

Full Length Research Paper

Microsatellites revealed genetic diversity and population structure in Colombian avocado (*Persea americana* Mill.) germplasm collection and its natural populations

Gloria Patricia Cañas-Gutiérrez^{1*}, Rafael Eduardo Arango-Isaza² and Clara Inés Saldamando-Benjumea²

¹Laboratorio de Control Biológico y Fitosanidad, Corporación para investigaciones Biológicas, Medellín, Antioquia, Colombia.

²Escuela de Biociencias, Facultad de Ciencias, Universidad Nacional de Colombia, sede Medellín, Antioquia, Colombia.

Received 28 November, 2018; Accepted 20 February, 2019

Avocado (*Persea americana* Mill.) is one of the most consumed fruits around the world. The species is differentiated into three botanical races: Mexican, Guatemalan and West Indian. A genetic characterization of a Colombian avocado germplasm collection (N = 105) preserved since 2006 and 92 avocado “*criollo*” trees sampled in Antioquia was made with 12 microsatellites. Colombian avocado exhibits higher genetic diversity (13.9 alleles/locus, $H_o = 0.79-0.38$, $N_a = 12.3-2.4$, $N_e = 5.3-1.8$, $I = 1.87-0.68$) than avocados from other germplasms as *criollo* trees are the product of free crossing (multiple hybridizations). Population structure was found within Antioquian *criollo* avocados ($F_{ST} = 0.13$, $p < 0.0001$) and in the germplasm collection ($F_{ST} = 0.068$, $p < 0.0001$). Divergence in Antioquia resulted from differences in elevation and climatic conditions. Southwest, East and Altiplano-norte Antioquian sub-regions are genetically close and produce *criollo* avocados in high elevations where grafting with Hass is likely to be successful. In Colombia, avocado genetic differences were enhanced by the Andean Chain of Mountains. STRUCTURE HARVESTER assignment revealed that germplasm avocados were distributed in $K = 2$. The first cluster was composed of samples from the south of Colombia (Valle del Cauca, Nariño), the second by samples from the north (Antioquia, Bolivar, Magdalena) and avocados from the rest of the departments were distributed in both. The results obtained are relevant for avocado certification because knowledge of genotypic and phenotypic variation is crucial for crop management and grafting with avocado cultivars such as Hass, which is economically important for Colombian avocado exportation.

Key words: Avocado, *Persea americana*, germplasm, genetic diversity, microsatellites.

INTRODUCTION

Genbank collections are of great importance because information about its structure is relevant for genetic

*Corresponding author. E-mail: maaroufbrahim@gmail.com.

Author(s) agree that this article remain permanently open access under the terms of the [Creative Commons Attribution License 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

resources conservation and utilization of plant material (Odong et al., 2011). Conservation and use of plant genetic resources include material, that is, both reproductively and vegetatively propagated. This material consist of: (a) varieties that are currently used as new plant cultivars, (b) obsolete cultivars, (c) races and cultivars traditionally used by agriculturists, (d) wild relatives of plant species and (e) genetic lineages that include elite materials and improved cultivars of species as well as aneuploids plants (Frankel et al., 1995; Karp et al., 1997; Corona-Jácome et al., 2016; Guzmán et al., 2017). Estimation of genetic structure of heterogeneous germplasms is essential for core collections sampling as partitioning before sampling ensures that genetic, phenotypic and ecological spectra are entirely represented (Odong et al., 2011).

In avocado (*Persea americana* Mill.) ($2n = 2x = 24$), the species has great genetic variability with almost limited leveraging since cross fertilization between trees occur freely in nature through open pollination enhanced by insects (Sánchez, 1999; Barrientos-Priego and López, 2007). Genetic characterization of avocado germplasm as well as knowledge of its genetic diversity allow researchers for agronomic improvement through the development of new varieties and cultivars that are resistant to diseases and soil salinity and cultivars that are more profitable in terms of fruit production, fruit quality and fruit maturation precocity (Guzmán et al., 2017; Sánchez, 1999). This species is a subtropical evergreen tree native to Central America and Mexico (Chanderbali et al., 2008; Chen et al., 2009; Ashworth and Clegg, 2003; Corona-Jácome et al., 2016; Guzmán et al., 2017). It belongs to the Lauraceae family and the Magnoliid clade of early-divergent angiosperms and is differentiated in three horticultural races (Chanderbali et al., 2008; Chen et al., 2009; Ashworth and Clegg, 2003; Gross-German and Viruel, 2013; Corona-Jácome et al., 2016) that are adapted to different climatic conditions and recognized as: Mexican [*P. americana* var. *drymifolia* (Schltdl. & Cham.) S.F.Blake], Guatemalan [*P. americana* var. *guatemalensis* L. Wms.], and West Indian [*P. americana* var. *americana* Mill.]. These races originated from highland Mexico, highland Guatemala, and lowland Mexico (Tierras bajas), respectively (Corona-Jácome et al., 2016) and are distinguishable on the basis of morphological, physiological and horticultural traits (Bergh, 1992; Alcaraz and Hormaza, 2007; Gross-German and Viruel, 2013). Mexican and Guatemalan races are adapted to cooler climates while West Indian races to tropical conditions (Gross-German and Viruel, 2013). In Colombia, most commercial avocados are the product of interracial hybrids developed from seedlings that are commonly recognized as “criollo” (Bernal and Bernal et al., (2014); Cañas et al., 2015). In this country, avocado production is mainly based on grafting selected *criollo* trees with commercial cultivars, particularly the variety Hass (Cañas et al., 2015; Rodríguez et al., 2009).

However, *criollo* avocados have not been certified yet and Colombian avocado exportation requires genetic and phenotypic characterizations as well as avocado fruit production free of pathogens (that is, *Phytophthora cinnamomi*) and pests (that is, Thrips such as *Selenothrips rubrocinctus* Giard and *Frankliniella gardeniae* Moulton, and individuals of the genus: *Leptothrips* and *Karnyothrips* species (Echeverri-Florez et al., 2004)). Previous molecular characterization of “criollo” trees and few cultivars including Hass, Fuerte and Papelillo from the Department of Antioquia (Northeast Colombia) was made using 38 amplified fragment length polymorphism (AFLP) (Cañas et al., 2015). They found genetic similarities between *criollo* avocado trees and the variety Hass (AFLP fingerprints shared between Hass and *criollo* trees) and population structure ($F_{st} = 0.137$, $p < 0.01$) between Antioquian municipalities (collection sites) according to differences in elevation and climate conditions. In avocado, as in other fruits, morphological traits have been traditionally used to identify genotypes (Guzmán et al., 2017; Alcaraz and Hormaza, 2007; IPGRI, 1995). Nevertheless, evaluation of these traits is laborious and inexact due to the influence of environmental effects, subjectivity and a limited number of discriminatory traits (Alcaraz and Hormaza, 2007). For this reason, molecular markers such as microsatellites have been used to clarify the genetic similarities between avocado races and cultivars (Sharon et al., 1997; Alcaraz and Hormaza, 2007; Borrone et al., 2007; Gross-German and Viruel, 2013; Guzmán et al., 2017). These markers, also known as simple sequence repeats (SSRs), are stretches of DNA that consist of tandem repeats of 1 to 6 bp. They are located throughout nuclear, mitochondrial, and chloroplast genomes and represent suitable molecular markers to perform genetic studies as they are highly reproducible and co-dominant (Freeland, 2005). Several microsatellite loci have been developed in avocado, as they have been used for linkage map construction (Sharon et al., 1997; Borrone et al., 2007), molecular germplasm characterizations (Alcaraz and Hormaza, 2007; Guzmán et al., 2017) and diversity analysis (Alcaraz and Hormaza, 2007; Borrone et al., 2007; Gross-German and Viruel, 2013; Corona-Jácome et al., 2016; Guzmán et al., 2017).

Given that no studies on the molecular characterization of an avocado germplasm from Colombia ($N = 105$, *ex situ* trees from diverse departments) preserved since 2006 together with *criollo* avocado trees sampled (from 2014 to 2015, $N = 92$) at the Department of Antioquia (Northeast Colombia), have been carried out in Colombia; the aim of this study was to employ 12 microsatellites previously standardized in the species (Alcaraz and Hormaza, 2007) to analyze its genetic diversity and its population genetics to determine if Colombian avocado is partitioned due to environmental effects such as elevation as previously found in Antioquia



Figure 1. Department of Antioquia with all sample sites (municipalities) used for leaves collections where leaves were collected.

(Cañas et al., 2015). Elevation plays an important role in avocado, as Mexican and Guatemalan races are usually found in highlands and cool environments while West Indian races in lowlands and tropical conditions (Gross-German and Viruel, 2013). The results obtained in this study are relevant as they represent the first step taken for Colombian avocado certification also for port grafting *criollo* trees with the variety Hass, which is a product of a cross between Mexican × Guatemalan races (Chen et al., 2009) and one of the most desired avocados in the world (Chanderbali et al., 2008).

MATERIALS AND METHODS

Plant

Fresh leaves collections were carried out in 197 cultivars including: 92 trees corresponding to *criollo* trees sampled at the Department of Antioquia (Northeast Colombia). These collections were made during 2014 and 2015, together with 105 cultivars held in an *ex situ* germplasm since 2006 in AGROSAVIA (formerly called CORPOICA, Corporación Colombiana de Investigación Agropecuaria) at the municipality of Palmira (Valle del Cauca Department, Southwest Colombia). Samples from Antioquia were divided in six sub-regions, and composed of 27 municipalities that are distributed in different elevations (Figure 1 and Table 1).

Germplasm collections were composed of 105 *criollo* trees that were collected in 12 Colombian Departments including: Bolívar, N = 9; Caldas, N = 4; Cauca, N = 4; Cesar, N = 2; Huila, N = 4; Magdalena, N = 12; Nariño, N = 10; Quindío, N = 9; Risaralda, N = 6; Santander, N = 15; Tolima, N = 7 and Valle del Cauca, N = 23 (Figure 2 and Table 2).

Genomic DNA extraction and amplification

Total genomic DNA was isolated from avocado leaves based on Cetyl Tri-methyl Ammonium Bromide (CTAB) extraction method previously standardized in Colombia (Cañas et al., 2015). Twelve microsatellites were selected from those used by Alcaraz and Hormaza (2007) based on their high polymorphism. Three multiplex PCR amplifications were performed in 10 µl volume containing 16 mM (NH₄)₂SO₄, 67 mM Tris-HCl pH 8.8, 0.01% Tween20, 2 mM MgCl₂, 0.1 mM each dNTP, 0.4 µM of each primer, 25 ng genomic DNA and GoTaq[®] ADN polymerase (Promega, WI, USA). Forward primers were labeled with WellRed fluorescent dyes on the 5' end (Proligo, France). Reactions were carried out in an I-cycler (Bio-Rad Laboratories, Hercules, CA, USA) thermo cycler using the following temperature profile for multiplex PCR 2 and 3: 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. For multiplex PCR 1, same temperature profiles were used with the exception of 55°C for annealing temperature was used instead. The PCR products were analyzed by capillary electrophoresis in the equipment ABI PRISM[®] 3130 Genetic Analyzer (Applied Biosystem, CA, USA).

Table 1. Avocado *criollo* trees sampled in six sub-regions of Antioquia during 2014 and 2015.

ID number	Municipality	Sub-region	Elevation (masl)
ANSS5	Sonsón	East	2238
ANSS6	Sonsón	East	2277
ANSS7	Sonsón	East	2244
ANSS8	Sonsón	East	1841
ANSS9	Sonsón	East	1823
ANSS10	Sonsón	East	1806
ANSS11	Sonsón	East	1918
ANSS45	Sonsón	East	2236
ANSS81	Sonsón	East	1905
ANAB12	Abejorral	East	847
ANAB13	Abejorral	East	978
ANAB22	Abejorral	East	1747
ANAB23	Abejorral	East	1843
ANAB24	Abejorral	East	1909
ANAB42	Abejorral	East	2121
ANAB43	Abejorral	East	2109
ANAB46	Abejorral	East	876
ANAB98	Abejorral	East	1902
ANRN84	Rionegro	East	2144
ANRN91	Rionegro	East	2133
ANRN92	Rionegro	East	2138
ANRN93	Rionegro	East	2140
ANRN94	Rionegro	East	2131
ANRN95	Rionegro	East	2136
ANRN96	Rionegro	East	2136
ANSB14	Santa Barbara	Southwest	1022
ANSB15	Santa Barbara	Southwest	1022
ANSB16	Santa Barbara	Southwest	1022
ANSB17	Santa Barbara	Southwest	883
ANSB18	Santa Barbara	Southwest	922
ANSB19	Santa Barbara	Southwest	971
ANSB20	Santa Barbara	Southwest	927
ANSB21	Santa Barbara	Southwest	863
ANSB47	Santa Barbara	Southwest	845
ANSB72	Santa Barbara	Southwest	1010
ANSB73	Santa Barbara	Southwest	1017
ANCA25	Caramanta	Southwest	2094
ANCA28	Caramanta	Southwest	1956
ANCA29	Caramanta	Southwest	1956
ANVA30	Valparaiso	Southwest	1921
ANVA32	Valparaiso	Southwest	1950
ANVA82	Valparaiso	Southwest	1954
ANVA83	Valparaiso	Southwest	1950
ANVE26	Venecia	Southwest	757
ANVE27	Venecia	Southwest	755
ANJA31	Jardín	Southwest	1993
ANJA35	Jardín	Southwest	1881
ANJA36	Jardín	Southwest	1904
ANJA37	Jardín	Southwest	1945
ANJA38	Jardín	Southwest	1921
ANJA40	Jardín	Southwest	1797

Table 1. Contd.

<i>ANJE33</i>	Jerico	Southwest	1909
<i>ANJE34</i>	Jerico	Southwest	1909
<i>ANRR85</i>	Urrao	Southwest	1813
<i>ANRR86</i>	Urrao	Southwest	1900
<i>ANRR87</i>	Urrao	Southwest	1933
<i>ANRR88</i>	Urrao	Southwest	1921
<i>ANRR89</i>	Urrao	Southwest	1979
<i>ANRR89M</i>	Urrao	Southwest	1979
<i>ANAR48</i>	Arboletes	Urabá	89
<i>ANAR49</i>	Arboletes	Urabá	100
<i>ANAR49M</i>	Arboletes	Urabá	96
<i>ANSJ50</i>	San Juan	Urabá	230
<i>ANNE55</i>	Necocli	Urabá	204
<i>ANNE56</i>	Necocli	Urabá	58
<i>ANTU57</i>	Turbo	Urabá	35
<i>ANTU57M</i>	Turbo	Urabá	35
<i>ANAP58</i>	Apartadó	Urabá	176
<i>ANAP58M</i>	Apartadó	Urabá	232
<i>ANCA59</i>	Carepa	Urabá	337
<i>ANCA66</i>	Carepa	Urabá	335
<i>ANMU63</i>	Mutata	Urabá	128
<i>ANCH64</i>	Chigorodo	Urabá	50
<i>ANCH65</i>	Chigorodo	Urabá	44
<i>ANCG53</i>	Cañas Gordas	West	957
<i>ANCG54</i>	Cañas Gordas	West	957
<i>ANDA62</i>	Dabeiba	West	471
<i>ANBU51</i>	Buriticá	West	1640
<i>ANGI52</i>	Giraldo	West	2040
<i>ANUR60</i>	Uramita	West	786
<i>ANUR61</i>	Uramita	West	804
<i>ANCI74</i>	Cisneros	Northwest	1057
<i>ANCI75</i>	Cisneros	Northwest	1055
<i>ANSR76</i>	San Roque	Northwest	890
<i>ANSR77</i>	San Roque	Northwest	873
<i>ANSR78</i>	San Roque	Northwest	1074
<i>ANCA80</i>	Caracolí	Northwest	631
<i>ANMA79</i>	Maceo	Northwest	652
<i>ANMA104</i>	Maceo	Northwest	642
<i>ANER100</i>	Entrerios	Altiplano norte	2457
<i>ANER101</i>	Entrerios	Altiplano norte	2493
<i>ANER102</i>	Entrerios	Altiplano norte	2353

Each reaction was repeated twice to minimize run-to-run variation. After amplifications, each putative genotype was identified with the program Peak Scanner Software v1.0 (Applied Biosystem, CA, USA).

Gene diversity and population genetics analyzes

Estimators such as: (a) N_a = number of different alleles, (b) N_e = number of effective alleles = $1/\sum p_i^2$ where p_i = allele frequencies,

(c) mean allele number per locus (microsatellite) = number of alleles/ N_x , N_x = number of genotyped individuals per population, (d) Observed (H_o) and expected heterozygosities: H_o = Observed Heterozygosity = No. of Hets / N , N = total number of genotyped individuals and H_e = Expected Heterozygosity = $1 - \sum p_i^2$ and finally (e) Shannon Weaver indexes ($I = \text{Shannon's Information Index} = -1 \times \sum (p_i \times \ln(p_i))$), were estimated from each sampled site (Antioquia and germplasm collection site separately) with the software GENALEX 6.501 (Peakall and Smouse, 2012). GENEPOP 4.2 (Raymond and Rousset, 1995) was used for exploring Hardy-

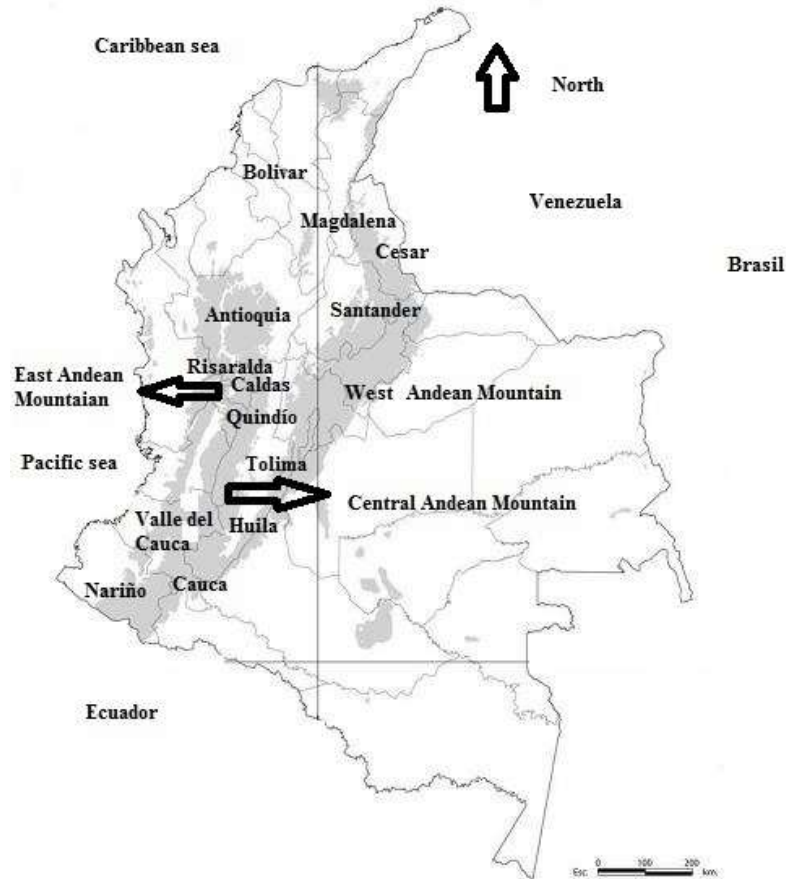


Figure. 2 Departments of Colombia where avocado samples were taken for the germplasm bank.

Weinberg equilibrium and linkage disequilibrium per population. Bonferroni corrections were carried out after obtaining HWE results and linkage disequilibrium estimations. These corrections were made to modify the significance levels (α) according to comparisons made amongst sub-regions or departments in the HW analysis and between paired loci for the linkage disequilibrium (Sokal and Rohlf, 1995). GENALEX 6.501 was also used to estimate analysis of molecular variance (AMOVA) to determine if avocados collected from different sampling sites (sub-regions) from Antioquia and/or different departments of Colombia (germplasm) were genetically different. This program also used to calculate Nei's genetic paired distances between sampled sub-regions of Antioquia and between departments of Colombia that are part of the germplasm bank. These distances were further used to graph neighbor joining (NJ) dendrograms with the program MEGA 4.0 (Kumar et al., 2008). No bootstrap was estimated for the NJ trees as the data entry consisted of Nei's pair distances calculated between sampled sites solely.

The model based clustering analysis STRUCTURE 2.3.4 (Pritchard et al., 2000) was used to assess the most probable avocado cluster membership from all samples maintained in the germplasm bank from Colombia (departments of Colombia) together with samples taken from Antioquia. This program was run for 15.000 Markov chain Monte Carlo (MCMC) steps after a burn-in period of 150.000 interactions from $K=1-15$. Each K was calculated from ten independent runs and also considering (a) no admixture

model and (b) allele frequency model independent following the procedure given by Chen et al. (2009). The ad hoc estimated likelihood of K (ΔK) (Evanno et al., 2015) was used to estimate the most likely number of populations (K) based on the rate of change in the log probability of the data [$\ln Pr(X/K)$]. The number of $K = 2$ genetic units was estimated with STRUCTURE HARVESTER web v.0.6.92 (Earl and vonHoldt, 2012).

RESULTS

Microsatellite polymorphisms and genetic diversity

A total of 147 fragments were detected with 12 microsatellites in avocados collected in the six sub-regions of Antioquia (Northeast Colombia). The number of alleles per locus ranged from 9 (AVT386, AVT372, AVD102) to 18 (AVD013). The mean allele number was 12.3 alleles per locus. The total number of alleles were high for the Antioquian Southwest (375 alleles) and West (244 alleles) sub-regions suggesting that both included the most diverse Antioquian avocados. On the contrary, Altiplano-norte only presented 13 alleles.

Table 2. *Criollo* cultivars hold in the Colombian avocado Genebank Genbank collection maintained at AGROSAVIA (Palmira, Valle del Cauca).

ID Number	Department
<i>BOCA-018</i>	Bolívar
<i>BOCA-015</i>	Bolívar
<i>BOCA-021</i>	Bolívar
<i>BOCA-022</i>	Bolívar
<i>BOCB67</i>	Bolívar
<i>BOCB68</i>	Bolívar
<i>BOSJ69</i>	Bolívar
<i>BOJN70</i>	Bolívar
<i>BOJN71</i>	Bolívar
<i>CAVI-030</i>	Caldas
<i>CAVI-041</i>	Caldas
<i>CAVI-030</i>	Caldas
<i>CAVI-041</i>	Caldas
<i>CACAJ-042</i>	Cauca
<i>CACAJ-044</i>	Cauca
<i>CACAJ-037</i>	Cauca
<i>CACAJ-036</i>	Cauca
<i>CEPU-008</i>	Cesar
<i>CELA-006</i>	Cesar
<i>HUGA-012</i>	Huila
<i>HUSU-004</i>	Huila
<i>HUAL-017</i>	Huila
<i>HUAL-018</i>	Huila
<i>MASA-026</i>	Magdalena
<i>MACI-005</i>	Magdalena
<i>MACI-018</i>	Magdalena
<i>MASA-001</i>	Magdalena
<i>MACI-010</i>	Magdalena
<i>MACI-017</i>	Magdalena
<i>MACI-025</i>	Magdalena
<i>MACI-019</i>	Magdalena
<i>MACI-021</i>	Magdalena
<i>MAFU-016</i>	Magdalena
<i>MACI-023</i>	Magdalena
<i>MAFU-048</i>	Magdalena
<i>NATU-008</i>	Nariño
<i>NATU-011</i>	Nariño
<i>NARIÑO 1</i>	Nariño
<i>NACHA-002</i>	Nariño
<i>NATU-009</i>	Nariño
<i>NATU-001</i>	Nariño
<i>NATU-003</i>	Nariño
<i>NACHA-003</i>	Nariño
<i>NATU-002</i>	Nariño
<i>NACHA-004</i>	Nariño
<i>QUIPI-003</i>	Quindío
<i>QUIAR-022</i>	Quindío
<i>QUICO-009</i>	Quindío
<i>QUIPI-001</i>	Quindío

Table 2. Contd.

<i>QUIQUI-029</i>	Quindío
<i>QUIBU-011</i>	Quindío
<i>QUICO-006</i>	Quindío
<i>QUICO-005</i>	Quindío
<i>QUICO-015</i>	Quindío
<i>RIPE-006</i>	Risaralda
<i>RIQUI-008</i>	Risaralda
<i>RIPE-007</i>	Risaralda
<i>RIPE-001</i>	Risaralda
<i>RIPE-003</i>	Risaralda
<i>RIPE-002</i>	Risaralda
<i>SALA-028</i>	Santander
<i>SACA-014</i>	Santander
<i>SASA-003</i>	Santander
<i>SASA-005</i>	Santander
<i>SASA-004</i>	Santander
<i>SASA-006</i>	Santander
<i>SASA-009</i>	Santander
<i>SARI-027</i>	Santander
<i>SARI-023</i>	Santander
<i>SACA-013</i>	Santander
<i>SACA-017</i>	Santander
<i>SACA-019</i>	Santander
<i>SALA-030</i>	Santander
<i>SARI-021</i>	Santander
<i>SALA-031</i>	Santander
<i>TOCHA-025</i>	Tolima
<i>TOMA-003</i>	Tolima
<i>TOFRE-019</i>	Tolima
<i>TOIB-050</i>	Tolima
<i>TOAL-004</i>	Tolima
<i>TOIB-030</i>	Tolima
<i>TOAL-005</i>	Tolima
<i>VAPA-066</i>	Valle del Cauca
<i>VAPA-063</i>	Valle del Cauca
<i>VAPA-061</i>	Valle del Cauca
<i>VABU-014</i>	Valle del Cauca
<i>VABU-010</i>	Valle del Cauca
<i>VARO-003</i>	Valle del Cauca
<i>VABU-024</i>	Valle del Cauca
<i>VAPA-062</i>	Valle del Cauca
<i>VARO-001</i>	Valle del Cauca
<i>VABU-021</i>	Valle del Cauca
<i>VABU-016</i>	Valle del Cauca
<i>VARO-006</i>	Valle del Cauca
<i>VARO-002</i>	Valle del Cauca
<i>VAPA-064</i>	Valle del Cauca
<i>VAPA-026</i>	Valle del Cauca
<i>VABU-022</i>	Valle del Cauca
<i>VAPA-021</i>	Valle del Cauca
<i>VARO-020</i>	Valle del Cauca
<i>VAPA-020</i>	Valle del Cauca

Table 2. Contd.

VAPA-070	Valle del Cauca
VAPA-060	Valle del Cauca
VAYO-098	Valle del Cauca
VAPA-056	Valle del Cauca

This result could be due to the fact that the sampled site had the smallest population size ($N = 3$) (Table 3). Likewise, the number of different alleles (N_a) were 8.50 for both West and Southwest regions and $N_a = 1.42$ for Altiplano-norte. The number of effective alleles (N_e) was 4.6 for West sub-region and 1.4 for Altiplano-norte. Shannon-Weaver indexes were from 1.7 for the West sub-region to 0.43 for Altiplano-norte. H_o (observed heterozygosity) ranged from 0.59 (West) to 0.52 (Altiplano-norte) and H_e (expected heterozygosity) from 0.76 (East) to 0.30 (Altiplano-norte) (Table 3). Diversity parameters obtained from the germplasm collection (12 Colombian departments) generated 167 amplified fragments with the following information: the number of amplified alleles was from 10 (AVD102) to 21 (AVD006), with an average of 13.9 alleles per locus. The total number of amplified alleles per population was $N = 969$ for Antioquia. This Colombian department presented the highest number of alleles. However, it was the most sampled department of Colombia. Valle del Cauca also had a high number of amplified alleles with $N = 217$ and Cesar Department only had 20 amplified alleles. The total number of different alleles (N_a) was 12.3 for Antioquia with the highest value and 2.3 for Caldas with the lowest value. Shannon-Weaver index ranges were from 0.7 for Valle del Cauca to 1.9 for Antioquia. Observed (H_o) and expected heterozygosities (H_e) ranges were from 0.4 (Cauca Department) to 0.70 (Risaralda Department) (Table 3).

HWE was found in 30% of the six Antioquian sub-regions after Bonferroni corrections and in 50% of the 13 departments (given that Antioquia was also included in the analysis) (Table 4). Analysis of linkage disequilibrium showed that microsatellites were in linkage equilibrium after Bonferroni corrections (data not shown).

Population genetics results

The AMOVA analyses estimated between six Antioquian sub-regions produced $F_{ST} = 0.084$, $p < 0.0001$ and between 13 Colombian departments (germplasm collection together with samples from Antioquia) generated $F_{ST} = 0.068$, $p < 0.0001$. Both values suggest population differentiation between sub-regions and between departments. NJ dendrograms obtained with Nei's genetic distances between sub-regions (Figure 3) and between departments' (Figure 4) clustered

populations that are geographically close. In Antioquia, clustering occurred between sites that are located in similar elevations and climatic conditions (Table 1). The NJ dendrogram obtained for Antioquia separated Urabá and Altiplano-norte from the rest of sub-regions. Elevations were from 35 to 337 masl in Urabá, from 804 to 2040 masl in West sub-region, from 873 to 1057 masl in Northwest sub-region, from 833 to 2094 masl in Southwest sub-region, from 876 to 2277 masl in East sub-region and finally in Altiplano-norte elevations were from 2353 to 2493 masl. Genetic differentiation between departments might also be influenced by the three Andean Mountains Chains as clustering between sampling sites that occurred in locations found in the same mountain chain suggesting these mountains are a barrier to gene flow in this species as avocado producers share seedlings between neighbor orchards (Figure 2). The only sampling site from Antioquia that was genetically apart from the rest was Altiplano-norte. This genetic differentiation might be due to its small size.

The program STRUCTURE HARVESTER assigned $K = 2$ from the 12 departments of Colombia that composed of the germplasm bank together with samples collected in Antioquia according to the following estimations: Mean $\ln(P|D) = -7267.10$, Stdev $\ln P(K) = 0.1870$, $\ln'(K) = 410.44$, and $\Delta K = 773.13$ (Figure 5). These results suggest that Colombian *criollo* avocado is sub-divided into two populations where most avocado samples taken from the south of Colombia were found in cluster 1, including the departments of Valle del Cauca and Nariño and samples taken at the north of Colombia in cluster 2, including the departments of Antioquia, Bolivar and Magdalena. The rest of the departments were evenly distributed in both clusters (Table 5). STRUCTURE HARVESTER also subdivided *criollo* avocados from Colombia in a similar clustering pattern to the NJ dendrogram obtained with Nei's distances as this dendrogram separated samples according to their geographic proximity and their distribution in the Andean Chain on mountains.

DISCUSSION

P. americana constitutes one of the most consumed fruits around the world (Guzmán et al., 2017; Gross-German and Viruel, 2013; Chen et al., 2009; Chanderbali et al., 2008). This study represents the first molecular analysis made with microsatellites in Colombian *criollo* trees collected in nature together with cultivars obtained from an *ex situ* germplasm preserved since 2006. In Colombia, few studies have focused on analyzing this species at the molecular level. In avocado, open pollination increases the genetic variability of the species (Sánchez, 1999; Barrientos-Priego and Lopez, 2007) as in the case of *criollo* cultivars in Colombia. *Criollo* cultivars are distributed in a wide range of thermic zones and are

Table 3. Genetic diversity estimators obtained from *criollo* avocados collected in (a) Antioquia department (six sub-regions) and (b) maintained at the Genebank collection in AGROSAVIA.

(a) Sub-region	Ni	Nt		N	Na	Ne	I	Ho	He	F
East	25	244	Mean	20.333	8.500	4.566	1.705	0.540	0.758	0.284
			SD	1.170	0.723	0.447	0.097	0.055	0.023	0.072
Southwest	36	375	Mean	31.250	8.500	4.288	1.605	0.518	0.722	0.287
			SD	1.181	0.866	0.613	0.119	0.049	0.030	0.059
Uraba	15	161	Mean	13.417	6.000	3.458	1.361	0.521	0.649	0.189
			SD	0.701	0.640	0.537	0.119	0.053	0.043	0.076
West	7	81	Mean	6.750	5.667	4.272	1.489	0.585	0.715	0.203
			SD	0.131	0.607	0.550	0.127	0.087	0.041	0.094
Northwest	8	76	Mean	6.333	4.000	2.662	1.003	0.568	0.530	-0.104
			SD	0.829	0.564	0.426	0.149	0.095	0.066	0.147
Altiplanonorte	3	13	Mean	1.083	1.417	1.400	0.433	0.528	0.301	-0.786
			SD	0.288	0.288	0.285	0.115	0.139	0.078	0.114
(b) Department										
Antioquia	96	969	Mean	80.750	12.250	5.262	1.873	0.537	0.785	0.319
			SD	3.908	0.970	0.611	0.100	0.044	0.020	0.050
Bolívar	9	85	Mean	7.083	3.750	2.604	1.057	0.434	0.589	0.283
			SD	0.633	0.305	0.196	0.078	0.074	0.034	0.109
Caldas	4	30	Mean	2.500	2.333	2.278	0.809	0.667	0.458	-0.483
			SD	0.557	0.527	0.522	0.183	0.142	0.100	0.070
Cauca	4	33	Mean	2.750	2.417	1.801	0.683	0.375	0.378	-0.003
			SD	0.538	0.514	0.390	0.166	0.104	0.090	0.112
Cesar	2	20	Mean	1.667	2.667	2.472	0.895	0.667	0.531	-0.267
			SD	0.188	0.333	0.319	0.131	0.112	0.074	0.127
Huila	4	34	Mean	2.833	2.750	2.249	0.868	0.618	0.512	-0.189
			SD	0.322	0.329	0.277	0.112	0.106	0.060	0.143
Magdalena	12	107	Mean	8.917	4.417	3.055	1.168	0.549	0.601	0.111
			SD	1.090	0.514	0.346	0.140	0.081	0.065	0.074
Nariño	10	95	Mean	7.917	5.833	4.471	1.570	0.576	0.754	0.242
			SD	0.712	0.458	0.389	0.087	0.071	0.024	0.086
Quindío	9	93	Mean	7.750	6.333	4.827	1.648	0.569	0.771	0.269
			SD	0.552	0.607	0.475	0.094	0.064	0.021	0.079
Risaralda	6	61	Mean	5.083	4.500	3.341	1.319	0.700	0.686	-0.012
			SD	0.379	0.261	0.217	0.060	0.053	0.021	0.059
Santander	15	138	Mean	11.500	5.750	3.400	1.315	0.453	0.628	0.271
			SD	0.839	0.566	0.531	0.130	0.059	0.052	0.063
Tolima	7	74	Mean	6.167	4.833	3.423	1.347	0.587	0.694	0.145
			SD	0.345	0.405	0.237	0.072	0.059	0.020	0.092
Valle del Cauca	23	217	Mean	18.083	8.167	5.205	1.774	0.599	0.790	0.246
			SD	1.454	0.787	0.518	0.093	0.041	0.017	0.046

Ni = Sample size, Nt = total alleles, N = mean number of alleles, Na = number of different alleles, Ne = number of effective alleles, I = Shannon Index, Ho = observed heterozygosity, He = expected heterozygosity, F = Fixation index, SD = Standard deviation.

Table 4. Hardy Weinberg equilibrium results obtained for (a) six sub regions of Antioquia and (b) the avocado Genbank maintained in AGROSAVIA.

(a) Sub-region	P	SD	Switches (ave.)	Bonferroni Test		Sig
				$\alpha=0.05/n$	α	
East	0.0000	0.0000	7012.33	0.05/6	0.0083	*
Southwest	0.0000	0.0000	7875.25	0.05/5	0.01	*
Urabá	0.0000	0.0000	8630.17	0.05/4	0.0125	*
West	0.0000	0.0000	8352.17	0.05/3	0.016	*
Northeast	0.0399	0.0044	11291.10	0.05/2	0.025	ns
Altiplano norte	0.5683	0.0036	15582.00	0.05/1	0.05	ns
(b) Department	P	SD	Switches (ave.)	$\alpha=0.05/n$	α	Sig.
Antioquia	0.0000	0.0000	7454.92	0.05/13	0.0038	*
Bolivar	0.0001	0.0000	16431.67	0.05/12	0.0041	*
Caldas	1.000	0.0000	14617.38	0.05/11	0.0045	ns
Cauca	0.1050	0.0043	9966.57	0.05/10	0.005	ns
Cesar	0.1758	0.0035	7024.22	0.05/9	0.0055	ns
Huila	0.6658	0.0044	12396.56	0.05/8	0.0062	ns
Magdalena	0.0125	0.0019	14420.91	0.05/7	0.0071	ns
Nariño	0.0000	0.0000	8394.33	0.05/6	0.0083	*
Quindio	0.0000	0.0000	7753.00	0.05/5	0.01	*
Risaralda	0.1048	0.0049	8489.08	0.05/4	0.0125	ns
Santander	0.0000	0.0000	7773.42	0.05/3	0.016	*
Tolima	0.0009	0.0002	10254.83	0.05/2	0.025	*
Valle del Cauca	0.0000	0.0000	8910.58	0.05/1	0.05	*

P = Probability value, SD = Standard deviation, α = significance value, Sig = significant probability.

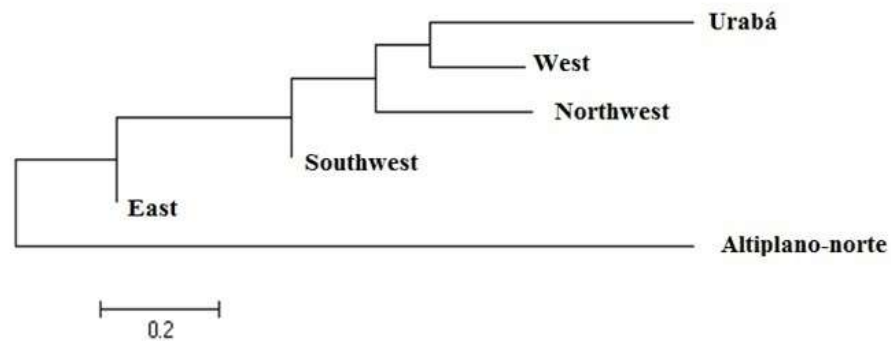


Figure 3. NJ dendrogram with Nei Nei's genetic paired distances for Antioquia sub-regions. East = 25 (Sonsón, Abejorral, Río Negro), Southwest = 34 (Caramanta, Valparaiso, Venecia, Jardín, Jerico, Urrao), West= 7 (Cañas Gordas, Dabeiba, Buriticá, Giraldo, Uramita), Northeast = 8 (Cisneros, San Roque, Caracolí, Maceo, Entre Ríos), Urabá 15 (Arboletes, San Juan, Necocli, Turbo, Apartadó, Carepa, Mutatá, Chigorodó) and Altiplano-norte = 3 (Entre Ríos).

greatly desired in this country for domestic consumption. Also, given their rusticity, adaptability and compatibility with several cultivars, they have been used as port grafting material, particularly the variety Hass (Cañas et al., 2015). Ninety two avocado *criollo* trees were sampled

from 6 Antioquian sub-regions. However, collection sizes were unevenly distributed as avocado production differs per municipality. These cultivars generated 147 microsatellites (alleles) that ranged from 9 to 18 alleles with a mean of 12.25 alleles per microsatellite whereas

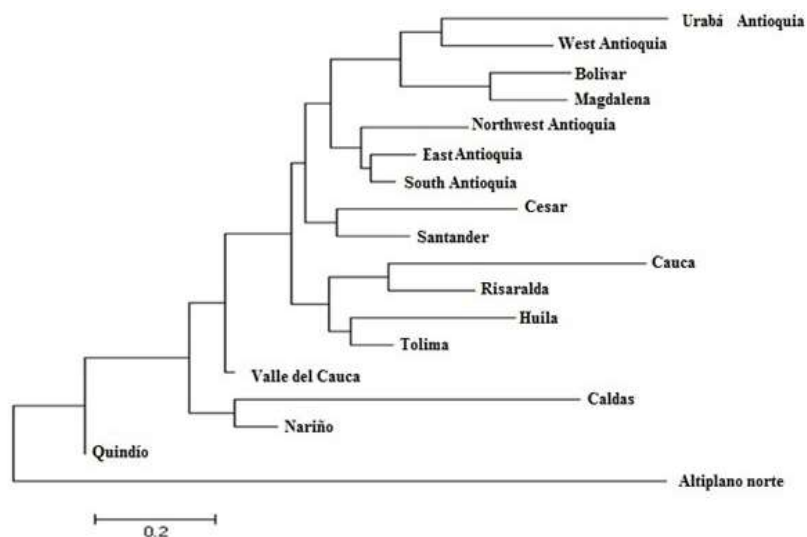


Figure 4. NJ dendrogram generated with Nei Nei's genetic paired distances for Colombian departments and Antioquian sub-regions (Urabá, West, Northwest, East, South, Altiplano-norte).

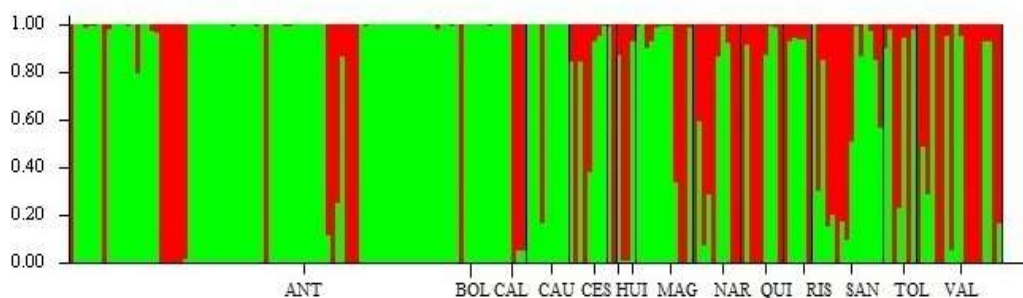


Figure 5 STRUCTURE sub- divisions generated for 12 departments of Colombia where avocado samples were taken, K = 2. Cluster 1 = Red, Cluster 2 = Green. ANT= Antioquia, BOL = Bolívar, CAL= Caldas, CES = Cesar, HUI = Huila, MAG = Magdalena, NAR = Nariño, QUI = Quindío, RIS = Risaralda, SAN = Santander, TOL = Tolima, VAL = Valle del Cauca.

Table 5. Assignment probability obtained with STRUCTURE for the Colombian avocado Genbank (N = 105).

Population	Department	Cluster 1 (Red)	Cluster 2 (Green)	N
1	Antioquia	0.197	0.803	96
2	Bolívar	0.092	0.908	9
3	Caldas	0.379	0.621	8
4	Cesar	0.501	0.499	2
5	Huila	0.543	0.457	4
6	Magdalena	0.237	0.763	12
7	Nariño	0.624	0.376	10
8	Quindío	0.58	0.42	9
9	Risaralda	0.373	0.627	6
10	Santander	0.43	0.57	15
11	Tolima	0.421	0.579	7
12	Valle del Cauca	0.623	0.377	18

criollo cultivars sampled from the germplasm (that consists of *ex situ* trees of 12 Colombian departments) generated 167 fragments with a range from 10 to 21 alleles and a mean value of 13.9 alleles per microsatellite. This result could be explained since a greater number of avocados were included in the germplasm. Similar outcomes were obtained by Abraham et al. (2014) from a microsatellite characterization of avocado from Ghana as they observed 115 alleles in total and in average 11.5 alleles per locus implying a great genetic diversity of this crop in Ghana as found in Colombia. Estimations of mean allele number obtained for Colombian avocado was higher than in Alcaraz and Hormaza (2007) as they found 9.75 alleles per locus with 16 microsatellites. They analyzed 75 *ex situ* accessions hold in the germplasm collection maintained at La Estación Experimental La Mayora in Málaga (Spain) that included cultivars from Canary Islands, Chile, Guatemala, Mexico, and USA. Other studies made with microsatellites in avocado by Borrone et al. (2007) showed values from 2 to 17 alleles in the three botanical races and in average 7.1 alleles per locus. In Borrone et al. (2007), they analyzed 24 individuals of *P. americana americana*, *P. a guatemalensis*, *P. a. drymifolia* as well as the species *Persea schiedeana*. Also, in Gross-German and Viruel (2013), they developed 47 new microsatellites in 42 avocado accessions from Spain showing a total of 455 alleles and in average 11.4 alleles per locus. One of the most recent studies made with microsatellites was carried out by Guzmán et al. (2017) in Mexico and they found a total of 548 alleles and from 4 to 34 alleles and in average 19.1 in 318 avocado accessions. The results obtained here suggest that *criollo* avocado from Colombia presents a high number of microsatellites alleles compared to other studies probably because avocados from this country are the product of multiple hybridizations that naturally occur through insect pollinations (Sánchez, 1999; Barrientos-Priego and Lopez, 2007). The number of samples used by Guzmán et al. (2017) was 318 avocados whereas in this study 197 avocados were analyzed. Although the number of samples used here are lower than in Guzmán et al. (2017), the number of alleles found in Colombian avocado are still very high.

Shannon-Weaver diversity index was between 1.87 and 1.16 in the germplasm collection meaning that genetic variability of avocado from Colombia is fairly homogenous amongst departments. Nevertheless, the highest estimation was obtained in Antioquia and the lowest in Cauca. These results could be explained because the majority of avocado samples came from Antioquia. Additionally, a comparison between observed (H_o) and expected heterozygosities (H_e) showed that municipalities from in Antioquia presented H_o from 0.52 (Altiplano-norte) to 0.59 (West) and H_e from 0.30 (Altiplano-norte) to 0.76 (East). In Colombia, H_o and H_e were from 0.4 (Cauca department) to 0.70 (Risaralda

department). Differences in heterozygosity estimations between departments could also be due to differences in sample sizes. These H_o and H_e values are relatively similar to the studies made by Ashworth and Clegg (2003), Alcaraz and Hormaza (2007), Borrone et al. (2007), Gross-German and Viruel (2013) and Guzmán et al. (2017). Ashworth and Clegg (2003) found $H_o = 0.61$ with 25 microsatellites standardized in the three botanical races and the species *Persea steyermarkii*. They also included the cultivars Fuerte and Bacon. In the study made by Borrone et al. (2007), they estimated H_o from 0.17 to 0.73 and included the three botanical races and *P. schiedeana*. Gross-German and Viruel (2013), produced H_o estimations were 0.67 in average from 42 accessions that were analyzed with 47 microsatellites from Spain and in the work made by Schnell et al. (2003) H_o values ranged from 0.73 to 0.94, they analyzed a larger sample size compared to the present work as they studied 224 accessions from a GenBank hold in Miami. Guzmán et al. (2017) estimated $H_o = 0.59$ and $H_e = 0.75$ from 318 accessions from Mexico where three botanical avocado races were included.

HW analysis showed that 67% of Antioquian sub-regions were in HWE including samples taken from East, Southeast, Urabá and West. Additionally, 53% departments of Colombia were also in HWE including samples collected at the departments: Antioquia, Bolívar, Nariño, Quindío, Santander, Tolima and Valle del Cauca. HWE is usually found when the studied marker is neutral and are widely distributed in a species genome (Freeland, 2005). Most avocado sampled sites from Colombia that were in HW disequilibrium had low sample sizes. Additionally, no linkage disequilibrium was found between pair of loci studied here. On the contrary, in Guzmán et al. (2017) HWE and linkage equilibrium were not found in the three botanical races. They observed high inter-locus linkage disequilibrium for two races: *drymifolia* (Mexican race) and *guatemalensis* (Guatemala race) and they suggest that this result was due to domestication bottleneck process experienced by each botanical race as in Mexico human intervention imposed selection intentionally to propagate certain lineages. In contrast, in Colombia most avocados that are produced and consumed are *criollo* and this type of crop is mostly composed of recombinant avocados (hybrids) as free pollination occurs freely in Colombian orchards.

Results obtained from AMOVA with six Antioquian sub-regions and between the Colombian departments that compose of the germplasm collection, were all significant, suggesting population structure in Colombian avocado. These results corroborate a previous study made with 111 *criollo* trees from Antioquia with 38 AFLP. They found genetic differentiation due to diverse agro-ecological conditions (elevation and climate) between sampled zones (Cañas et al., 2015). NJ dendrograms found here showed clustering between sampled sites that were geographically close at the six Antioquian sub-

regions and the 13 Colombian departments that compose of the avocado germplasm (including Antioquian samples). Genetic structure within Antioquia might have been due to differences in elevation and climate conditions between sites. Avocado samples collected in Southwest, East and Altiplano-norte were mainly found in high elevations meaning they are likely to be genetically similar to the variety Hass which is a product of a cross between Mexican × Guatemalan races. These two races are usually found in highlands (Gross-German and Viruel, 2013). Guzmán et al. (2017) demonstrated that Mexican avocado is genetically split in two groups due to differences in ecological conditions suggesting that not only human intervention but adaptation process is relevant for the genetics and production of this crop. Avocado genetic structure found in Colombia might also be enhanced by the three Andean Mountains Chains as the first NJ cluster was composed of avocados collected at the North of Colombia including the departments of Antioquia, Magdalena, Cesar and Santander. The second cluster was composed of samples collected in Cauca, Risaralda, Huila, Tolima and Valle del Cauca. These departments are found between the Central and the West Andean mountains. The third cluster composed of samples from Caldas, Nariño and Quindío, which are found between the Central and East Andean Mountains. On the other hand, avocados found in Altiplano-norte were genetically distant from other sub-regions of Antioquia.

Results obtained with STRUCTURE HARVESTER produced $K = 2$ from 12 departments analyzed here together with avocado samples collected in Antioquia ($N = 13$), this outcome suggests that Colombian avocado is split in two main clusters: the first group is mainly represented by avocados from the south of Colombia at Valle del Cauca and Nariño departments and the second cluster by departments found at the north of Colombia, including Antioquia, Bolivar and Magdalena. The rest of the departments: Caldas, Cesar, Huila, Nariño, Quindío, Risaralda, Santander, and Tolima were found in both clusters. Similar results were obtained with the NJ dendrogram of the germplasm collection as *criollo* avocados were grouped according to their geographic proximity and their distribution in the Andean Chain on Mountains. These mountains are a barrier to seedling movement as farmers share plants between neighbor orchards. These results differ to the study made by Gross-German and Viruel (2013) as they also used STRUCTURE in 42 Spanish avocado accessions genotyped with 47 microsatellites and found $K = 3$. However, they analyzed avocados from the three botanical races and did not analyzed different thermic zones whereas in this work only *criollo* trees were studied and both elevations and Andean Mountains sampling locations were considered; these *criollo* trees represent hybrids between these three races as they are the product of multiple hybridizations between them enhanced by insect pollinators (Sánchez, 1999;

Barrientos-Priego and López, 2007). Nevertheless, the results obtained here agreed with the outcome found by Guzmán et al. (2017) as they studied the three botanical races from 318 avocado accessions and estimated $K = 2$ that corresponded to two ecological regions of Mexico. The first cluster coincided with humid semi warm to humid semi cold conditions and the second cluster to humid semi warm to hot semi dry conditions.

Conclusion

This study represents the first step taken in Colombia to analyze the genetic diversity and population genetics with twelve microsatellites of an avocado germplasm ($N = 105$) together with samples obtained from *criollo* trees collected in Antioquia department (Northeast) ($N = 92$). Microsatellites used here were very useful as they were easily standardized, highly polymorphic among cultivars and established different genetic groups according to their geographic origin and elevation differences. Although new techniques based on next generation sequencing have been recently used to characterize crops, microsatellites are still very useful as in a sequence only four nucleotides are studied per nucleotide position, whereas in the case of microsatellites, many alleles can be found per locus meaning their polymorphism is high. Population genetic differentiation was found in Antioquia and in the germplasm avocado accessions suggesting the importance of climatic conditions and elevation to choose selecting genotypes for grafting purposes, particularly *criollo* avocados found in highlands are likely to be easily grafted with Hass cultivar as this avocado is a cross between Mexican × Guatemalan races that are mainly adapted to highlands (Gross-German and Viruel, 2013). Guzmán et al. (2017) analyzed 318 avocados from the three botanical races from Mexico and found that ecological conditions are important for genetic differentiation between these races as they might have evolved in two climatic condition regions. Population differentiation found here between Colombian departments was also influenced by the Andean Chain of Mountains suggesting that is a barrier to gene flow as avocado farmers share plant material between neighbors. Further studies are necessary to evaluate the results of grafting Hass variety with Colombian *criollo* trees including gene expression analysis, resistance to pathogens such as *P. cinnamomi*, resistance to soil salinity and soil drainage effects amongst other characteristics that are important for the improvement of avocado production in Colombia as a certified crop should be characterized at genotypic and phenotypic levels and free of pathogens and pests.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests

REFERENCES

- Alcaraz ML, Hormaza JI (2007). Molecular Characterization and genetic diversity in an avocado collection of cultivars and local Spanish genotypes using SSRs. *Hereditas* 144:244-253.
- Ashworth VE, Clegg MT (2003). Microsatellite markers in avocado (*Persea americana* Mill.): genealogical relationships among cultivated avocado genotypes. *Journal of Heredity* 94:407-415.
- Barrientos-Priego A, López-López L (2007). Historia y genética del aguacate. In: Memoria de la Fundación Salvador Sánchez Colín. Centro de Investigaciones Científicas y Tecnológicas del aguacate en el estado de México. Coatepec de Harinas, México 1:100-121.
- Borrone JW, Schnell RJ, Violi HA, Ploetz RC (2007). Seventy microsatellite markers from *Persea americana* Miller (avocado) expressed sequence tags. *Molecular Ecology Notes* 7:439-444.
- Bergh B (1992). The origin, nature and genetic improvement of the avocado. In: California Avocado society yearbook 76:61-75.
- Bernal J, Díaz-Diez C, Osorio-Toro C, Tamayo-Vélez A, Osorio-Vega W, Córdoba-Gaona O, Londoño-Zuluaga ME, Kondo-Rodríguez D, Tamamasuka D, Carabalí-Muñoz A, Varón-Devia E, Caicedo-Vallejo AM, Tamayo-Molano PJ, Sandoval-Aldana A, Forero-Longas F, García-Lozano J, Londoño-Bonilla M (2014). Actualización Tecnológica y Buenas Prácticas Agrícolas (BPA) en el Cultivo de Aguacate. 410 p. Ed. Corpoica. SUIMAGEN CREATIVA S.A.S. Colombia
https://conectarural.org/sitio/sites/default/files/documentos/Manual%20Actualizacion%20Tecnologica%20y%20BPA%20Cultivo%20de%20Aguacate_GOBERNACION%20PDF%20BAJA%20con%20caratulas.pdf
- Cañas-Gutiérrez GP, Galindo-López LF, Arango-Isaza RE, Saldamando-Benjumea CI (2015). Diversidad genética de aguacate (*Persea americana*) en Antioquia. Colombia. *Agronomía Mesoamericana* 226:129-143.
- Corona-Jácome 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* 9:80-85.
- 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, Ashworth VE, De la Cruz M, Clegg MT (2009). Tracing the geographic origins of major avocado cultivars. *Journal of Heredity* 100:56-65.
- Earl DA, vonHoldt BM (2012). Structure Harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources* 4:359-361.
- Evanno G, Regnaut S, Goudet, J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14(8):2611-2620.
- Frankel OH, Brown AHD, Burdon JJ (1995). The conservation of plant biodiversity. Cambridge Univ. Press, Cambridge.
- Freeland J (2005). *Molecular Ecology*. Ed. John Wiley and Sons. West Sussex. England.
- 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 Genetics and Genomes* 9:539-555.
- Guzmán LF, Machida-Hirano R, Borraro E, Cortés-Cruz M; Espíndola-Barquera MC, Heredia-García E (2017). Genetic structure and selection of a core collection for long term conservation of avocado in Mexico. *Frontiers on Plant Science* 8:243.
- International Plant Genetic Resources Institute (IPGRI) (1995) Descriptors for Avocado (*Persea* spp.). International Plant Genetic Resources Institute, Rome, Italy 106 p.
- Karp A, Edwards K, Bruford M, Vosman B, Morgante M, Seberg O, Kremer A, Boursot P, Arctander P, Tautz D, Hewitt G (1997). Newer molecular technologies for biodiversity evaluation: Opportunities and challenges. *Nature Biotechnology* 15:625-628.
- Kumar S, Dudley J, Nei M, Tamura K (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief in Bioinformatics* 9:299-306.
- Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL, van Eeuwijk FA (2011). Determination of genetic structure of genebank collections: are traditional hierarchical clustering methods appropriate for molecular marker data? *Theoretical Applied Genetics* 123:195-205.
- Peakall R, Smouse PE (2012). GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28:2537-2539.
- 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. *Journal of Heredity* 86:248-249.
- Rodríguez M, Jaramillo G, 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. Nov 11-13.
- Sánchez-Pérez J (1999). Recursos genéticos de aguacate (*Persea americana* Mill.) y especies afines en México. *Revista Chapingo Serie Horticultura* 5:7-18.
- Schnell RJ, Brown JS, Olano CT, Power EJ, Krol CA, Kuhn DN, Motamayor JC (2003). Evaluation of avocado genebank using microsatellite markers. *Journal of the American Society Horticultural Science* 128:881-889.
- Sharon D, Cregan PB, Mhameed S, Kusharska M, Hillel J, Lahav E, Lavi U (1997). An integrated genetic linkage map of avocado. *Theoretical Applied Genetics* 95:911-921.
- Sokal RR, Rohlf FJ (1995). *Biometry: the principles and practice of statistics in biological research*. WH Freeman, New York