriente	1980
388	Abejorral	Oriente altiplano	1800	118	Caramanta	Sur oriente	1994
390	Abejorral	Oriente altiplano	1761	119	Caramanta	Sur oriente	1977
399	Abejorral	Oriente altiplano	1636	293	Caramanta	Sur oriente	2027
400	Abejorral	Oriente altiplano	1647	294	Caramanta	Sur oriente	2026
404	Abejorral	Oriente altiplano	1650	295	Caramanta	Sur oriente	2020
411	Abejorral	Oriente altiplano	1642	297	Caramanta	Sur oriente	2011
412	Abejorral	Oriente altiplano	1644	298	Caramanta	Sur oriente	2005
413	Abejorral	Oriente altiplano	1655	312	Caramanta	Sur oriente	2092
417	Abejorral	Oriente altiplano	1658	314	Caramanta	Sur oriente	2085
1	El Retiro	Oriente altiplano	1975	315	Caramanta	Sur oriente	2091
2	El Retiro	Oriente altiplano	1920	99	Valparaiso	Sur oriente	1894
179	El Retiro	Oriente altiplano	2199	100	Valparaiso	Sur oriente	1906
193	El Retiro	Oriente altiplano	2112	101	Valparaiso	Sur oriente	1907
194	El Retiro	Oriente altiplano	2100	256	Montebello	Sur oriente	1965
35	San Vicente	Oriente altiplano	2084	257	Montebello	Sur oriente	1960
37	San Vicente	Oriente altiplano	2173	258	Montebello	Sur oriente	1954
38	San Vicente	Oriente altiplano	2175	259	Montebello	Sur oriente	1952
39	San Vicente	Oriente altiplano	2176	263	Montebello	Sur oriente	1932
40	San Vicente	Oriente altiplano	2197	264	Montebello	Sur oriente	1931
41	San Vicente	Oriente altiplano	2194	265	Montebello	Sur oriente	1930
44	Marinilla	Oriente altiplano	2049	266	Montebello	Sur oriente	1926
45	Marinilla	Oriente altiplano	2100	273	Montebello	Sur oriente	1915
46	Marinilla	Oriente altiplano	2124	276	Montebello	Sur oriente	1890
47	Marinilla	Oriente altiplano	2183	279	Montebello	Sur oriente	1867
62	Marinilla	Oriente altiplano	2149	280	Montebello	Sur oriente	1868
64	Marinilla	Oriente altiplano	2149	21 Hass	Sonsón	Oriente páramo	2181
170	Rionegro	Oriente altiplano	2150	307 Hass	Caramanta	Sur oriente	2376
171	Rionegro	Oriente altiplano	2150	428 Hass	Rionegro	Oriente altiplano	2135

Table 2. List of the 67 avocado accessions evaluated with SSRs in this study.

Accession	Country of origin	Race	Accession	Country of origin	Race	Accession	Country of origin
Acueductos	Spain (Málaga)	Unknown	G-6	Guatemala	M^{1}	OA184	USA
Adi	Israel	$G \times M^{10}$	G-755 A	Guatemala	G × P. *schiedeana*	Pinkerton	USA
Alhaurin	Spain (Málaga)	Unknown	Gem (3-29-5)	USA	$G \times M^{1}$ / G^{3}	San Javier 13	Spain (Málaga)
Bacon	USA	$G \times M^{1}$ / $M^{2,3}$	Gvar 13	Israel	M^{9}	San Javier 14	Spain (Málaga)
Bentavol gar	Spain (Málaga)	Unknown	Harvest	USA	$G \times M^{1}$ / G^{3}	San Javier 19	Spain (Málaga)
Bentavol mv	Spain (Málaga)	Unknown	Hass	USA	$G^{1,3}$ / $G \times M^{2}$	San Javier 28	Spain (Málaga)
BL 5552	USA	$G \times M^{1}$	Hass Motril	Spain (Málaga)	Unknown	San Javier 29	Spain (Málaga)
Bueno	Spain (Málaga)	Unknown	Iriet	Israel	$G \times M^{2}$	S. Chinea	Spain (Canary Islands)
Clavero	Spain (Málaga)	Unknown	La Consula 6	Spain (Málaga)	Unknown	Schmidt	Mexico
Coin	Spain (Málaga)	Unknown	La Consula 9	Spain (Málaga)	Unknown	Scott	USA
Colin V-33	Mexico	$M \times G^{3}$	La Consula 12	Spain (Málaga)	Unknown	Shepard	USA
Cupanda	Mexico	Unknown	Lamb Hass (BL122	USA	$G \times M^{1,2}$	Sir Prize	USA
Duke 6	Mexico	M^{1}	La Piscina	Spain (Málaga)	Unknown	Thomas	USA
Duke 7	Mexico	M^{1}	Lonjas	Mexico	M^{4}	Topa Topa	USA
Duke Parent	Mexico	M^{1}	Lula	USA	$G \times M^{3}$ / $G \times WI^{2,5}$	Toro Canyon	USA
Eden	Israel	$G \times M^{6}$	Maoz	Israel	WI^{8}	Villena	Spain (Málaga)
El Viso	Spain (Málaga)	Unknown	Marvel (BL 516)	USA	$G \times M^{1}$		
Fito Chinea	Spain (Canary Islands)	Unknown	Mexicola	USA	$M^{2,3}$		
Fuerte	Mexico	$G \times M^{1}$	Monsalve	Spain (Málaga)	Unknown		
Fundación 2	Mexico	Unknown	Negra de la Cruz	Chile	M^{5}		
Fuchs 20	Israel	$Wi \times G^{7}$	Nobel (BL667)	USA	$G \times M^{1}$		

Table 3. List of microsatellites employed in this study.

Locus	Reference	Marker	Allelic size (bp)	Number of alleles per locus
AVT 226	Ashworth et al., 2004	VIC	292–327	7
AVT 372	Ashworth et al., 2004	HEX	164–200	9
AVT 386	Ashworth et al., 2004	PET	220–234	6
AVD 102	Ashworth et al., 2004	PET	150–201	11
AVD 017	Ashworth et al., 2004	HEX	168–223	10
AVD 022	Ashworth et al., 2004	FAM	216–260	11
AVAG 21	Sharon et al., 1997	FAM	158–221	10
AVD.001	Ashworth et al., 2004	HEX	208–267	16
AVD006	Ashworth et al., 2004	FAM	301–348	16
AVA 12	Sharon et al., 1997	HEX	130–142	5
AVMIX03	Sharon et al., 1997	PET	143–190	16
AVMIX04	Sharon et al., 1997	VIC	121–194	16

Ln (pi)), Ho = observed heterozygosity = no. of hets. / N, He = expected heterozygosity = 1 – sum pi^2, and fixation index ((F = He – Ho) / He = 1 – (Ho / He)) for all avocado populations from Antioquia and the accessions maintained in the germplasm collection in Spain were calculated with GENALEX 6.501 software (Peakall and Smouse 2012).

An analysis of molecular variance (AMOVA) test was used to determine whether criollo populations from Antioquia were genetically different by employing GENALEX 6.501 (Peakall and Smouse, 2012). Genepop 4.2 (Raymond and Rousset, 1995) was used to explore the Hardy–Weinberg equilibrium and linkage disequilibrium

per population. GENALEX 6.501 was also used to calculate F_{ST} paired distances between the Antioquian localities and the avocado accessions maintained in the germplasm collection. These distances were used to obtain a dendrogram with the unweighted pair group method with arithmetic mean (UPGMA) algorithm (Sneath and Sokal, 1973). The dendrogram was obtained with the Mega 4.0 program (Kumar et al., 2008). To complement this last analysis, another UPGMA dendrogram was calculated per individual sample for the Antioquian avocado samples using Nei–Li genetic distances with the multivariate statistical package MVSP 3.22 (Kovach, 1998).

Additionally, the model-based clustering analysis tool STRUCTURE 2.3.4 was used (Pritchard et al., 2000) to assess the most probable cluster membership for each individual, first from all of the avocado populations from Antioquia and then combining the populations from Antioquia and the accessions maintained in the germplasm bank. This program was run for 10,000,000 Markov chain Monte Carlo steps after a burn-in period of 100,000 interactions for K = 1–15 under an admixture model. Each K was calculated from 10 independent runs. The ad hoc estimated likelihood of K (delta K) (Evanno et al., 2005) was employed to estimate the most likely number of populations (K) based on the rate of change in the log probability of the data [ln Pr (X/K)]. Finally, population genetics demographic analysis was performed using BOTTLENECK 1.2 software (Piry et al., 1999) to determine if the Colombian samples suffered from a bottleneck effect, as it estimates the distribution of He from the observed number of alleles (*k*), given the sample size (*n*), assuming equilibrium between the mutation and drift for each locus. This distribution simulates the coalescent process of *n* genes under 3 possible mutation models: the infinite alleles model, step mutation model, and 2-phase model. Therefore, the value of He calculated from the allele number by a coalescence procedure (Heq) was lower than the obtained He estimated directly from the allele frequencies (He). For neutral markers, in a population in gene mutation drift equilibrium, there is an equal probability that a given locus has a slight excess or deficit of heterozygosity with regard to the heterozygosity calculated from the number of alleles. In contrast, in a bottlenecked population, a large fraction of the loci analyzed will exhibit a significant excess of He. To measure this probability, 3 procedures were used in this study: a) sign test, b) standardized difference test, and c) Wilcoxon's signed rank test.

3. Results

In this work, 2 groups of avocado accessions were analyzed with 14 SSRs; the first group consisted of 90 local criollo genotypes sampled from 11 municipalities of Antioquia (Colombia) and the second group consisted of genetic data obtained by Alcaraz and Hormaza (2007) from an avocado collection that contains 67 genotypes representative of the overall diversity of avocado from different countries (N = 7).

3.1. SSR polymorphism and genetic diversity

The *P. americana* genetic diversity from the 18 populations generated a total of 143 fragments with 14 SSRs. The mean number of different alleles found in the Antioquian avocados was 4.33 ± 2.0. The highest number of total alleles was found in the Málaga (Spain) accessions with 350 alleles, followed by the USA with 266 alleles. In Antioquia, the highest number of alleles was found in Sonsón with 248 alleles and the municipality with the lowest number of alleles was Santa Bárbara with 28 alleles. Differences in the allele number between these sites were explained by differences in the samples sizes, as the lower the number of alleles, the smaller the sample size. Moreover, in the germplasm bank, accessions with the lowest number of alleles were Guatemala and Chile. However, their sample sizes were very small. Similar patterns of gene diversity were found between the mean values estimated for Na and Ne that are inherited in a population, as the population with the highest value was Málaga with 7.214 for Na, while the lowest value was 2 for Santa Barbara. In the case of the Ne, in Antioquia, the municipality with the highest Ne was Rio Negro, where Ne = 4.01, and the population with the lowest Ne was Santa Barbara with Ne = 1.81. The I value for the Antioquian avocados was between 0.582 (Santa Barbara) and 1.485 (Rio Negro and Caramanta). Ho and He ranged from 0.393 (Santa Bárbara) to 0.686 (Mexico) for the observed ones and from 0.384 (Santa Bárbara) to 0.724 (Rio Negro) for He. F_{ST} was positive for most populations, with the exception of Guatemala and Santa Bárbara, where its value was negative (Table 4).

Hardy–Weinberg equilibrium (HWE) was found for all of the microsatellites after Bonferroni corrections were tested per population (Table 5). The analysis of linkage disequilibrium showed that most microsatellites were in linkage equilibrium after Bonferroni corrections at α = 0.000357 (0.05/140) (data not shown).

3.2. Genetic relationship among accessions

The AMOVA result estimated for Antioquia was F_{ST} = 0.05439, $P < 0.0001$, suggesting genetic structuring among the municipalities sampled. Moreover, the AMOVA test showed that most of the genetic variation occurred within (94.5%) rather than between (5.44%) municipalities. Additionally, the UPGMA dendrogram based on F_{ST} paired distances between the populations from Antioquia and samples from the germplasm bank produced 2 main

Table 4. Sample size (N), no. of different alleles (Na), no. of effective alleles (Ne), information index (I), observed heterozygosity (Ho), expected heterozygosity (He), and fixation index (F) estimators obtained from all of the avocado populations from Antioquia (A) and the accessions of the gene bank from La Mayora (G).

Pop.		Mean	SD	Pop.		Mean	SD	Pop.		Mean	SD
Sonsón (A)	N	17.71	0.61	Caramanta (A)	N	13.64	0.50	San Vicente (A)	N	5.86	0.36
	Na	5.79	2.15		Na	6.43	1.91		Na	3.86	1.66
	Ne	3.34	1.28		Ne	3.74	1.29		Ne	2.78	1.10
	I	1.34	0.37		I	1.48	0.32		I	1.07	0.46
	Ho	0.49	0.16		Ho	0.43	0.19		Ho	0.40	0.27
	He	0.66	0.12		He	0.70	0.10		He	0.57	0.20
	F	0.26	0.23		F	0.36	0.29		F	0.32	0.36
Abejorral (A)	N	8.93	0.27	Montebello (A)	N	11.57	0.51	Marinilla (A)	N	5.93	0.27
	Na	4.14	1.99		Na	5.43	1.50		Na	3.64	1.28
	Ne	2.82	1.04		Ne	3.12	0.90		Ne	2.63	1.15
	I	1.12	0.40		I	1.31	0.33		I	1.01	0.41
	Ho	0.53	0.19		Ho	0.52	0.21		Ho	0.43	0.30
	He	0.60	0.13		He	0.65	0.12		He	0.55	0.19
	F	0.12	0.31		F	0.21	0.27		F	0.21	0.44
El Retiro (A)	N	4.79	0.58	Chile (G)	N	2.00	0.00	Rionegro (A)	N	8.79	0.58
	Na	3.14	1.03		Na	2.71	0.91		Na	5.64	1.50
	Ne	2.69	0.96		Ne	2.54	0.94		Ne	4.01	1.27
	I	0.98	0.39		I	0.90	0.39		I	1.49	0.31
	Ho	0.56	0.34		Ho	0.61	0.40		Ho	0.54	0.19
	He	0.57	0.20		He	0.54	0.20		He	0.72	0.10
	F	–0.01	0.55		F	–0.10	0.60		F	0.23	0.31
La Ceja (A)	N	6.00	0.00	Guatemala (G)	N	2.00	0.00	Santa Barbara (A)	N	2.00	0.00
	Na	4.79	1.58		Na	2.79	0.80		Na	2.00	0.68
	Ne	3.72	1.29		Ne	2.54	0.82		Ne	1.81	0.59
	I	1.37	0.33		I	0.92	0.36		I	0.58	0.36
	Ho	0.54	0.20		Ho	0.61	0.29		Ho	0.39	0.40
	He	0.70	0.09		He	0.55	0.19		He	0.38	0.23
	F	0.22	0.30		F	–0.12	0.40		F	–0.03	0.72
Islas Canarias (G)	N	3.00	0.00	Israel (G)	N	6.00	0.00	Valparaiso (A)	N	3.00	0.00
	Na	3.57	1.45		Na	4.79	1.42		Na	3.21	0.80
	Ne	3.11	1.50		Ne	3.67	1.20		Ne	2.66	0.84
	I	1.09	0.49		I	1.35	0.38		I	1.02	0.29
	Ho	0.55	0.31		Ho	0.65	0.25		Ho	0.57	0.33
	He	0.59	0.21		He	0.69	0.16		He	0.59	0.13
	F	0.09	0.41		F	0.04	0.28		F	0.03	0.51
Malaga (G)	N	25.00	0.00	Mexico (G)	N	10.00	0.00	USA (G)	N	19.00	0.00
	Na	7.21	2.75		Na	5.50	1.45		Na	6.86	2.21
	Ne	3.67	1.19		Ne	3.65	1.16		Ne	3.96	1.23
	I	1.47	0.36		I	1.42	0.29		I	1.53	0.35
	Ho	0.58	0.16		Ho	0.69	0.17		Ho	0.64	0.16
	He	0.70	0.10		He	0.70	0.09		He	0.72	0.10
	F	0.18	0.19		F	0.03	0.18		F	0.12	0.19

Table 5. HWE analysis performed on the avocado populations sampled in Antioquia.

Population	P-value	SE	Switches (average)
Sonson	1.000	0.0000	14299.71
Abejorral	0.9995	0.0002	22094.43
El Retiro	0.9839	0.0012	17810.15
La Ceja	1.000	0.0000	12670.14
San Vicente	1.000	0.0000	14988.69
Marinilla	0.9998	0.0001	13880.15
Rionegro	1.000	0.0000	9659.57
Santa Barbara	0.9854	0.0004	23649.25
Valparaiso	0.9898	0.0007	10294.38
Caramanta	1.000	0.0000	9671.36
Montebello	1.000	0.0000	11768.43
Guatemala	0.9936	0.0004	9437.91
Islas Canarias	0.9998	0.0001	8203.83
Israel	0.9940	0.0010	9430.29
Málaga	1.000	0.0000	8402.29
Mexico	0.9510	0.0053	11402.14
USA	1.000	0.0000	9547.07

clusters (Figure 1). The first cluster was composed of all the Antioquian samples and the second comprised the accessions maintained at the IHSM La Mayora. The Antioquian avocados were closer to the Guatemala accessions kept in the Spanish germplasm bank, suggesting a possible genetic proximity. Furthermore, the second UPGMA dendrogram based on the individual Nei–Li distances estimated from each individual sample from Antioquia produced 3 groups, where genetic associations between the cultivars were not explained by the vicinity of the municipalities where the samplings were made (Figure 2).

3.3. Population assignment results

When only the Antioquian avocado samples (N = 11 populations) were analyzed with STRUCTURE, 3 groups were clustered (K= 3), producing a likelihood value of ln (P|D) = –3839.8 and a likelihood improvement value of ΔK = 6.79 (Figure 2; Table 6). Cluster 1 was composed of avocado cultivars from Valparaíso, Sonsón, San Vicente, Marinilla, and Caramanta, with percentages of assignment between 88% and 63% and lower assignment percentages in clusters 2 and 3. Cluster 2 mostly contained avocados from Abejorral, Retiro, Rio Negro, Montebello, and Santa Bárbara, with assignment percentages between 78% and 43% and lower percentages in clusters 1 and 3. La Ceja was the only municipality where the avocados were almost evenly distributed in the 3 clusters, with percentages of assignment of 42% in cluster 1, 28% in cluster 2, and 29% in cluster 3.

When the Colombian genotypes were analyzed (N = 90) with the worldwide collection (N = 67), the Bayesian analysis of population structure produced 2 clusters (K = 2). The maximum likelihood value was ln (P|D) = –7705.3, producing a likelihood improvement value of ΔK = 496.15 (Figure 2). All of the Antioquian samples were grouped together within cluster 2 and separately from the genotypes maintained in the germplasm bank that were grouped in cluster 1 (Table 6). These results further supported the outcome obtained by the UPGMA tree, demonstrating that criollo avocado from Colombia is genetically differentiated from the avocado accessions of the germplasm bank.

3.4. Bottleneck results

Finally, in the bottleneck simulations, the sign test, the standardized difference test, and Wilcoxon's signed rank test produced no significant P-values, with the exception of the El Retiro and Rio Negro samples (Table 7). Tests were carried out given the genetic differentiation between criollo avocados and the germplasm materials, as shown by the UPGMA tree and STRUCTURE results. The bottleneck simulations suggested that avocado populations from Antioquia have not undergone recent population bottleneck effects through avocado artificial selection; thus, the Colombian avocado populations have not suffered a genetic reduction due to agriculturist manipulation.

4. Discussion

In this study, molecular characterization of 90 cultivars of criollo avocado from Antioquia (11 populations) was carried out with 14 microsatellites that were previously tested by Alcaraz and Hormaza (2007). This work was performed to compare these samples with 67 accessions found in a germplasm collection in Spain. These accessions were from the USA, Spain, Mexico, Israel, Guatemala, the Canary Islands, and Chile. In general, the results showed that the population with the highest allele diversity was Málaga (Spain), with 350 alleles genotyped in 11 individuals, followed by the USA accessions (N = 3) with 266 alleles and Sonsón (Antioquia) (N = 7) with 248 alleles. These results demonstrated that populations with high gene diversities originated from hybrid avocado samples, as they represented crosses between avocado varieties and races. For example, the USA accessions were representatives of crosses from Guatemalan × Mexican races, whereas cultivars from Antioquia were the result of natural crosses that occurred through insect pollinations

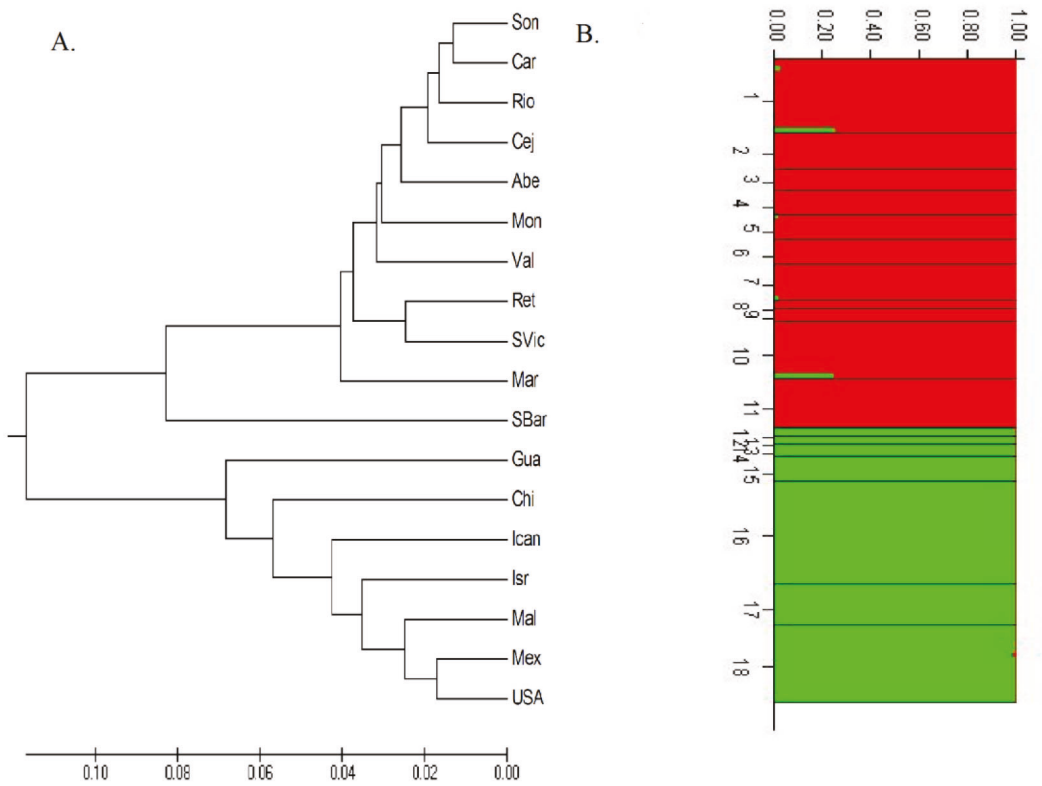

Figure 1. A) UPGMA dendrogram produced with F_{ST} paired genetic distances of avocado populations from Antioquia and avocado accessions maintained at the avocado germplasm collection of IHSM La Mayora in Spain. B) Number of K subpopulations obtained with STRUCTURE after performing the Evanno test with Antioquia avocado cultivars and avocado accessions maintained at the avocado germplasm collection of IHSM La Mayora in Spain. Son (1) = Sonsón, Car (2) = Caramanta, Rio (3) = Rionegro, Cej (4) = La Ceja, Abe (5) = Abejorral, Mon (6) = Montebello, Val (7) = Valparaiso, Ret (8) = Retiro, SVic (9) = San Vicente, Mar (10) = Marinilla, SBar (11) = Santa Bárbara, Gua (12) = Guatemala Chi (13) = Chile, Ican (14) = Canary Islands, Isr (15) = Israel, Mal (16) = Málaga, Mex (17) = Mexico, and USA (18) = United States.

within the avocado orchards. On the other hand, the lowest allele numbers were observed in Santa Bárbara (Antioquia), Guatemala, and Chile. All of these populations had low sample sizes, suggesting an effect of genetic drift in the sampling (Freeland, 2005). Differences in the Na and Ne estimators were also high for Málaga and low for Santa Bárbara, coinciding with the low number of alleles found for them. With respect to Ne, the highest value was found in Rio Negro, while the lowest was in Santa Bárbara. The I value was also lower in Santa Bárbara (0.582) and higher in Rio Negro and Caramanta (1.485). These results suggest higher gene diversity in Rio Negro and Caramanta than in the other municipalities of Antioquia. Ho values ranged between 0.393 (Santa Bárbara) and 0.686 (Mexico), while He ranged between 0.384 (Santa Bárbara) and 0.724 (Rio Negro). These results also show the relation between heterozygosity and allele number, since the more alleles, the more heterozygotes found in a natural population (Freeland, 2005). Wright's (F) fixation index (ratio between Ho and He) was positive for most of the populations, with the exception of Guatemala and Santa Bárbara, where the value was negative. These results suggested that most of the avocado accessions studied here were the product of inbreeding (when the inbreeding coefficient (F_{IS}) was positive) and few of them were the product of exogamy (F_{IS} was negative). Similar results were found by Alcaraz and Hormaza (2007), who suggested that these types of outcomes are expected from genetically related genotypes.

Another important result was related to the HWE, as it was found in all of the microsatellites after Bonferroni corrections, and no linkage disequilibrium between loci was obtained in this work. These results suggest that no

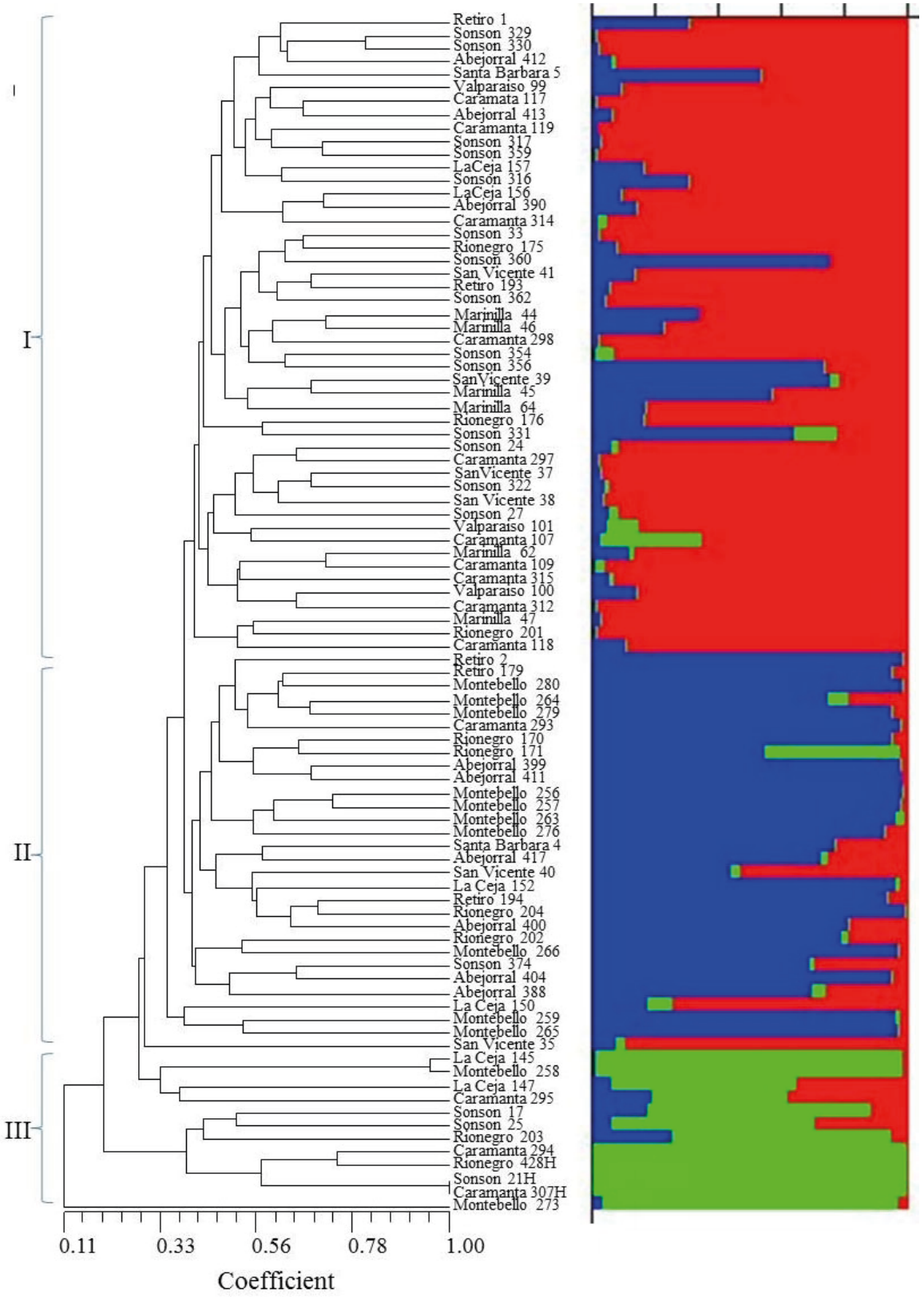

Figure 2. Clustering of 90 local avocado accessions from Antioquia based on a UPGMA dendrogram and Nei–Li distances, and a Bayesian analysis with STRUCTURE.

Table 6. Bayesian statistics analyzed with STRUCTURE: a) inferred clusters in Antioquia and b) inferred clusters obtained for Antioquia and germplasm bank.

a)

Pop.	Inferred clusters							N
	1	2	3	4	5	6	7	
Sonsón	0.117	0.063	0.237	0.059	0.447	0.059	0.018	18
Abejorral	0.005	0.250	0.240	0.018	0.174	0.292	0.022	9
Retiro	0.003	0.428	0.061	0.036	0.239	0.215	0.019	5
La Ceja	0.046	0.217	0.341	0.020	0.106	0.026	0.243	6
San Vicente	0.003	0.169	0.355	0.136	0.282	0.023	0.032	6
Marinilla	0.003	0.064	0.400	0.160	0.333	0.026	0.014	6
Río Negro	0.196	0.293	0.116	0.030	0.235	0.106	0.026	9
Santa Bárbara	0.002	0.662	0.158	0.011	0.106	0.041	0.019	2
Valparaíso	0.008	0.034	0.514	0.017	0.252	0.032	0.143	3
Caramanta	0.144	0.033	0.478	0.080	0.068	0.083	0.114	14
Montebello	0.006	0.014	0.031	0.092	0.018	0.742	0.097	12

b)

Pop.	Inferred clusters		N
	1	2	
Sonsón	0.016	0.984	18
Abejorral	0.001	0.999	9
El Retiro	0.001	0.999	5
La Ceja	0.001	0.999	6
San Vicente	0.006	0.994	6
Marinilla	0.001	0.999	6
Río Negro	0.003	0.997	9
Santa Bárbara	0.001	0.999	2
Valparaíso	0.001	0.999	3
Caramanta	0.019	0.981	14
Montebello	0.001	0.999	12
Chile	0.999	0.001	2
Guatemala	0.999	0.001	2
Islas Canarias	0.999	0.001	3
Israel	0.999	0.001	6
Málaga	0.999	0.001	25
Mexico	0.999	0.001	10
USA	0.998	0.002	19

evolutionary forces are influencing these markers and no genetic association between them exist, demonstrating that they are useful for population genetic studies (Freeland, 2005).

AMOVA estimated F_{ST} = 0.05439 (P < 0.0001) for the 90 avocado cultivars analyzed with 14 SSRs. This result demonstrated that *P. americana* is genetically structured among the 11 municipalities from Antioquia. Although this differentiation was only 5.4%, its value was significant, suggesting that avocados from this part of Colombia are genetically distinct according to sampled site. This outcome obtained in this work corroborates a previous study of 111 Antioquian criollo avocados genotyped with 38 AFLP markers by Cañas-Gutierrez et al. (2015). In this study, the authors suggested that avocado structuring in Antioquia was due to differences in the agroecological conditions of the studied zone. These ecological conditions were apparently related to the 3 horticultural races of the species, according to elevation sampling sites that were from 1400 to 1800, 1900 to 2000, and 2000 to 2400 m a.s.l. The relevance of using microsatellites with the same avocados employed by Cañas-Gutierrez et al. (2015) was that, in this work, heterozygous genotypes were discovered (Ho). AFLP markers are not very useful when studying crosses between individuals, as knowledge of the genotype of the parental and F1 generations is partial because they represent dominant markers, whereas microsatellites are codominant. Additionally, results obtained with the UPGMA dendrogram with the F_{ST} paired distances between the municipalities and accessions were separated into 2 main clusters. One was composed of avocados maintained in the germplasm from La Mayora (Spain) and the other comprised cultivars from Antioquia. These avocados were also separated and

Table 7. Bottleneck analysis performed in avocado samples from Antioquia. Sign = Sign test, SD = standard deviation test, W = Wilcoxon test, 1t = one tail, 2t = two tail, Prob = probability, IAM = infinite allele model, SMM = step mutation model, TPM = two phase model.

Population	Mean_N	Mean_k	Mean_He	P_sign_IAM
Sonson	35.43	5.86	0.683	0.241
Abejorral	17.86	4.21	0.641	0.070
El Retiro	9.57	3.14	0.638	0.009
La Ceja	12	4.79	0.766	0.236
San Vicente	11.71	3.86	0.628	0.496
Marinilla	11.86	3.64	0.599	0.376
Rionegro	17.57	5.79	0.772	0.114
Valparaiso	6	3.14	0.700	0.357
Caramanta	27.29	6.43	0.730	0.553
Montebello	23.14	5.43	0.677	0.434
Population	P_sign_TPM	P_sign_SMM	P_stdv_IAM	P_stdv_TPM
Sonson	0.323	0.056	0.057	0.468
Abejorral	0.187	0.343	0.022	0.181
El Retiro	0.010	0.003	0.001	0.004
La Ceja	0.495	0.526	0.020	0.114
San Vicente	0.534	0.350	0.212	0.477
Marinilla	0.412	0.245	0.305	0.388
Rionegro	0.565	0.297	0.019	0.160
Valparaiso	0.502	0.574	0.242	0.369
Caramanta	0.313	0.018	0.237	0.135
Montebello	0.169	0.004	0.314	0.117

analyzed in another UPGMA dendrogram and by Nei–Li distances. This dendrogram divided the 90 cultivars into 3 main groups that coincided with the number of K = 3 clusters estimated with STRUCTURE. These cultivars did not group together according to geographical proximity, suggesting no isolation by distance (Freeland, 2005). Moreover, these Antioquian cultivars were closely related to the Guatemalan accessions found in the germplasm bank.

When the 90 avocado cultivars from Antioquia were analyzed with the 67 accessions from La Mayora with STRUCTURE, K = 2. These results further suggested that Antioquian avocados were genetically different from the accessions of the germplasm bank. This could be due to avocados from Colombia having been naturally crossed between trees by insect pollinators (Yin et al., 2009), whereas avocados from La Mayora were the result of crosses between races and varieties of this species for its genetic improvement. The Bottleneck program showed that none of the tests (sign test, standardized difference test, and Wilcoxon's signed rank analysis) were significant for most of the Antioquian municipalities, with the exception of cultivars from El Retiro and Rio Negro. These tests showed that, in general, the number of alleles and Ho values from Antioquia did not suffer from the bottleneck effect. Differences in the allele composition and low gene diversity found in the Antioquian avocados versus the germplasm avocados might have been due to avocados being introduced into Colombia in pre-Columbian times from a limited number of genotypes. This, combined with the multiple crosses and germplasm exchanges of avocado trees within Antioquia, particularly between neighbor growers, would explain the development of a genetically homogeneous population. On the other hand, avocados from the germplasm bank were from different countries and varieties producing more alleles and heterozygous genotypes per microsatellite.

In conclusion, the results obtained here represent the first Colombian characterization of criollo avocados based on microsatellites markers. These outcomes showed that,

in general, Antioquia criollo cultivars probably clustered together due to the exchange of plant material among neighboring farms, where insect pollination influenced the chances of interbreeding between close avocado trees. This work also revealed population structuring within Antioquia, supporting previous AFLP analyses (Cañas-Gutierrez et al., 2015). Cañas-Gutierrez et al. (2015) suggested that agroclimatic conditions (Holdridge zones) played a role in the genetic differentiation of the avocado material evaluated and stressed the need to molecularly and morphologically characterize this local avocado material that existed before the intensification of the cultivation of Hass in this region of Colombia. Colombia represents the fifth largest avocado producer in the world through grafting of Hass with avocado criollo trees for fruit commercialization, while the European Union and the USA are the largest importers of this fruit (Chanderbali et al., 2008; Cañas-Gutierrez et al., 2015). However, the selection of avocado varieties has been a difficult task because avocado vegetative propagation has not been licensed yet, as seeds used to produce rootstocks are of unknown genetic origin (Rodríguez et al., 2009). To extend this type of study it would be desirable to analyze additional local avocados from other areas of Colombia and other South American countries to determine their genetic similarities and associate gene diversity with the agroecological conditions of this crop.

Acknowledgments

This work was financially supported in 2008 by the Colombian Ministry of Agriculture, under grant number 2008L72183-6387 and the project entitled "Caracterización Morfológica y molecular de materiales de aguacate para su aplicación en producción de material vegetal", as well as the Spanish Ministerio de Economía y Competitividad – European Regional Development Fund, European Union (AGL2013-43732-R and AGL2016-77267-R), INIA (RFP2015-00009), and the Ibero-American Program for Science, Technology and Development - CYTED (Natifrut project). Gloria Patricia Cañas-Gutierrez's internship at the IHSM La Mayora was supported by the Program for Enhancement Research Centers of Departamento de Administrativo de Ciencia, Tecnología e Innovación, Colciencias Colombia, and the Faculty of Agricultural Sciences at Universidad Nacional de Colombia, Sede Medellín.

References

Alcaraz M, Hormaza I (2007). Molecular characterization and genetic diversity in an avocado collection of cultivars and local Spanish genotypes using SSRs. Hereditas 144: 244-253.

Ashworth VETM, Clegg MT (2003). Microsatellite markers in avocado (*Persea americana* Mill.): genealogical relationships among cultivated avocado genotypes. J Hered 94: 407-415.

Ashworth VETM, Kobayashi MC, De la Cruz M, Clegg MT (2004). Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. Sci Hortic 101: 255-267.

Bergh BO, Lahav E (1996). Avocadoes. In: Janick J, Moore JN (editors). Fruit Breeding. Vol. 1. Tree and Tropical Fruits. New York, NY, USA: John Wiley, pp. 113-166.

Borrone JW, Schnell RJ, Violi HA, Ploetz RC (2007). Seventy microsatellite markers from *Persea americana* Miller (avocado) expressed sequence tags. Mol Ecol Notes 7: 439-444.

Bost JB, Smith NJH, Crane JH (2013). History, distribution and uses. In: Schaffer B (editor). The Avocado: Botany, Production, and Uses. 2nd ed. Wallingford, UK: CABI Publishing, pp. 10-30.

Cañas-Gutiérrez GP, Galindo-López LF, Arango-Isaza RE, Saldamando-Benjumea CI (2015). Diversidad genética de aguacate (*Persea americana*) en Antioquia. Colombia. Agronomía Mesoamericana 26: 129-143 (in Spanish).

Chanderbali AS, Albert VA, Ashworth VTEM, Clegg MT, Litz RE, Soltis DE, Soltis PS (2008). *Persea americana* (avocado): bringing ancient flowers to fruit in the genomics era. Bioassays 30: 386-396.

Chen H, Morrell PL, De la Cruz M, Clegg MT (2008). Nucleotide diversity and linkage disequilibrium in wild avocado (*Persea americana* Mill.). J Hered 99: 382-389.

Corona-Jacome EC, Galindo-Tovar ME, Lee-Espinosa HE, Landero-Torres I (2016). Diversidad genética del aguacate (Persea americana Mill.) en cuatro zonas de su área de dispersión natural. Agroproductividad 6: 80-85 (in Spanish).

Davis J, Henderson D, Kobayashi M, Clegg MT (1998). Genealogical relationships among cultivated avocado as revealed through RFLP analyses. J Hered 89: 319-323.

Douhan GW, Fuller E, McKee B, Pond E (2011). Genetic diversity analysis of avocado (*Persea americana* Miller) rootstocks selected under greenhouse conditions for tolerance to phytophthora root rot caused by *Phytophthora cinnamomi*. Euphytica 182: 209-217.

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

Fiedler J, Bufler G, Bangerth F (1998). Genetic relationships of avocado (*Persea americana* Mill.) using RAPD markers. Euphytica 101: 249-255.

Freeland J (2005). Molecular Ecology. Chichester, UK: John Wiley.

Galindo-Tovar ME, Arzate-Fernández MA (2010). Consideraciones sobre el origen y primera dispersión del aguacate (*Persea americana*, Lauraceae). Cuadernos de Biodiversidad 33: 11-15.

Galindo-Tovar ME, Ogata-Aguilar N, Arzate-Fernández AM (2008). Some aspects of avocado (*Persea americana* Mill.) diversity and domestication in Mesoamerica. Genet Resour Crop Evol 55: 441-450.

Gross-German E, Viruel MA (2013). Molecular characterization of avocado germplasm with a new set of SSR and EST-SSR markers: genetic diversity, population structure, and identification of race-specific markers in a group of cultivated genotypes. Tree Genet Genomes 9: 539-555.

Kovach WL (1998). MVSP: A Multivariate Statistical Package for Windows. Version 3.0. Pentraeth, UK: Kovach Computing Services.

Kumar S, Dudley J, Nei M, Tamura M (2008). MEGA: A biologist centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9: 299-306.

Kwame A, Akromah R, Ofori FA, Takrama JF, Saada D, Bitton I, Lavi U (2008). Genetic characterization of Ghanaian avocadoes using microsatellite markers. J Am Soc Hortic Sci 133: 801-809.

Mhameed S, Sharon D, Kaufman D, Lahav E, Hillel J, Degani C, Lavi U (1997). Genetic relationships within avocado (*Persea americana* Mill.) cultivars and between *Persea* species. Theor Appl Genet 94: 279-286.

Peakall R, Smouse PE (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28: 2537-2539.

Piry S, Luikart G, Cornuet JM (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90: 502-503.

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

Raymond M, Rousset F (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86: 248-249.

Rodríguez M, Jaramillo JG, Orozco J (2009). Colecta de Aguacates criollos colombianos como base para iniciar programas de fitomejoramiento que contribuyan a su competitividad. In: Memorias III Congreso Latinoamericano de Aguacate (in Spanish).

Schaffer B, Gil PM, Mickelbart MV, Whiley AW (2013). Ecophysiology. In: Schaffer B (editor). The Avocado: Botany, Production, and Uses. 2nd ed. Wallingford, UK: CABI Publishing, pp. 168-199.

Schnell RJ, Brown JS, Olano CT, Power EJ, Krol CA, Kuhn DN, Motamayor JC (2003). Evaluation of avocado germplasm using microsatellite markers. J Am Soc Hortic Sci 128: 881-889.

Sharon D, Cregan PB, Mhameed S, Kuharska M, Hillel I, Lahav E, Lavi U (1997). An integrated genetic linkage map of avocado. Theor Appl Genet 95: 911-921.

Sneath PH, Sokal RR (1973). Numerical Taxonomy. San Francisco, CA, USA: Freeman.

Storey WB, Bergh BO, Zentmyer, GA (1986). The origin, indigenous range, and dissemination of the avocado. California Avocado Society Yearbook 70: 127-133.

Torres AM, Bergh BO (1978). Isozymes as indicators of out-crossing among 'Pinkerton' seedlings. California Avocado Society Yearbook 62: 103-110.

Ying Z, Davenport TL, Faber B, Zhang T (2009). Reevaluation of the role of honeybees and wind on pollination of avocado. J Hortic Sci Biotechnol 83: 255-260.