JMBFS

Journal of Microbiology, Biotechnology and Food Sciences

International peer-reviewed scientific online journal

Published by Faculty of Biotechnology and Food Sciences

SSRs MARKERS REVEAL HIGH GENETIC DIVERSITY AND LIMITED DIFFERENTIATION AMONG POPULATIONS OF NATIVE GUATEMALAN AVOCADO

José Alejandro Ruiz-Chután[1,2], Julio Ernesto Berdúo-Sandoval[1], Marie Kalousová[2], Eloy Fernández[2], Jana Žiarovská[3], Amílcar Sánchéz-Pérez[1], Bohdan Lojka[2]*

Address(es):
[1] Department of Crop Sciences and Agroforestry, Czech University of Life Sciences Prague, Kamýcká 129 165 00 Praha – Suchdol, Czech Republic.
[2] Facultad de Agronomía, Universidad de San Carlos de Guatemala, Ciudad Universitaria Zona 12, Guatemala.
[3] Department of Genetics and Plant Breeding, Research Centre AgroBioTech, Slovak University of Agriculture in Nitra, Slovak Republic.

*Corresponding author: lojka@ftz.czu.cz

https://doi.org/10.55251/jmbfs.6134

ARTICLE INFO

Received 5. 6. 2022
Revised 10. 6. 2022
Accepted 14. 6. 2022
Published 1. 10. 2022

Regular article

OPEN ACCESS

ABSTRACT

Even though Guatemala is considered one of the three domestication centres of the avocado (*Persea americana* Mill), little is known about the genetic diversity of its native populations. This is the first study to explore avocado genetic diversity and population structure in native dispersal areas using microsatellite markers. We sampled 189 avocado trees grown from seeds in eight geographical populations, and 12 microsatellite loci were tested for diversity. There were a total of 289 alleles found throughout the 12 loci, with an average of 23.83 alleles per locus. The average observed and expected heterozygosity were 0.53 and 0.83, respectively. The analysis of molecular variance revealed that the eight populations comprised around 2% of the variation. Five of the 28 $G'_{ST(Nei)}$ pairwise comparisons indicated no genetic difference. We discovered a mixture of avocado trees from various populations using unweighted pair group method with arithmetic mean (UPGMA) hierarchical cluster analysis. The sampled individuals were classified into three main genetic clusters by the model-based STRUCTURE and discriminant analysis of principal components (DAPC). The weak genetic structure exhibited by mixing avocado trees from different populations could be mainly attributed to high human-mediated gene flow (*Nm* = 12.25) due to avocado's importance in food, culture, and Guatemala's religion. The high genetic diversity found in the avocado germplasm suggests that it might be a valuable source of variable alleles that could be used in breeding programs for this crop in Guatemala.

Keywords: Persea americana, population structure, genetic admixture, discriminant analysis of principal components, analysis of molecular variance

INTRODUCTION

Avocado (*Persea americana* Mill.) is a (sub)tropical evergreen tree native to Mesoamerica (**Schaffer *et al.*, 2012**). It is currently one of the most commercially significant fruit crops worldwide (**Liu *et al.*, 2020**). Botanists classify avocado into three horticultural races based on ecological preferences and fruit characteristics (**Schaffer *et al.*, 2012**). The Mexican race contain cold-tolerant, early-maturing fruit with thin skin, the Guatemalan race is slightly cold-tolerant due to its origin in a tropical highland climate, and its fruit has thick skin, whereas the West Indian race is cold-sensitive and adapted to humid tropical conditions, and its thin-skinned fruits have higher sugar and lower oil content than the Mexican and Guatemalan races (**Galindo-Tovar *et al.*, 2008; Schaffer *et al.*, 2012; Gross-German & Viruel, 2013**).

The Mesoamerican topographical conditions, climatic barriers, and the large size of the avocado seed contributed to the low mobility of the genetic material among regions, which caused the three races to remain well separated until the arrival of Spanish explorers, who promoted movement and closer contact of the distinct races (**Popenoe & Zentmyer, 1997**). Furthermore, avocado flowering presents a protogynous dichogamy opening model that favours cross-pollination. Besides, sterility barriers do not exist between or among the three races (**Alcaraz & Hormaza, 2011; Gross-German & Viruel, 2013**). These three reasons have resulted in a current situation in which races have been mixed in various regions of America, and numerous collections demonstrate clear racial introgression (**Reyes-Alemán *et al.*, 2013**). Most commercial varieties have been developed as crosses between these three races (**Bergh, 1992; Fiedler *et al.*, 1998).**

Due to the steady increase of interest in avocado production, many studies have been developed to explore the genetic diversity of global avocado resources with molecular markers, like RAPD (**Fiedler *et al.*, 1998**), RFLP (**Davis *et al.*, 1998**), AFLP (**Cañas-Gutiérrez *et al.*, 2015**), SSRs (**Ashworth *et al.*, 2004; Gross-German & Viruel, 2013; Boza *et al.*, 2018**), and SNP (**Chen *et al.*, 2009**). More recent studies have reported using Next Generation Sequencing (NGS) to evaluate the genetic diversity of avocado germplasm more precisely, resulting in substantial contributions to breeding programs (**Ge *et al.*, 2019; Rubinstein *et al.*, 2019; Talavera *et al.*, 2019**).

Nevertheless, studies have mainly focused on characterising cultivated commercial varieties and cultivars. On the other hand, the genetic diversity of avocado populations in their native dispersal areas, like Guatemala, remain poorly understood. In Guatemala, the molecular characterisation of native avocado is crucial because the populations have exhibited a dramatic size reduction, mainly by land use changes and deforestation. From 2000 to 2017, around 854,137 ha were deforested, and 1,012,947 ha of forest were disturbed in Guatemala (**Bullock *et al.*, 2020**), including areas of natural dispersion of native avocado populations. Another factor contributing to the decline of native genotypes is establishing improved cultivars or high productivity hybrids (**Rincón *et al.*, 2011**). Altogether, these factors cause genetic erosion damage to avocado's germplasm, which is highly valuable because it provides excellent traits of fruit quality and taste and resistance to soil-borne diseases to be used in future breeding programs (**Zentmyer *et al.*, 1988; Guzmán *et al.*, 2017**).

In Guatemala, native avocado populations can be found in the country's central, western, and northern parts (**Knight & Campbell, 1999**). They can be found in a wide altitudinal range (1,300 to 2,900 masl) and various vegetation types, including deciduous bushland, horticultural landscapes, and forest (**Wolstenholme & Whiley, 1999**). The genetic diversity of Guatemala's native avocado populations is unknown. As a result, the objective of this study was to use highly variable microsatellites markers (SSR) to examine the genetic diversity and genetic structure of native avocado populations in Guatemala's main natural dispersal areas. SSR markers can be used to investigate pedigrees, population structure, genomic variation, evolutionary processes, and fingerprinting (**Abdul-Muneer, 2014**). SSRs are still relevant and cost-effective markers to utilise, even when newer molecular approaches such as genotyping by sequencing or restriction site-associated DNA sequencing are available (**Hodel *et al.*, 2016**). We hypothesise that the diverse ecological conditions and geographical differences contributed to creating a distinct genetic structure in the studied populations. The findings of this study will lead to a better understanding of Guatemala's avocado genetic pool and will be valuable in future conservation and breeding programs.

MATERIAL AND METHODS

Study site, tree sampling, and data collection

The present study was carried out in several departments representing the main areas of natural dispersal of native avocado. The ecological characteristics of each department included in the study are shown in the S1 Table. In total, 189 individual trees were sampled from eight populations (Fig 1). We collected three fresh leaves from each individual tree, dried with silica gel, packed in plastic bags, labelled, and transported to Prague, Czech Republic, for genetic analysis at the Czech University of Life Sciences Prague's (CZU) Molecular Genetics Laboratory.

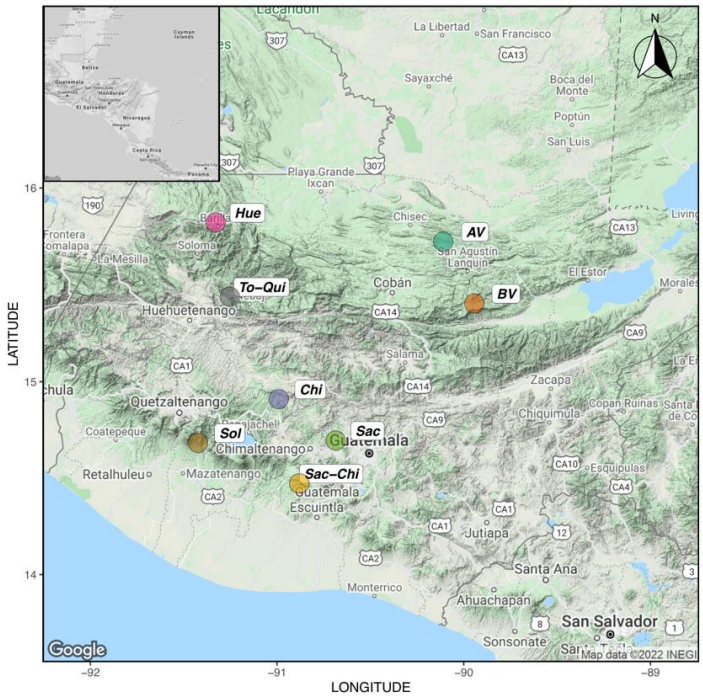

Figure 1 Map of Guatemala, displaying the geographical location of sampled avocado populations.

DNA isolation and SSR analyses

The cetrimonium bromide (CTAB) technique was used to extract DNA (**Doyle & Doyle, 1987**). A NANODROP™ (THERMOFISHER SCIENTIFIC, MA, USA) spectrophotometer was used to determine the concentration and purity of DNA. For the polymerase chain reaction (PCR), DNA samples were diluted to a final concentration of 25 ng µL-1. Twelve microsatellite primer pairs earlier designed for *P. americana* (**Sharon *et al*., 1997; Ashworth *et al*., 2004**) were utilised to amplify all DNA samples. We used four different colours to mark forward primers for detection fluorescently. Three multiplex PCRs were performed with varying annealing temperatures for each one and optimised concentrations for each primer (S2 Tab).

The PCR reaction mixtures were prepared in a total volume of 10 µL, containing 1 µL of DNA (25 ng L-1), primers at the concentrations listed in the S2 Table, and Multiplex PCR Plus (1 X) (QIAGEN®, DUS, DE). The Thermal Cycler T 100 (BIO-RAD, CA, USA) was used to perform PCR amplification with the following profile: 95 °C for 15 min, followed by 35 cycles at 95 °C for 30 s, either 63.4 °C (M1), 57.6 °C (M2), or 65 °C (M3) for 1 min, 72 °C for 1 min, and a final extension at 72 °C for 10 min. The PCR products were separated using electrophoresis on a Genetic Analyzer 3500 (APPLIED BIOSYSTEMS, CA, USA). A 1 µL aliquot of PCR products was combined with 0.2 µL of GeneScan-500 LIZ (APPLIED BIOSYSTEMS, CA, USA) and 12 µL of Hi-Di formamide (APPLIED BIOSYSTEMS, CA, USA). GeneMarker v.2.4.0 was used to score the microsatellite alleles (SOFTGENETICS, PA, USA).

Data analysis

Genetic diversity and population genetic divergence

We used GenoDive v.3.05 (**Meirmans, 2020**) to calculate basic statistics such as the number of alleles (*Na*), number of effective alleles (*Ne*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), corrected fixation index ($G'_{ST(Nei)}$), and inbreeding coefficient (G_{IS}) to assess genetic diversity per locus and population. FSTAT software was used to calculate allelic richness (*Ar*) using a rarefaction method (**Goudet, 1995**). The formula for calculating gene flow (*Nm*) was $Nm = (1\ G'_{ST(Nei)}/4\ G'_{ST(Nei)}$ (Slatkin & Barton, 1989). Shannon's information index (*I*) was carried out in GenAlEX v.6.5 (**Peakall & Smouse, 2012**), Hardy-Weinberg equilibrium tests and Analysis of Molecular Variance (AMOVA) were performed in GenoDive.

Individual samples were grouped according to populations, regions, and fixation indices (F_{IT}, F_{IS}, F_{ST}, F_{SC}, and F_{CT}) were computed using hierarchical global AMOVA. We used GenoDive to do a Mantel test to see if isolation by distance was significant. To examine pairwise population differentiation ($G'_{ST(Nei)}$), we also used GenoDive, and its visualisation was conducted using the package superheat (**Barter & Yu, 2018**).

Population structure

Using the poppr package (**Kamvar *et al*., 2014**) implemented in R software v.4.0.3 (**R Core Team, 2020**), the allele dataset in GenAlEx format was converted into a genind object. Avocado samples with similar genetic features across the 12 loci were assembled using a hierarchical cluster analysis. We used the software poppr to compute Bruvo's genetic distance and create the UPGMA hierarchical cluster. The dendrogram was viewed and customised in the Interactive tree of life (iTOL) v.4 program (**Letunic & Bork, 2019**).

Afterwards, to infer the genetic clusters (subpopulations) and explore the population structure of the sampled trees, we employed a model-based Bayesian clustering method implemented by software STRUCTURE v.2.3.4 (**Pritchard *et al*., 2000**). The program was run with 100,000 burn-in steps, followed by 1,000,000 MCMC iterations for the number of clusters K=1 to 10, with ten repetitions for each K. The model assumed correlated allele frequency for the populations and mixed ancestry (admixture) for the individuals. Because weak structuring was assumed, the LOCPRIOR model was implemented, which uses sampling locations as prior information to assist the clustering (**Hubisz *et al*., 2009**). Due to the uneven sampling across geographical populations, the parameters alpha were set to 1/K, following the recommendations of **Wang (2017)**. For collating results generated by the program STRUCTURE, the output files were analysed by the web-based program StructureSelector (**Li & Liu, 2018**), which implements Puechmaille estimators (**Puechmaille, 2016**) to detect the optimal number of K that best fits the data. The resulting replicate Q-matrices for the optimal K were aligned in CLUMPP (**Jakobsson & Rosenberg, 2007**), and a bar graph was visualised using ggplot2 (**Wickham, 2016**).

We used a Discriminant Analysis of Principal Components (DAPC) implemented in the program adegenet (**Jombart, 2008**) to confirm the population assignment obtained by STRUCTURE. The method required utilising the find.clusters function from the adegenet package to determine the ideal number of genetic clusters (K), and then using the Bayesian Information Criterion (BIC) to choose the optimal number of genetic clusters using the elbow method. DAPC was used to describe the clusters that were discovered. The correct number of principal components and discriminant functions to be preserved were confirmed using a cross-validation function (Xval.dapc).

RESULTS

Population genetic diversity

Across the 189 avocado trees, 286 alleles were found at twelve loci. The average mean number of alleles per locus was 23.83 (Tab 1). The highest alleles per locus (*Na*) was 32 (AUCR418), whereas the locus with the fewest alleles per locus (*Na*) was AVT436 (9). The allelic richness (*Ar*) varied from 3.93 (AVAG07) to 8.66 (AVD001). Expected heterozygosity (*He*) and observed heterozygosity (*Ho*) ranged from 0.60 (AVAG07) to 0.92 (AVD001) and from 0.29 (AVAG07) to 0.68 (AVMIX04), respectively.

Overall, the mean of *Ho* was lower than that of *He*, showing that avocado populations have a homozygote excess. Inbreeding coefficient (G_{IS}) varied from 0.15 (AVAG22) to 0.51 (AVAG07). The Shannon's information index (*I*) ranged from 1.18 to 2.42 (AVAG07) (AVAG22). All the loci were found to be significantly out of Hardy-Weinberg equilibrium.

Table 2 shows the genetic diversity of each of the eight avocado populations. To-Qui had the greatest *Na* (14.16), while BV had the lowest (*Na* = 7.16). *Ar* varied from 5.47 (Chi) to 6.80 (Hue). With averages of 0.53 and 0.83, respectively, the *Ho* varied from 0.41 to 0.65 and the *He* from 0.78 to 0.89. When compared to *He*, all of these populations exhibited lower *Ho*. Population BV had the lowest Ho and population Sac had the highest, whereas population Sol had the lowest *He* and population Sac had the most. The G_{IS} (genetic inbreeding coefficient) ranged from 0.16 (Sac) to 0.49 (Sac) (BV). All populations exhibited a significant deviation from Hardy-Weinberg equilibrium.

Table 1 Characterization of 12 simple sequence repeat loci in avocado (*P. americana*) based on 189 trees representing eight populations in Guatemala.

Locus name	Repeats	*Na*	*Ne*	*Ar*	*Ho*	*He*	$G'_{ST(Nei)}$	G_{IS}	I	*Nm*	HW
AVAG05 [a]	$(AG)_{10}$	25	4.59	6.18	0.49	0.81	0.037	0.39	1.82	6.51	*
AVAG11 [a]	$(CT)_{18}$	24	3.71	5.54	0.38	0.76	0.049	0.49	1.60	4.85	*
AVAG13 [a]	$(AG)_{20}$	31	6.66	7.66	0.61	0.88	0.019	0.30	2.18	12.91	*
AVAG07 [a]	$(TC)_{15}$	16	2.38	3.93	0.29	0.60	0.031	0.51	1.18	7.81	*
AVAG21 [a]	$(CT)_{22}$	20	4.66	5.85	0.64	0.81	0.009	0.21	1.74	27.53	*
AVAG25 [a]	$(TC)_{14}$	19	5.33	6.17	0.46	0.84	0.017	0.45	1.86	14.46	*
AVD022 [a]	$(TC)_{13}$	24	5.23	6.17	0.46	0.84	0.022	0.45	1.88	11.11	*
AVMIX04 [a]	$(AG)_{12}$ $(CAA)_5$ $(ACAG)_{10}$	26	8.03	7.84	0.68	0.90	0.018	0.24	2.27	13.64	*
AVT436 [b]	$(ATC)_9$	9	3.35	4.53	0.41	0.73	0.017	0.43	1.39	14.46	*
AUCR418 [b]	$(CT)_{22}$	32	6.01	7.03	0.53	0.87	0.01	0.38	2.09	24.75	*
AVD001 [b]	$(CT)_{12}$	31	9.03	8.66	0.62	0.92	0.002	0.32	2.38	124.75	*
AVAG22 [b]	$(GA)_{15}$	29	9.00	8.27	0.77	0.91	0.016	0.15	2.42	15.38	*
Mean		23.83	5.67	6.49	0.53	0.82	0.02	0.35	1.90	12.25	

* indicates significance of p value at ≤ 0.01 a= from (Sharon et al., 1997) b= from (Ashworth et al., 2004), *Na*: observed number of alleles, *Ne*: effective number of alleles, *Ar*: allelic richness, *Ho*: observed heterozygosity, *He*: expected heterozygosity; $G'_{ST(Nei)}$: corrected fixation index, G_{IS}: inbreeding coefficient, *I*: Shannon information index. *Nm*: gene flow, HW: Hardy-Weinberg equilibrium test.

Table 2 Genetic diversity of 189 avocado (*P. americana*) trees representing eight populations in Guatemala as detected by allele sizes at 12 simple sequence repeat loci.

Population	Region	*N*	*Na*	*Ne*	*Ar*	*Ho*	*He*	G_{IS}	*I*	HW
Sac	Central	36	10.66	5.24	5.56	0.65	0.78	0.16	1.82	*
Sac-Chi		32	10.58	5.41	5.75	0.53	0.80	0.33	1.87	*
Chi		23	8.25	5.12	5.47	0.64	0.79	0.18	1.74	*
Sol		8	7.58	5.66	6.70	0.47	0.89	0.47	1.82	*
To-Qui	Western	36	14.16	7.23	6.61	0.60	0.84	0.28	2.15	*
Hue		23	12	7.42	6.80	0.49	0.85	0.41	2.11	*
BV	Northern	11	7.16	5.07	5.87	0.41	0.82	0.49	1.69	*
AV		20	10.75	7.05	6.60	0.44	0.84	0.47	2.01	*
Mean			10.15	6.03	6.17	0.53	0.83	0.35	1.90	

* indicates significance of *p* value at ≤ 0.01, *N*: number of sampled trees, *Na*: observed number of alleles, *Ne*: effective number of alleles, *Ar*: allelic richness, *Ho*: observed heterozygosity, *He*: expected heterozygosity, G_{IS}: inbreeding coefficient, *I*: Shannon information index, HW: Hardy-Weinberg equilibrium test.

Population genetic divergence

$G'_{ST(Nei)}$, a genetic differentiation metric, ranged from 0.049 (AVAG11) to 0.002 (AVD001) among the 12 loci, with a mean of 0.02 (Tab 1). The average gene flow *Nm* in 12 SSR loci was 12.25 (Tab 1), far greater than 1, indicating that gene flow among eight avocado populations was relatively frequent. It was consistent with the low level of genetic differentiation among populations.

The AMOVA indicated that the genetic diversity mainly was within individuals (67.4%), followed by within populations (30.7%) and only 1.8% occurring among populations (Tab 3). When the populations were further grouped into regions, the percentage of variation indicated that the genetic diversity mainly was within individuals (67.1%) and within populations (30.6%), followed by among regions (1.8%) and finally populations within regions (0.6%) (Tab 3).

Table 3 Analysis of molecular variance using 10,000 permutations for 189 avocado (*P. americana*) trees from eight populations

Source of Variation	Sum of squares	Variance component	Percentage of variation	F-stat	p-value
A: when the samples trees were grouped according to geographical populations					
Within individuals	615.5	3.381	67.4	F_{IT} = 0.326	< 0.001
Among individuals within populations	1121.5	1.54	30.7	F_{IS} = 0.313	< 0.001
Among populations	74.0	0.092	1.8	F_{ST} = 0.018	< 0.001
B: when the geographical populations were further grouped according to regions					
Within Individuals	615.5	3.381	67.1	F_{IT} = 0.329	< 0.001
Among individuals within populations	1121.5	1.54	30.6	F_{IS} = 0.313	< 0.001
Among populations within regions	38.8	0.03	0.6	F_{SC} = 0.006	0.007
Among regions	35.2	0.089	1.8	F_{CT} = 0.018	< 0.001

The genetic divergence between the eight avocado populations was established by calculating pairwise $G'_{ST(Nei)}$ comparisons (Fig 2A). Only 5 of the 28 pairs of populations did not show a significant differentiation ($p > 0.05$). The $G'_{ST(Nei)}$ values ranged from 0 (AV and BV) to 0.06 (Sac and BV). The Mantel test revealed a significant association between genetic and geographical distance ($r = 0.420$, $p < 0.05$) (Fig 2B).

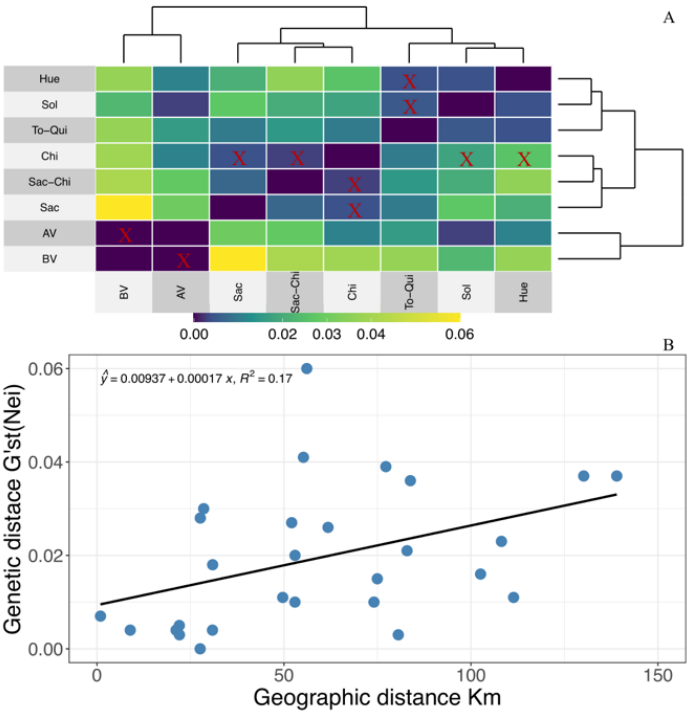

Figure 2 A) Pairwise $G'_{ST(Nei)}$ heatmap and dendrogram based on $G'_{ST(Nei)}$ values among the eight sampling sites. The $G'_{ST(Nei)}$ matrix is represented by the heatmap colour code, which includes discrete $G'_{ST(Nei)}$ bins ranging from low to high genetic differentiation. At $p < 0.05$, the X indicates that the p value is not significant. B) Mantel test demonstrating a link between geographic and genetic distances between eight sampling sites. Based on 10,000 permutations of 28 pairwise comparisons, the genetic and geographical distances were associated (Rxy = 0.420, p = 0.038).

Population genetic structure

The UPGMA-based revealed three distinct groupings, each including samples from different populations, pointing at a weak genetic structure (Fig 3).

Figure 3 Dendrogram constructed with UPGMA showing the genetic relationship of the 189 avocado trees from the eight analysed populations in Guatemala based on a dataset of twelve microsatellite loci. The same font colour is used to signify samples obtained from the same population. The inner circle represents the three main clusters identified on the dendrogram depicted by each colour revealing high admixture between samples from different populations

Based on Puechmaille estimators obtained from StructureSelector, the true number of clusters among all samples was K = 3 (S1 Fig). It suggests that the 189 avocado tree samples can be divided into three groups or clusters (Fig 4). When K = 3, the genetic structure implies a significant similarity between the Sac, Sac-Chi, and Chi populations, depicted by blue colour. Individuals from the Sol population had a genetic makeup similar to that of some individuals from other populations, such as BV and AV. To-Qui and Hue's populations exhibited an ancestry mainly represented by red colour and blue colour in less proportion. BV and AV populations showed an ancestry mainly represented by orange colour and in less proportion by blue colour.

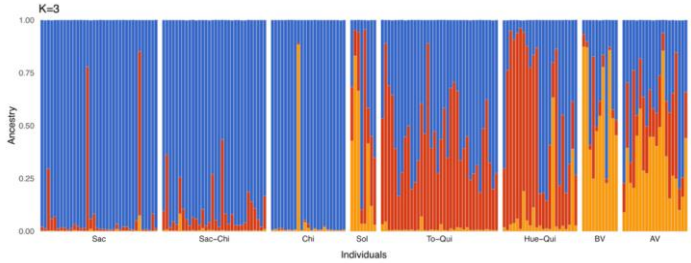

Figure 4 Bayesian clustering analysis based on twelve SSR markers data for avocado populations from Guatemala performed using STRUCTURE where three (K =3) main clusters were identified. Each individual is shown by a vertical bar divided into colored segments that indicate the estimated membership percentages of that individual.

The data were described using 80 principal components amounting to 92.5% of the total variance and two discriminant functions. The find.cluster function found three clusters that were linked to the BIC value (S2 Fig). The DAPC plot (Fig 5A) revealed three clusters, with clusters 1 and 2 (to the left) and cluster 3 (to the right) separated by the first discriminant function. Cluster 1 and 2 were distinguished only by the second discriminant function. Cluster 1 had the most individuals (76), followed by cluster 3 (56) and cluster 2 (52). The suggested cluster by DAPC showed a weak structure between the clusters (Fig 5 B) consistent with STRUCTURE and UPGMA dendrogram.

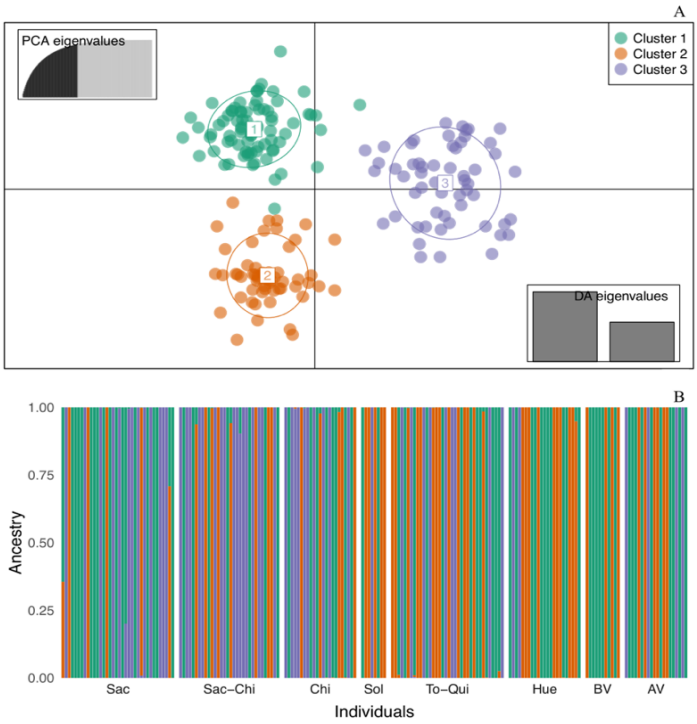

Figure 5 Analysis of population structure using DAPC based on 189 avocado genotypes from eight populations. (A) Scatterplot of the first two discriminant functions, based on 80 retained principal components. Each colour represents a genetic cluster (B) Barplot showing assignment of individuals to the three clusters recovered by the DAPC without prior information.

DISCUSSION

In the present study, a total of 286 alleles were identified, showing high genetic diversity among the 189 sampled avocado trees of the eight populations using 12 SSR loci with the number of alleles ranging from 9 to 32 per locus (Tab 1). For comparison, **Juma *et al.* (2020)** detected 167 alleles employing 10 SSR loci across 226 Ghanian genotypes, with the number of alleles ranging from 10 to 23 per locus, whereas **Schnell *et al.* (2003)** detected 8 to 30 alleles per locus. The mean number

of alleles per locus obtained in the present study was 23.83. A slightly lower number of alleles per locus was reported by **Juma *et al*. (2020)**, **Schnell *et al.* (2003)** and **Guzmán *et al*. (2017)**, 16.7, 18.8 and 19.5, respectively. **Gross-German & Viruel (2013)** and **Abraham & Takrama (2014)** reported lower numbers, 11.4 and 11.5, respectively. The differences between the results from this study and the previously cited results could be attributed to many factors such as sample size, different markers used, the diversity of the germplasm investigated, or the accuracy of the platform for genotyping (**Vieira *et al*., 2016**). Other factors that could explain the discrepancies include the quality of genomic DNA used for SSR PCR amplification and PCR procedure optimisation (**Juma *et al*., 2020**).

The mean observed heterozygosity across SSR loci obtained in the present study was 0.53, which was similar to the 0.56 reported by **Boza *et al.* (2018)**, indicating that both gene pools have similar levels of genetic diversity. Higher *Ho* values have been reported by **Juma *et al.* (2020)**: 0.65, **Schnell *et al.* (2003)**: 0.64, and **Guzmán *et al.* (2017)**: 0.61. Lower *Ho* values in avocado genetic reservoirs have been reported by **Abraham & Takrama (2014)**: 0.48 and **Liu *et al*. (2020)**: 0.39, exhibiting lesser genetic diversity in the germplasm studied than ours. Lower *Ho* values could also be due to the use of different SSR loci among the studies.

The diversity analysis among populations exposed that the observed and expected heterozygosity, Shannon's information index, and allelic richness were the highest for Sac, Sol, To-Qui, and Hue, respectively (Tab 2). It means that populations from the central and western regions are more diverse than those from the north. It hence may provide elite genotypes for breeding programs incorporating the genetic diversity into new commercialised cultivars. In addition, western and central populations may also adapt to climate fluctuation better than northern populations. Most of the diversity measures were the lowest in the BV population, indicating lower genetic diversity. This finding could be explained by the widespread replacement of native avocado genotypes with commercial cultivars, owing to rising demand for the Hass variety, which reduces genetic diversity within the genetic pool.

The inbreeding level identified within the analysed populations may lead to Hardy-Weinberg disequilibrium exhibited across the 12 SSR loci (Tab 1). Wright's fixation indices (F_{IT}, F_{ST}, and F_{IS}) are useful for estimating genetic variation within and among populations. F_{IT} calculates the heterozygosity deficit ($F_{IT} > 0$) or excess ($F_{IT} < 0$) across all populations (**Lachance, 2009**). When AMOVA was computed without considering regions, the global heterozygosity deficit (F_{IT}) was 0.326 ($p < 0.001$; Tab 3). It denotes around 33% more observed homozygotes which is concordant with the G_{IS} values (Tab 1), indicating a significant reduction of heterozygotes.

In the same order, F_{IS} is a measure of deviation from panmixia at local scales (e.g., within subsamples). It indicates the deficit ($F_{IS} > 0$) or excess ($F_{IS} < 0$) of heterozygotes (**de Meeûs et al., 2007**). The average F_{IS} was 0.313 ($p < 0.001$), which indicates that within subpopulations, there is about a 31% deficit of heterozygotes. The high frequency of homozygotes can be attributed to a variety of factors. The presence of null alleles, which are induced by PCR failure in microsatellites amplification (**Wattier *et al*., 1998; Stadhouders *et al*., 2010**), causes heterozygote individuals for these alleles to be wrongly categorised as homozygotes for the dominant alleles (**Lemer *et al*., 2011**). Another factor that can reduce the observed frequency of heterozygotes in a population is the Wahlund effect which can be defined as the apparent excess of homozygotes and the deficit of heterozygotes observed in a sample of individuals due to the existence of population subdivision (**de Meeûs, 2018**).

F_{ST} measures the mean deficiency of expected heterozygotes among subpopulations with respect to that expected for the total population. Usually, it measures differentiation between subpopulations. The genetic differentiation among the eight avocado geographical populations was low ($F_{ST} = 0.018$, $p < 0.001$), however, a significant population-based subdivision of Guatemalan avocados was observed. Thus, genetic variation mostly existed within populations, as shown through AMOVA, where close to 98% of the genetic variation was shared within and among the populations. It shows that gene flow was quite common among avocado populations, minimising genetic differentiation (**Wright, 1931**), consistent with the high gene flow value (*Nm* = 12.25). The F_{ST} value detected in this study is close to 0.061 and 0.054 published by **Juma *et al.* (2020)** and **Cañas-Gutiérrez *et al.* (2019)**, but lower than the 0.19, 0.25, and 0.63 previously reported by **Boza *et al.* (2018)**, **Gross-German & Viruel (2013)** and **Talavera *et al.*, (2019)**, respectively. The obvious difference is because, as **Juma *et al.* (2020)** and **Cañas-Gutiérrez (2015)**, the avocado trees genotyped in our study corresponded only to native avocados, excluding commercial varieties. **Boza *et al.* (2018)** presented a different situation, who genotyped trees of *P. americana*, *P. nubigena*, and *P. krugii;* and **Gross-German & Viruel (2013)** genotyped 42 avocado accessions including rootstocks, commercial varieties, and Spanish local selections.

AMOVA showed that the genetic variation among the eight populations was 1.8% of the total variation (Tab 3). When the samples were grouped according to the three main geographical regions, the genetic variation among groups was also 1.8% with a significant F_{CT} value ($F_{CT} = 0.018$, $p < 0.001$), pointing to substructure but in very a slight way. These results are consistent with **Juma *et al.* (2020)**, who reported that the genetic variation among groups was 1.98% ($F_{CT} = 0.019$, $p < 0.05$) when the geographical populations were grouped according to regions. The substructure found with the AMOVA is consistent with the UPGMA dendrogram at a population level (Fig 2A) which is grouped according to geographical region, except for the Sol population, which a priori was included in the central region but was more related to the western population. However, it is not surprising because the geographical distance between Sol and western populations is relatively short. The substructure and organisation displayed in the dendrogram (Fig 2A) are also supported by the Mantel test with a significant correlation between geographical and genetic distances among the eight populations.

Nonetheless, when we explored the genetic relationships among all the sampled trees in-depth, with UPGMA cluster analysis at individuals' level (Fig 3), it was not possible to separate the trees according to their populations or regions. The pairwise population $G'_{ST(Nei)}$ (Fig 2A) confirmed the dendrogram findings, revealing an absence of differentiation between pairs of populations, such as AV versus BV, Sac versus Chi, and Sol versus To-Qui. As one of the three domestication centres, the religious, mythological, economic, and medicinal significance of the avocado in Guatemala could explain the genetic admixture observed among avocado populations (**Galindo-Tovar *et al*., 2008**). This situation persists nowadays and has become the main factor affecting the genetic diversity and structure of the populations.

To confirm the population structure of the 189 avocado genotypes, we used the model-based STRUCTURE. The results showed that the sampled trees could be regrouped into three clusters based on the information of the 12 studied loci. When K = 3, high similarity in population structure was observed between populations from the central region, except by Sol population, which exhibited genetic similarity with northern populations. In addition, populations from western and northern regions showed a differentiated genetic ancestry. This result is sustained by the absence of differentiation among populations from the same region found with the population pairwise $G'_{ST(Nei)}$ (Fig 2A). These findings were corroborated by DAPC, which proved that the 189 sampled trees might be organised into three genetic groups. The DAPC was accurate to assign the ancestry of each genotype according to the three groups (clusters) inferred.

Nevertheless, the bar plot of DAPC (Fig 5B) exhibited a weak structure among populations and pointed out that each geographical population has alleles originating from at least three clusters. Diverse analyses such as AMOVA, UPGMA, STRUCTURE, and DAPC revealed that avocados grown in different Guatemalan populations show high genetic similarity with low genetic differentiation among them. Here, we demonstrate that the diverse ecological conditions and geographical differences do not contribute to creating a distinct genetic structure in the studied populations, probably due to high human-mediated gene flow (*Nm* = 12.25) as suggested by **Galindo-Tovar *et al*. (2008)**. It could be considered as the main factor affecting the diversity and genetic structure of Guatemalan avocado populations.

CONCLUSION

This study represents the first characterisation of native Guatemalan avocado trees based on microsatellites markers. We found a high genetic diversity in the analysed germplasm, which was described based on various diversity indices. It also demonstrates the potential of the native germplasm as a precious source of diverse alleles that must be conserved for its exploitation in breeding programs. Therefore we recommend establishing a core collection of the native genotypes from the whole range as populations genetic differentiation is not very high. The hierarchical cluster analysis revealed an admixture of trees from different regions and populations exhibiting a weak but significant population structure, denoting a high gene flow among the populations. It is corroborated by STRUCTURE and DAPC, exposing that the avocado genotypes can be regrouped into three genetic populations. Today's Guatemalan avocado germplasm richness is a product of gene flow between populations. However, this flow is threatened by deforestation and the extensive introduction of commercial varieties that displace native genotypes, reducing their population size. Thus, we suggest avoiding the decreasing population size and keeping the connectivity among them to ensure the conservation of genetic diversity.

Acknowledgments: This research had financial support from the budget assigned to Dirección General de Investigación (Digi) through the 4.8.63.4.41project and The Internal Grant Agency of CULS Prague (No. 20213110)

REFERENCES

Abdul-Muneer, P. M. (2014). Application of microsatellite markers in conservation genetics and fisheries management: Recent advances in population structure analysis and conservation strategies. *Genetics Research International*, 2014, 691759. https://doi.org/10.1155/2014/691759

Abraham, J. D., & Takrama, J. F. (2014). Genetic characterisation of avocado (*Persea americana* Mill.) in two regions of Ghana. *African Journal of Biotechnology*, *13*(51), 4620–4627. https://doi.org/10.5897/ajb2014.14023

Alcaraz, M. L., & Hormaza, J. I. (2011). Influence of physical distance between cultivars on yield, outcrossing rate and selective fruit drop in avocado (*Persea americana*, Lauraceae). *Annals of Applied Biology*, *158*(3), 354–361. https://doi.org/10.1111/j.1744-7348.2011.00469.x

Ashworth, V. E. T. M., Kobayashi, M. C., Cruz, M. D. L., & Clegg, M. T. (2004). Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. *Scientia Horticulturae 101*(204), 255–267. https://doi.org/10.1016/j.scienta.2003.11.008

Barter, R. L., & Yu, B. (2018). Superheat: An R package for creating beautiful and extendable heatmaps for visualizing complex data. *Journal of Computational and Graphical Statistics*, *27*(4), 910–922. https://doi.org/10.1080/10618600.2018.1473780

Bergh, B. (1992). The origin, nature, and genetic improvement of the avocado. *California Avocado Society*, *76*, 61–75.

Boza, E. J., Tondo, C. L., Ledesma, N., Campbell, R. J., Bost, J., Schnell, R. J., … & Gutiérrez, O. A. (2018). Genetic differentiation, races and interracial admixture in avocado (*Persea americana* Mill.), and *Persea spp*. evaluated using SSR markers. *Genetic Resources and Crop Evolution*, *65*(4), 1195–1215. https://doi.org/10.1007/s10722-018-0608-7

Bullock, E. L., Nolte, C., Segovia, A. R., & Woodcock, C. E. (2020). Ongoing forest disturbance in Guatemala's protected areas. *Remote Sensing in Ecology and Conservation*, *6*(2), 141–152. https://doi.org/10.1002/rse2.130

Cañas-Gutiérrez, G. P., Alcaraz, L., Hormaza, J. I., Arango-Isaza, R. E., Saldamando-Benjumea, C. I. (2019). Diversity of avocado (*Persea americana* mill.) cultivars from antioquia (Northeast Colombia) and comparison with a worldwide germplasm collection. *Turkish Journal of Agriculture and Forestry*, *43*(4), 437–449. https://doi.org/10.3906/tar-1807-25

Cañas-Gutiérrez, G. P., Galindo-López, L. F., Arango-Isaza, R., & Saldamando-Benjumea, C. I. (2015). Genetic diversity of avocado (*Persea americana* Mill) cultivars in Antioquia, Colombia. *Agronomía Mesoamericana, 26*(1), 129–143. https://doi.org/10.15517/am.v26i1.16936

Chen, H., Morrel, P., Ashworth, V., De la Cruz, M., & Clegg, M. (2009). Tracing the geographic origins of major avocado cultivars. *The American Genetic Association*, *100*(1), 56–65. https://doi.org/10.1093/jhered/esn068

Davis, J., Henderson, D., Kobayashi, M., Clegg, M. T., & Clegg, M. T. (1998). Genealogical relationships among cultivated avocado as revealed through RFLP analyses. *The Journal of Heredity*, *89*(4), 319–323. https://doi.org/10.1093/jhered/89.4.319

De Meeûs, T. (2018). Revisiting FIS, FST, Wahlund effects, and null alleles. *Journal of Heredity*, *109*(4), 446–456. https://doi.org/10.1093/jhered/esx106

De Meeûs, T., McCoy, K. D., Prugnolle, F., Chevillon, C., Durand, P., Hurtrez-Boussès, S., … & Renaud, F. (2007). Population genetics and molecular epidemiology or how to "débusquer la bête." *Infection, Genetics and Evolution*, *7*(2), 308–332. https://doi.org/10.1016/j.meegid.2006.07.003

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh tissue. *Phytochemical Bulletin*, *19*, 11–15.

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

Galindo-Tovar, M. E., Ogata-Aguilar, N., & Arzate-Fernández, A. (2008). Some aspects of avocado (*Persea americana* Mill.) diversity and domestication in Mesoamerica. *Genetic Resources and Crop Evolution*, *55*(3), 441–450. https://doi.org/10.1007/s10722-007-9250-5

Ge, Y., Zhang, T., Wu, B., Tan, L., Ma, F., Zou, M., … & Wang, T. (2019). Genome-Wide assessment of avocado germplasm determined from specific length amplified fragment sequencing and transcriptomes: Population structure, genetic diversity, identification, and application of race-specific markers. *Genes*, *10*(3), 215. https://doi.org/10.3390/genes10030215

Goudet, J. (1995). Computer program to calculate F-statistics. *Journal of Heredity*, *86*(6), 485–486. https://doi.org/10.1093/oxfordjournals.jhered.a111627

Gross-German, E., & Viruel, M. A. (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*(2), 539–555. https://doi.org/10.1007/s11295-012-0577-5

Guzmán, L. F., Machida-Hirano, R., Borrayo, E., Cortés-Cruz, M., Espíndola-Barquera, M. del C., & Heredia García, E. (2017). Genetic structure and selection of a core collection for long term conservation of avocado in Mexico. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00243

Hodel, R. G. J., Segovia-Salcedo, M. C., Landis, J. B., Crowl, A. A., Sun, M., Liu, X., … & Soltis, P. S. (2016). The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences*, *4*(6), 1600025. https://doi.org/10.3732/apps.1600025

Hubisz, M. J., Falush, D., Stephen, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology*, *9*(5), 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Juma, I., Geleta, M., Nyomora, A., Saripella, G. V., Hovmalm, H. P., Carlsson, A. S., … & Ortiz, R. (2020). Genetic diversity of avocado from the southern highlands of Tanzania as revealed by microsatellite markers. *Hereditas*, *157*(1). https://doi.org/10.1186/s41065-020-00150-0

Kamvar, Z. N., Tabima, J. F., & Gŕunwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Kimura, M., & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, *49*(4), 725–738.

Knight, R. J., & Campbell, C. W. (1999). Ecological adaptation and the evolution of modern avocado cultivars. *Revista Chapingo Serie Horticultura* 5, 49–54.

Lachance, J. (2009). Detecting selection-induced departures from Hardy-Weinberg proportions. *Genetics Selection Evolution*, *41*(1). https://doi.org/10.1186/1297-9686-41-15

Lemer, S., Rochel, E., & Planes, S. (2011). Correction method for null alleles in species with variable microsatellite flanking regions, a case study of the black-lipped pearl oyster *Pinctada margaritifera*. *Journal of Heredity*, *102*(2), 243–246. https://doi.org/10.1093/jhered/esq123

Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(1), 256–259. https://doi.org/10.1093/nar/gkz239

Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources*, *18*(1), 176–177. https://doi.org/10.1111/1755-0998.12719

Liu, Y., Ge, Y., Zhan, R., Lin, X., Zang, X., Li, Y., … & Ma, W. (2020). Molecular markers and a quality trait evaluation for assessing the genetic diversity of avocado landraces from China. *Agriculture*, *10*(4), 102. https://doi.org/10.3390/agriculture10040102

Meirmans, P. G. (2020). Genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, *20*(4), 1126–1131. https://doi.org/10.1111/1755-0998.13145

Peakall, R., & Smouse, P. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research an update. *Bioinformatics*, *28*, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Popenoe, W., & Zentmyer, G. A. (1997). Early history of the avocado. *California Avocado Association Yearbook*, 81, 163–171.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources*, *16*(3), 608–627. https://doi.org/10.1111/1755-0998.12512

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reyes-Alemán, J. C., Valadez-Moctezuma, E., Simuta-Velázco, L., Barrientos-Priego, A., & Gallegos-Vázquez, C. (2013). Distinction of species of the genus *Persea* by RAPD and ISSR DNA markers. *Revista Mexicana de Ciencias Agrícolas, 4*(4), 517–529 (in Spanish with an abstract in English).

Rincón, C., Sánchez, J., & Espinosa, F. (2011). Foliar chemical caracterization of the creole avocado trees (*Persea americana* var. *drymifolia*) in the germplasm banks from Michoacán, México. *Revista Mexicana de Biodiversidad*, *82*, 395–412 (in Spanish with an abstract in English).

Rubinstein, M., Eshed, R., Rozen, A., Zviran, T., Kuhn, D. N., Irihimovitch, V., … & Ophir, R. (2019). Genetic diversity of avocado (*Persea americana* Mill.) germplasm using pooled sequencing. *BMC Genomics*, *20*(1). https://doi.org/10.1186/s12864-019-5672-7

Schaffer, B., Wolstenholme, B., & Whiley, A. (2012). The Avocado: Botany, production and uses (2nd ed.). Croydon: CPI Group (UK) Ltd.

Schnell, R. J., Brown, J. S., Olano, C. T., Power, E. J., Krol, C. A., Kuhn, D. N., … & Motamayor, J. C. (2003). Evaluation of avocado germplasm using microsatellite markers. *Journal of the American Society for Horticultural Science, 128*(6), 881–889. https://doi.org/10.21273/JASHS.128.6.0881

Sharon, D., Cregan, P. B., Mhameed, S., Kusharska, M., Hillel, J., Lahav, E., … & Lavi, U. (1997). An integrated genetic linkage map of avocado. *Theoretical and Applied Genetics*, *95*(5-6), 911–921. https://doi.org/10.1007/s001220050642

Slatkin, M., & Barton, N. H. (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution, 43*(7), 1349. https://doi.org/10.2307/2409452

Stadhouders, R., Pas, S. D., Anber, J., Voermans, J., Mes, T. H. M., & Schutten, M. (2010). The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay. *The Journal of Molecular Diagnostics*, *12*(1), 109–117. https://doi.org/10.2353/jmoldx.2010.090035

Talavera, A., Soorni, A., Bombarely, A., Matas, A. J., & Hormaza, J. I. (2019). Genome-Wide SNP discovery and genomic characterization in avocado (*Persea americana* Mill.). *Scientific Reports*, *9*, 20137. https://doi.org/10.1038/s41598-019-56526-4

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, *39*(3), 312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Wang, J. (2017). The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Molecular Ecology Resources, 17*(5), 981–990. https://doi.org/10.1111/1755-0998.12650
Wattier, R., Engel, C. R., Saumitou-Laprade, P., & Valero, M. (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci: Experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Molecular Ecology, 7*(11), 1569–1573. https://doi.org/10.1046/j.1365-294x.1998.00477.x
Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org
Wolstenholme, B. N., & Whiley, A. W. (1999). Ecophysiology of the avocado (*Persea americana* Mill.) tree as a basis for pre-harvest management. *Revista Chapingo Serie Horticultura, 5*, 77–78.
Wright, S. (1931). Evolution in mendelian populations. *Bulletin of Mathematical Biology, 53*(1/2), 241–295. https://doi.org/10.1007/BF02459575
Zentmyer, G., Schieber, G., & Guillemet, F. (1988). History of the Martin Grande rootstock. *California Avocado Society, 72*, 121–125.