Original Research Article||

Genetic Diversity of Avocado (*Persea americana Mill.*) From Southern Ethiopia Using SSR Markers

Almaz Tadesse[1], Marie Kalousová[2], Hewan Demissie Degu[3*]

[1]*Mizan Tepi University, Mizan Aman, 260, South West Ethiopia, Ethiopia,*

[2]*Czech University of Life Sciences Prague, Kamýcká 129, Prague, 165 00, Prague, Czech Republic,*

[3]*Hawassa University, College of Agriculture, School of Plant and Horticultural Sciences, Hawassa, 5, Sidama, Ethiopia,*

Abstract

Avocado (*Persea americana* M.) Is a vital cash crop in Ethiopia. Understanding its population structure and allelic poly morphism is crucial for genetic improvement. However, the genetic diversity of Ethiopian avocado remains underexplor ed. This study investigated the genetic diversity of avocado in Southern Ethiopia using Simple Sequence Repeat (SSR) markers. A total of 109 avocado trees from 16 districts were sampled, and 12 SSR markers were employed for analysis. The study detected 140 alleles across the 12 loci, averaging 11.7 alleles per locus. The average expected heterozygosity was 0.63 ± 0.12, while the observed heterozygosity was 0.48 ± 0.19, with all loci showing significant deviation from har dyweinberg Equilibrium (HWE). The Analysis of Molecular Variance (AMOVA) indicated that approximately 5% of th e genetic variation was among the 16 populations. Pairwise comparisons of population FST values revealed a lack of gen etic differentiation in seven out of 48 paired comparisons. The Unweighted Pair Group Method with Arithmetic Mean (UPGMA), based on the Nie and Li similarity index, grouped the genotypes into three major clusters with bootstrap value s of 100 and 5. Genetic distance analyses showed mixing of avocado trees from different districts. Discriminant Analysis of Principal Components (DAPC) categorized the samples into three groups, while model-based structure analysis subdivided them into two main genetic clusters. The moderate genetic diversity observed in the avocado germplasm is promising for the future of avocado cultivation in Ethiopia, suggesting the germplasm is a valuable source of alleles for genetic improvement. The mixing of avocado trees across districts indicates strong gene flow among populations. Furthermore, the significant variation among tree populations from different districts offers numerous opportunities for avocado breeding programs, providing hope for potential advancements in avocado cultivation.

Key words: Avocado, Genetic diversity, SSR markers, Population structure, Southern Ethiopia

Original submission: August 12, 2024; **Revised submission**: December 09, 2024; **Published online**: December 30, 2024

***Corresponding author's address:** Hewan Demissie, Email: hewand@hu.edu.et

Authors: Almaz Tadesse: almintadesse@gmail.com; Marie Kalousová: Kalousovam@ftz.czu.cz

INTRODUCTION

Avocado (*Persea americana* Mill) originated in Mexico, Central or South America, and was first cultivated in Mexico as early as 500 BC (Galindo-Tovar et al., 2008). Avocado is a highly heterozygous diploid species with 24 chromosomes (2n = 24) and a genome size of 841.6 Mbp (Yang et al., 2024). It is a cross-pollinating species with outcrossing rates ranging from 74% to 96% (Borrone et al., 2008). The plant produces edible, nutritious, and commercially used fruits.Avocado is a polymorphic species with numerous taxa adapted to different climates and altitudinal ranges. These include P. americana var. drymifolia, P. Americana var. Guatemalans, and P. Americana var. Americana is commonly called the Mexican, Guatemalan, and West Indian horticultural races, respectively (Galindo-Tovar et al., 2008). The Mexican and Guatemalan races are adapted to cooler

climates, whereas the West Indian race requires warmer weather for optimum development. The three avocado races are cross-compatible, and hybridization can occur between trees of different races when grown near each other (Juma et al., 2020). Most commercial avocado cultivars are interracial hybrids developed from chance seedlings. Thus, agriculturally important cultivars in subtropical climates, such as "Hass," "Bacon," and "Fuerte" are Guatemalan-Mexican hybrids with different degrees of hybridization (Boza et al., 2018). Genetic characterization and diversity analysis in avocado enable researchers to achieve agronomic improvement by developing disease-resistant new varieties and cultivars that are more profitable in terms of fruit production, fruit quality, and fruit maturation precocity (Guzman et al., 2017).

Avocado was first introduced to Ethiopia around 1938 by privet orchardists in Hirna (Eastern Highlands of Ethiopia) and Wando-genet (Southern Highlands of Ethiopia) and has since been distributed to different ago-ecological zones in Ethiopia (Berhanu, 2013). Despite its long history of introduction and the diverse agroecology of Ethiopia, its distribution is limited to a few areas of the country (Jalata, 2021). Currently, the main avocado-producing areas in Ethiopia include the Sidama and Wolaita areas in the south, the Jimma and Mizan areas in the southwest, and the Hararge area in the country's eastern region.

In Ethiopia, avocado trees are mainly grown as an integral component of coffee (Coffea arabica L.) and enset (Ensete ventricosum) production systems (Biazin et al., 2018). Smallholder producers consider avocados to be shade trees and benefit from the sale of avocado fruits (Biazin et al., 2018).

The potential for commercial avocado production in Ethiopia is promising, with a recent surge in interest (Terheggen, 2019). Although several high yielding andbiotic stressresistant varieties are suitable for commercial farming, they are currently scarce in Ethiopia. This underscores the need to gather, charact erize, and select accessions from the existing germpla sm to achieve higher yield and quality. However, ther e is limited information available on the genetic diver sity of avocados in Ethiopia, which is a crucial factor for successful breeding and conservation efforts. This lack of comprehensive data remains a significant cha llenge for advancing avocado breeding programs.

Molecular markers are used for the genetic characterization of germplasm because they exhibit higher heritability and greater polymorphism than morphological markers, making them more effective for distinguishing between genotypes (Bunjkar et al., 2024). Additionally, breeding perennial crops poses challenges due to their long life spans.

Molecular markers, such as Random Amplified Polymorphic DNA (RAPD), Simple Sequence Repeats (SSRs), and Single Nucleotide Polymorphisms (SNPs), are employed to identify accessions and variations at the earliest stages of development. These markers have been used to analyze the genetic diversity of avocado accessions globally (Degu et al., 2024). In Ethiopia, SSR markers are particularly preferred for avocado diversity analysis because of their high polymorphism and their ability to detect numerous alleles at specific loci. This makes SSRs highly suitable for genetic diversity assessments.

As co-dominant markers, SSRs allow for the differentiation of both homozygous and heterozygous individuals, providing a comprehensive view of the genetic structure within populations. Moreover, SSR markers are reproducible, reliable, relatively cost-effective, and robust tools for genetic mapping and population structure analysis. Their versatility is invaluable for identifying genomic regions associated with key traits and exploring gene flow and genetic differentiation across avocado populations in Ethiopia.

Although avocado cultivation in Ethiopia has a long history, comprehensive reports have yet to be made on avocado diversity. This research was aimed to fill this gap by using SSR markers to estimate the population structure and diversity of Ethiopian avocados. The findings of this study will enhance the understanding of avocado diversity in Ethiopia and will be helpful for future research and breeding programs.

MATERIALS AND METHODS

One hundred nine avocado accessions, from three regions in southern Ethiopia and from a germplasm bank at the Wondo-Genet Research Center (WGRC), were used for genetic diversity analysis (Fig. 1 and Table S1). Young leaves were collected in a falcon tube (15 ml) filled with the color indicator Silica Gel Desiccant and shipped to the Czech University of Life

Sciences Prague, Faculty of Tropical Agrisciences, Prague, for SSR marker molecular analysis.

Total genomic DNA was isolated using a modified cetyltrimethylammonium bromide (CTAB) extraction protocol developed by (Doyle and Doyle, 1987). The concentration and purity of DNA were measured on a nanodrop spectrophotometer. The DNA samples were diluted to final concentrations of 20 ng/µl for polymerase chain reactions (PCR).

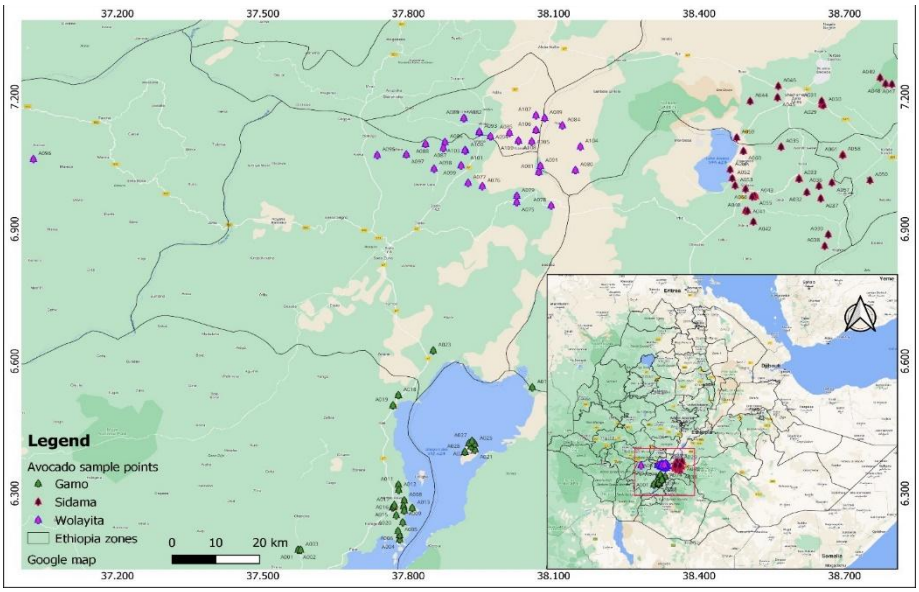

Figure 1. Map of Southern Ethiopia, displaying the geographical location of sampled avocado populations.

Twelve microsatellite primers (Ashworth, 2003; Sharon et al., 1997) (Table 1) were used to analyze the collected avocado samples. The methodology was adapted from Ruiz-chutan et al. 2022. Forward primers were fluorescently labeled for detection in four colors: 6-FAM, PET, NED, and VIC. PCR amplifications were conducted separately for each multiplex in a T100 Thermal Cycler (Bio-Rad, USA) by performing the following steps: Denaturation at 95 °C for 5 minutes, followed by 35 cycles of 30 seconds each, a specific alignment temperature for each multiplex (one minute), elongation at 72 °C for one minute, and a final extension of 10 minutes at 72 °C. Fragment analysis was performed with the Genetic Analyzer 3500 (Applied Biosystems, USA). Microsatellite alleles were scored using GENEMARKER v. 2.4.0 (Soft Genetics, USA).

Table 1. Diversity Statistics of the 12 SSR loci in 109 avocado samples in sixteen districts in southern Ethiopia.

Locus	SSR Motif	Na	Ne	Ho	He	I	HW	PIC
AVAG05	(AG)10	10	2.03	0.3	0.51	1.19	***	0.72
AVAG13	(CT)18	16	5.15	0.72	0.8	1.92	***	0.87
AVT436	(AG)20	6	2.33	0.58	0.57	1.01	***	0.68
AVAG11	(ATC)9	11	2.38	0.28	0.58	1.22	***	0.78
AVAG07	(CT)22	11	2.39	0.62	0.58	1.21	***	0.7
AVMIX04	(GA)15	13	2.8	0.49	0.64	1.34	***	0.82
AVAG21	(AG)12 (CAA)5 (ACAG)10	9	4.48	0.62	0.78	1.68	***	0.85
AUCR418	(TC)14	15	5.11	0.58	0.8	1.88	***	0.89
AVAG22	(TC)15	10	1.76	0.42	0.43	0.94	***	0.65
AVAG25	(GT)12(GA)13	12	1.8	0.07	0.44	0.93	***	0.68
AVD001	(CT)12	15	4.65	0.46	0.78	1.81	***	0.97
AVD022	(TC)13	12	3.41	0.67	0.71	1.46	***	0.89
Mean		11.67	3.19	0.48	0.63	1.38	***	0.79

NA, Number of alleles; Ne, Effective number of alleles; Ho, observed heterozygosity; He, expected Observed heterozygosity; I, Shannon's Information Index; HWE, Hardy Weinberg Equilibrium; PIC, Polymorphic Information content

Data Analysis

The corresponding peaks' peak features and fragment sizes were counted in the data. The GenAlEx (Peakall and Smouse, 2012) program (version 6.5.03) was used to estimate locus-based diversity indices such as the number of alleles (Na), an adequate number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity (He), or gene diversity (GD). HWE, or inter-population diversity, was calculated with Popgene32 software version 1.32 (Yeh and Boyle, 1997). AMOVA and average gene diversity were estimated across the loci for each population using the GenAlEx version 6.5.03 program (Peakall and Smouse, 2012). The GenAlEx was used to compute fixation indices (FST, FIT, FIS, FCT, and FSC) and pairwise comparisons between populations. Nei's genetic distance was computed in GenAlEx and imported into MEGA6 (Tamura et al., 2013), where the dendrogram in the Newick format was produced using the unweighted pair group method with arithmetic mean UPGMA. The dendrogram was visualized and customized in the Interactive Tree of Life (iTOL) version 4 (Letunic and Bork, 2019). Population structure is constructed using a Bayesian algorithm implemented in Structure software, version 2.3.4 (Pritchard et al., 2000). The analysis was based on an admixture ancestral model with correlated allele frequencies. To determine the number of population clusters (K), a burn-in period of 50,000 and a run length of the Monte Carlo Markov chain (MCMC) of 100,000 were used for K = 1 to K = 10, using ten replications for each K in the STRUCTURE software. The optimum K value was predicted using POPHELPER, an R package (Francis, 2017).

RESULTS AND DISCUSSION

SSR Polymorphism and Genetic Diversity

In this study 12 loci on 109 sampled avocado plants yielded a significant discovery of 140 distinct alleles (Table 1). The average number of alleles per locus was 11.70 ± 2.84 (Table 1). Each marker detected at least six alleles and was polymorphic, a finding that adds to our understanding of avocado plant diversity. AVAG13 was the locus with the most alleles (16), followed by 15 (AUCR418 and AVD001). Locus AVT436 has the lowest alleles (6). Similar analyses detected 11.5 alleles using 10 SSR loci across 71 avocado plants, with alleles ranging from 5 to 22 per locus (Janice and Jemmy, 2014), and 167 alleles using 10 SSR loci across 226 plants, with the number of alleles per locus ranging from 10 to 23 (Juma et al., 2020). The smallest number of alleles per locus, 3.1, was reported by (Liu et al., 2020) for 56 avocado trees examined in Hainan Province, China. In Gross-German and Virue2013) and Janice and Jemmy (2014) reports, it was 11.5. The differences between the present study and the previously reported results might be due to the markers' polymorphism, sample size, population size, the diversity of the germplasm investigated, and the platforms employed for quantifying amplified products. The quality of genomic DNA used in PCR amplification, optimization of PCR protocols, and differences in allele scoring accuracy could also account for these differences. Capillary electrophoresis might have brought about the difference in the number of alleles

discovered in avocado trees from previous studies, except published by Schnell et al., (2003).

The effective allele count varied from 5.15 (AVAG13) to 1.76 (AVAG22). Shannon's information index (I) range was 1.19 (AVAG05) to 0.93 (AVAG25). The average observed heterozygosity was 0.63 ± 0.19, with minimum and maximum values of 0.43 (AVAG22) and 0.83 (AVAG13 and AUCR418), respectively. The average polymorphism information content was 0.79 ± 0.04, indicating a wide range of genetic variation from 0.65 (AVAG22) to 0.97 (AVD001). All loci deviated from the Hardy-Weinberg equilibrium, suggesting the collected avocado samples do not represent random sampling. Potential factors such as non-random mating, genetic drift, or selection may influence the avocado populations' genetic structure. The wide range of polymorphism information content is a significant finding that highlights the genetic diversity among the sixteen avocado populations. In the current analysis, there were 12 alleles on average per locus. According to Juma et al. (2020), Schnell et al., and Guzmán et al. there are more alleles per locus (16.7, 18.8, and 19.5, respectively).

The average observed heterozygosity was 0.48, comparable to the levels of genetic diversity shown for the studied samples (Boza et al., 2018; Janice and Jemmy, 2014). The current observed heterozygosity (0.48), were lower than 0.65 (Schnell et al., 2003), and 0.6 (Guzmán et al., 2017) respectively. This suggests that avocado from Ethiopia has a limited genetic foundation. This finding coincides with the fact that the primary introduction of avocado in Ethiopia was by orchids (Berhanu, 2013). Observed heterozygosity with a lower value (0.39) has been reported by Liu et al. (2020). The finding suggests in the number of populations, and types of accessions used may account for the variation in the allele's number and heterozygosity.

Shannon's information index (I) is another metric for analyzing gene diversity. If markers change in their score and score close to one on the Shannon information index, they indicate the presence of variability and their suitability for genetic diversity investigations in that group (Nassiry et al., 2009). The results of this study's measurements (0.93 to 1.92), which supported the existence of variability and the usefulness of the markers for examining genetic diversity in the Ethiopian avocado population, were thus confirmed. The PIC value of each marker employed in this study was more informative and scored 0.65 and above, indicating that they are all highly informative and enable the diverse investigation of Ethiopian avocado accessions (Smith et al., 1997). Overall, the findings demonstrated the presence of increased genetic diversity, the strength of marker polymorphisms, and their capacity to provide pertinent information that aids in developing effective conservation and avocado improvement programs.

Genetic Diversity among the Collected Population

When avocado samples were analyzed at the population level, the WGRC exhibited the highest number of alleles (Na), 8.1± 0.6, followed by the Haroma (4) and Sake (3.9) with the average allele of 3.5 ± 0.1 (Table 2). On the other hand, Sadeka population had the lowest (2.7 ± 0.2) number of alleles (Table 2). The effective number of alleles also showed similar pattern where WGRC has the highest (5.8) followed by Yirbaduwanch (2.9) and Sake (2.8) with an average value of 2.6±0.1 among the population. The observed heterozygosity (Ho) varied from 0.6 for the WGRC, Sadake, and Sake population to 0.4 for the majority of the population. Additionally, the expected heterozygosity (He) was higher in the WGRC (0.8), followed by Adehora, Haroma, and Sake (0.6), respectively (Table 3). The Shannon's Information Index (I) value for WGRC (1.9) was the highest (Table 3) followed by the Adeyhora population (1.7 ± 0.01), and the lowest for Sadamodicha, Youwo, Sadeka, Xarmesa, Dalboatuawaro, and Warzalasho, at 0.8 ±0.1. Since gene diversity expresses the probability in which two randomly chosen alleles from the population are different, a Nei's unbiased gene diversity was calculated (Table 2). Genetic diversity was high for WGRC (0.8), followed by Haroma (0.7), and Yirbaduwanch (0.7) with an average value of 0.6±0.01 (Table 2).

Table 2. Diversity information among the sixteen geographic populations (districts)

Population	N	Na	Ne	I	Ho	He	Gene Diversity
Shara	10	3.6±0.5	2.3±0.3	0.9±0.13	0.4±0.1	0.5±0.1	0.5±0.1
Mole	10	3.5±0.5	2.6±0.3	1.0±0.1	0.5±0.8	0.6±0.5	0.6±0.5
Delbo	8	3.2±0.5	2.6±0.4	0.9±0.2	0.5±0.1	0.5±0.1	0.5±0.1
Haroma	6	4.0±0.3	2.7±0.2	1.1±0.1	0.6±0.1	0.6±0.02	0.7±0.03
Youwo	6	2.8±0.3	2.1±0.2	0.8±0.1	0.4±0.1	0.5±0.1	0.5±0.1
Sadeka	5	2.7±0.2	2.2±0.2	0.8±0.1	0.6±0.1	0.5±0.05	0.6±0.06
Xarmesa	5	2.8±0.3	2.1±0.2	0.8±0.1	0.4±0.1	0.5±0.05	0.5±0.06
Sadamodicha	5	2.9±0.3	2.1±0.2	0.8±0.1	0.5±0.1	0.5±0.05	0.5±0.06
Yirbaduwancho	6	3.5±0.3	2.9±0.3	1.1±0.1	0.5±0.1	0.6±0.05	0.7±0.05
WGRC	12	8.1±0.6	5.8±0.5	1.9±0.1	0.6±0.07	0.8±0.02	0.8±0.02
Adeyhora	6	3.7±0.4	2.6±0.2	1.0±0.1	0.6±0.1	0.6±0.1	0.6±0.1
Adeykoisha	6	3.4±0.4	2.5±0.3	1.0±0.1	0.5±0.1	0.5±0.05	0.6±0.05
Dalboatuawaro	6	2.8±0.2	2.2±0.2	0.8±0.1	0.4±0.1	0.5±0.05	0.6±0.05
Warzalasho	8	2.8±0.3	2.2±0.2	0.8±0.1	0.4±0.1	0.5±0.1	0.5±0.1
Galacha	5	2.8±0.2	2.2±0.2	0.9±0.1	0.5±0.7	0.5±0.4	0.6±0.5
Sake	5	3.9±0.4	2.8±0.3	1.1±0.1	0.6±0.1	0.6±0.1	0.6±0.1
Mean	7	3.5±0.1	2.6±0.1	1.0±0.03	0.5±0.02	0.5±0.01	0.6±0.01

N, Number of samples; NA, number of allele; Ne, Effective number of alleles; Ho, observed heterozygosity; He, expected Observed heterozygosity; I, Shannon's Information Index.

Molecular Variance and Population Divergence

The AMOVA revealed a low but significant ($p < 0.001$) difference among avocado samples collected from different districts, and region which accounted for 5% the total variation (Table 3). However, a lower differentiation (1.0; p value < 0.048) was obtained among groups of samples collected from the two agroecology zones in the study site (Table 3). The variation among individuals within districts and region (20%), and 75% within individuals in the analysis (Table 3) were highly significant ($P < 0.0001$).

Although avocado is usually considered an out-crossing plant, the findings might suggest it does much self-pollination. Another explanation could be that the significant divergence from HWE may be due to linkage disequilibrium with the studied loci under selection in the form of heterozygote disadvantages. The results are more significant than those of Juma et al. (2020), who discovered a difference (1.98%) among avocado population groups in the collection site.

The fixation index, or FIS, is a measure of inbreeding that determines whether a sub population has an excess or deficit of homozygotes (Chakraborty, 1993). Calculation of fixation index indices, based on a prior population hierarchy for the district, region and agroecology of the avocado, using AMOVA, resulted in FST (0.05), FIS (0.02) and FIT (0.205) values (Table 3). The fixation index (FST), which runs from 0 to 1, measures sub-population differentiation. A value near 0 indicates a complete panmixia population, while a number close to 1 indicates fully differentiated populations (Pearse and Crandall, 2004). The current result is indicating the existence of a low district and region based subdivision. This could result from genetic drift, and pressure from indirect selection, often leading to the loss of specific alleles or a change in frequency.The FST values found in the current inquiry were 0.19, 0.22, and 0.25, respectively which were lower than those previously reported by Boza et al. (2018), Guzmán et al. (2017) and Gross-German & Viruel (2013), In contrast to our research, Boza et al. (2018) examined avocado samples from Mexico and the United States that represented Persea americana, P. nubigena, and P. kruguii. Only P. americana that had been gathered in a single country, including regional cultivars, rootstocks, and local selections, was investigated by Guzmán et al. (2017). Samples characterized by Gross-Guzman et al. (2013) comprised P. longipes, P. nubigens, and P. schiedeana. However, the current study's average FST is comparable to Gomez, and Rivera (2017) for 226 avocado samples from Tanzania. The sixteen districts were responsible for 5% of the total genetic diversity (Table 3).

Table 3. AMOVA using 1000 permutations for avocado trees from different districts, region and altitudes

Source	DF	SS	MS	Est. Var.	Perc. Var.	F-Statistics	P-value
Between Districts	15	105.355	7.024	0.192	5	FST = 0.05	0.001
Between samples within District	93	412.063	4.431	0.757	20	FIS = 0.206	0.001
Within Samples	109	318.0	2.917	730.5	75	FIT = 0.245	0.001
Total	217	835.417		3.866	100	Nm =4.8	
Between Region	3	44.183	14.7	0.2	5	FST =0.05	0.001
Between samples within region	105	473.6	4.5	0.8	20	FIS = 0.2	0.001
Within Samples	109	318	2.9	2.9	75	FIT = 0.2	0.001
Total	217	835.4		3.9	100	Nm =4.76	
Between Agroecology	1	6.872	6.872	0.025	1	FST = 0.006	0.048
Between samples within Argo-ecology	107	510.545	4.771	0.927	24	FIS = 0.241	0.001
Within Samples	109	318.0	2.9	2.9	75	FIT = 0.246	0.001
Total	217	835.4		3.9	100	Nm =38.9	

The pairwise FST analysis reveals significant genetic differentiation among the studied populations (Table 4). With 69 out of 120 population pairs exhibiting significant differences ($p<0.001$), it is clear that there is substantial genetic structuring within the sampled groups. The highest FST value of 0.48 observed between the Shara and Warzalasho population indicating a high level of genetic differentiation. Such a high value suggests limited gene flow and genetic isolation between these populations. This could be due to geographical barriers, behavioral differences, or historical separations that have prevented cross pollination. Similarly, the second-highest FST value of 0.47 between Adekoisha and Yaowo ($p<0.05$) further supports the presence of significant genetic differentiation within these groups. The lower p-value threshold ($p<0.05$) indicates that while the differentiation is significant, it may not be as robust as those with $p<0.001$.

The pairwise FST analysis reveals significant genetic differentiation among the studied populations (Table 4). With 69 out of 120 population pairs exhibiting significant differences ($p<0.001$), it is clear that there is substantial genetic structuring within the sampled groups. The highest FST value of 0.48 observed between the Shara and Warzalasho population indicating a high level of genetic differentiation. Such a high value indicates limited gene flow and genetic isolation between these populations. This could be due to geographical barriers, behavioral differences, or historical separations that have prevented cross pollination. Similarly, the second-highest FST value of 0.47 between Adekoisha and Yaowo ($p<0.05$) further supports the presence of significant genetic differentiation within these groups. The lower p-value threshold ($p<0.05$) indicates that while the differentiation is significant, it may not be as robust as those with $p<0.001$.

Table 4. Population pairwise FST comparisons between Sixteen avocado populations

Sha	Sha	Mol	Del	Har	You	Sa	Xar	Sad	Yir	WG	Ad	Ade	Dal	War	Gal
Mol	0.02														
Del	0.47*	0.21*													
Har	0.13*	0.42*	0.09												
You	0.41*	0.37*	0.39*	0.08											
Sa	0.31*	0.43*	0.25*	0.38*	0.30*										
Xara	0.34*	0.44*	0.32*	0.08	0.41*	0.33*									
Sad	0.35*	0.08	0.13	0.31*	0.19*	0.11	0.11								
Yir	0	0.15*	0	0.02	0	0.08	0.02	0							
WG	0	0.17*	0.01	0.37*	0.09	0.35*	0.32*	0.26*	0.25*						
Ad	0.39*	0.44*	0.40*	0.14*	0.41*	0.33*	0.41*	0.11	0.01	0.05					
Ade	0.31*	0.15*	0.46*	0.25*	0.47*	0.48*	0.48*	0.35*	0	0.01	0.46*				
Dal	0.41*	0.02	0.36*	0	0.41*	0.02	0.11	0	0	0.01	0.41*	0.46*			
War	0.48*	0.03	0.38*	0.02	0.40*	0.04	0.12	0.14	0	0	0.40*	0.40*	0.16*		
Gal	0.27*	0.44*	0.40*	0.05	0.41*	0.24*	0.27*	0.02	0.01	0.08	0.40*	0.47*	0.11	0.09	
Sak	0.09	0.44*	0.03	0.15	0.05	0.25*	0.26*	0.01	0.39*	0.34*	0.16*	0.23*	0	0	0

Conversely, the lowest FST value of 0.14 between Haroma and Adehora populations suggests a much less pronounced genetic differentiation. This value indicates a higher level of gene flow between these populations, with significant implications. It suggests that these populations are geographically closer, have fewer barriers to gene flow, or share a more recent common ancestry, leading to a more homogenized genetic landscape.

The observed range of FST values underscores the complex population structure and varying levels of genetic connectivity among the studied groups. High FST values suggest that some populations are evolving independently, a process that could lead to local adaptations (De Villemereuil and Gaggiotti, 2015). On the other hand, low FST values show regions of connectivity where gene flow is sufficient to homogenize genetic differences (Izaguirre-Toriz et al., 2024). The influence of geographical barriers and historical events on these values is significant, shaping the genetic structure of these populations. Further research, including detailed ecological studies, is necessary to fully understand the mechanisms driving these genetic patterns and to develop effective conservation and breeding plans.

Hierarchical Cluster Analysis

The genetic distance matrix and the subsequent UPGMA-based dendrogram (Fig. 2) shows the genetic relationships and population structure among the studied avocado samples. These tools help elucidate the underlying genetic differentiation and connectivity within the population by grouping various districts and regions. Populations with lower genetic distances cluster together, indicating closer genetic relationships. Conversely, populations with higher genetic distances form separate clusters, reflecting their genetic isolation. The dendrogram highlights distinct genetic groupings, offering insights into historical and contemporary gene flow patterns (Rheindt and Edwards, 2011). The clustering pattern may reflect ecological niches, or breeding practices that have shaped the current genetic structure.

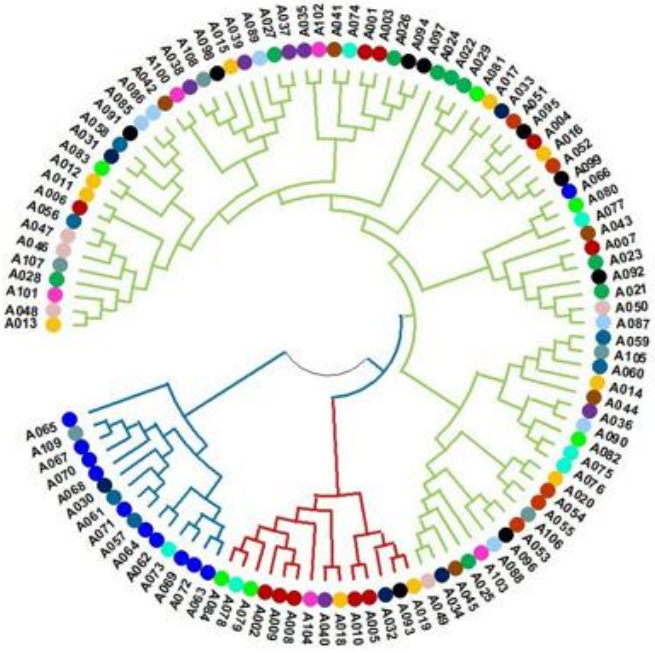

Figure 2. Dendrogram of the 109 avocado trees constructed with UPGMA showing genetic relationships between the analyzed samples. Samples collected from a common district are represented by the same circle color. The different phylogenetic groups are represented by red, blue, and green line colors for Group I, Group II, and Group III phylogenetic groups, respectively

The Discriminant Analysis of Principal Components (DAPC) method, a powerful tool in genetic analysis, has unveiled three distinct clusters among the avocado accessions (Fig. 3). This revelation provides profound insights into the genetic relationships and population structure. The first cluster, an exclusive domain of WGRC accessions, signifies a unique genetic identity for these accessions, possibly owing to their distinct genetic background, selective breeding practices, or specific adaptations to their native environments. The second and third clusters, housing accessions from various districts, hint at a more intricate genetic landscape. These clusters likely represent accessions that have experienced different evolutionary pressures, gene flow, and breeding practices compared to the WGRC accessions. The distinct clustering of WGRC accessions underscores the genetic differentiation between regionally diverse accessions. This differentiation could be attributed to several factors: WGRC accessions may originate from geographically isolated areas (Berhanu, 2013), reducing gene flow with other populations and increasing genetic differentiation. WGRC accessions might have been subject to specific selective breeding practices to preserve certain traits, further distinguishing them from other accessions. Unique environmental conditions and selective pressures in the WGRC regions could have led to the development of distinct genetic traits.

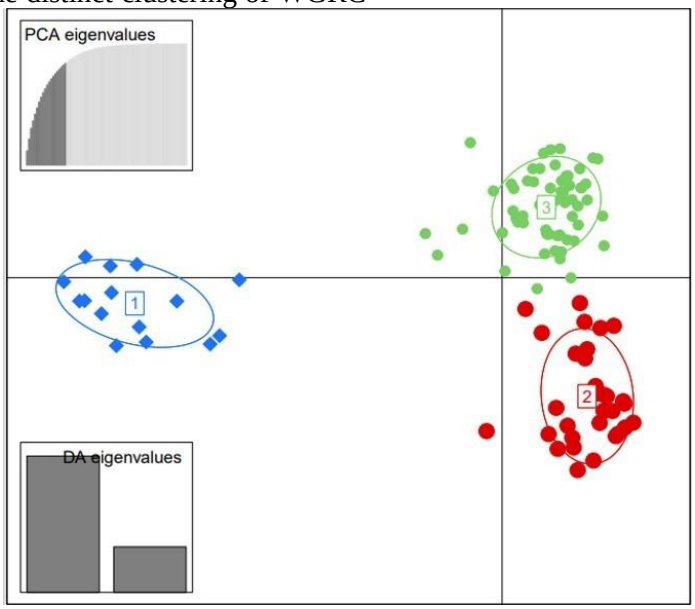

Figure 3. Principal component analysis with discrimination (DAPC) for 109 avocado samples. The first two Linear Discriminants are represented by the axes (LD). Each symbol stands for individual sample, and each circle for a group. The different subpopulations identified by DAPC are represented by numbers.

The clustering of accessions from diverse districts into three separate groups is a testament to the varying levels of genetic differentiation within these regions. This understanding is crucial as it can help us comprehend the different historical events that contributed to the genetic diversity within these clusters. The varied ecological niches across districts could drive adaptation to different environmental conditions, leading to genetic divergence (Sexton et al., 2017). The extent and direction of gene flow between districts can significantly influence genetic structure. These clusters reflect areas with more or less gene exchange. The distinct genetic identity of the WGRC accessions suggests they are valuable genetic resources. Conservation efforts should prioritize preserving their unique genetic traits. Understanding the genetic differentiation within the second and third clusters is not just a scientific endeavor, but a call to action to help identify genetically diverse populations and prioritize them for conservation to maintain overall genetic diversity. The distinct clusters, as revealed by the DAPC method, can serve as a beacon for breeding programs. They identify genetically diverse accessions that can be harnessed to introduce new traits and enhance genetic diversity. The unique genetic traits of the WGRC accessions, a treasure trove for breeding programs, could be instrumental in improving specific traits, such as disease resistance or environmental adaptability. This potential for

enhancement and diversification is a promising prospect for the future of avocado cultivation.

The dendrogram and DAPC could not separate trees based on their districts or regions. This was in line with the findings of AMOVA (Table 3), which showed that 93% of the total genetic variation was between the groups. The dendrogram and DAPC findings were further supported by the population pairwise FST results, which indicated no differences between Shara against Mole, Haroma, WGRC, and Sake. A possible explanation is the trading of avocado seeds, which may have caused genetic mixing in avocado populations similar to the findings from Tanzania (Juma et al., 2020; Juma et al., 2019). The genetic blending of avocado populations is brought by the exchange of seedlings between growers in different districts and the exportation of avocado produce from one area to another, where seeds are later sown or spontaneously sprouted. It might also be due to the introduction of genetic material that is considerably similar to different districts or regions (Juma et al., 2019).

The model-based STRUCTURE showed two ancestral groups (Fig. 4(a)) in the Ethiopian avocado accession. This finding is supported by the population pairwise FST values, which show a lack of difference between the pairs from the same structured population (Table 4). Accordingly, each population (district) has at least two clusters. Numerous indices, including AMOVA, FST, UPGMA, and DAPC, have demonstrated the genetic similarity and low but significant genetic divergence of avocados grown in avocado-producing areas in Ethiopia.

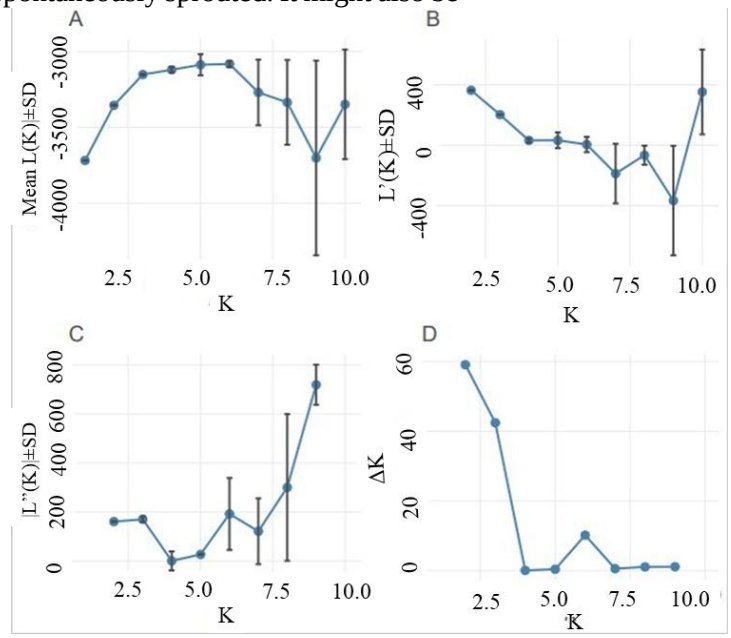

Figure 4a. Estimation of the number of groups in the avocado collection

Figure 4b. Population structure of 109 avocado accessions using STRUCTURE, with each individual represented by a vertical bar, using the expected ideal K = 2, and K =3 populations.

The genetic structure of avocado accessions was obtained using STRUCTURE analysis (Fig.4). The model-based Bayesian algorithm grouped the accessions into two clusters (K=2). This population structure bar plot shows no clear geographic-origin-based structure, indicating a complex genetic structure of avocado populations that may have implications for breeding and conservation strategies.

These findings were supported by DAPC and the UPGMA dendrogram tree, which showed that the 109 examined trees could be separated into three genetic clusters. Due to the genetic material sharing, populations from different growing areas showed less differentiation, indicating population mixing. A shared ancestor, extensive planting material trade, and high levels of gene flow could all contribute to this explanation. Regardless of the geographic distance between populations, the interchange of plant material between locations increases gene flow and the dispersion of alleles. Such a trend maximizes genetic variability among individuals but reduces diversity among the population.

CONCLUSIONS

This study is the first to use micro-satellites to analyze avocado trees introduced to Ethiopia with only three trees. We found a low level of genetic variation in the studied germplasm, as indicated by several diversity indicators. We recommend creating a core collection of the genotypes from the entire range because the genetic diversity of populations is low. The UPGMA tree, DAPC, and STRUCTURE analysis showed mixed trees from different regions and populations, indicating a moderate but significant population structure and a high degree of gene flow between the groups. Ethiopia has diverse avocado germplasm due to gene transfer between populations.

CONFLICTS OF INTEREST

Authors declare that they have no conflicts of interest.

ACKNOWLEDGEMENT

The authors would like to acknowledge the smallholder farmers of the Sidama and the Southern Ethiopia regional governments for allowing us to collect the avocado accessions' leaves for this study.

REFERENCES

Ashworth, V.E.T.M. 2003. Microsatellite Markers in Avocado (*Persea americana* Mill.): Genealogical Relationships Among Cultivated Avocado Genotypes. J. Hered., 94(5): 407–415. doi:10.1093/jhered/esg076. [Scholar Google]

Berhanu, M. 2013. The role of avocado production in coffee-based farming systems of South Western Ethiopia: The case of Jimma zone. African Journal of Agricultural Research, 8(19): 2285–2296. doi:10.5897/AJAR11.2332. [Scholar Google]

Biazin, B., Haileslassie, A., Zewdie, T., Mekasha, Y., Gebremedhin, B., Fekadu, A., Shewage, T. 2018. Smallholders' avocado production systems and tree productivity in the southern highlands of Ethiopia. Agroforestry Systems, 92: 127–137. doi:10.1007/s10457-016-0020-2. [Scholar Google]

Borrone, J.W., Olano, C.T., Kuhn, D.N., Brown, J.S., Schnell, R.J., Violi, H.A. 2008. Outcrossing in Florida Avocados as Measured Using Microsatellite Markers. Journal of the American Society for Horticultural Science, 133(2): 255–261. doi:10.21273/JASHS.133.2.255. [Scholar Google]

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: 1195–1215. doi:10.1007/s10722- 018-0608-7. [Scholar Google]

Bunjkar, A., Walia, P., Sandal, S.S. 2024. Unlocking Genetic Diversity and Germplasm Characterization with Molecular Markers: Strategies for Crop Improvement. Journal of Advances in Biology & Biotechnology, 27(6): 160–173. doi:10.9734/jabb/2024/v27i6873. [Scholar Google]

Chakraborty, R. 1993. Analysis of Genetic Structure of Populations: Meaning, Methods, and Implications, in: Majumder, P.P. (Ed.), Human Population Genetics. Springer US, Boston, MA, pp. 189–206. doi:10.1007/978-1-4615-2970-5_14.

De Villemereuil, P., Gaggiotti, O.E. 2015. A new F_{ST} -based method to uncover local adaptation using environmental variables. Methods in Ecology and Evolution 6: 1248–1258. doi:10.1111/2041-210X.12418. [Scholar Google]

Degu, H.D., Ruiz-Chután, J.A., Kalousová, M. 2024. Genetic diversity and population structure analysis of avocados (Persea americana Mill.) from Southern Ethiopia and Guatemala using polymorphic simple sequence repeat (SSR) markers. Genetic Resources and Crop Evolution, 71: 3331-3346. doi:10.1007/s10722-023- 01831-1. [Scholar Google]

Doyle, J.J., Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull., 19(1): 11–15. [Scholar Google]

Francis, R.M., 2017. R package and web app to analyses and visualize population structure. Molecular Ecology Resources 17, 27–32. doi:10.1111/1755-0998.12509. [Scholar Google]

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

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 & Genomes, 9(2): 539–555. doi:10.1007/s11295-012-0577-5. [Scholar Google]

Guzmán, L.F., Machida-Hirano, R., Borrayo, E., Cortés-Cruz, M., Espíndola-Barquera, M.D.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. doi:10.3389/fpls.2017.00243. [Scholar Google]

Izaguirre-Toriz, V., Aguirre-Liguori, J.A., Latorre-Cárdenas, M.C., Arima, E.Y., González-Rodríguez, A. 2024. Local adaptation of Pinus leiophylla under climate and land use change models in the Avocado Belt of Michoacán. Molecular Ecology, 33, e17424. doi:10.1111/mec.17424. [Scholar Google]

Jalata, Z., 2021. Current Status, Potentials and Opportunities of Avocado Production as an Alternative Crop: The Case of Ethiopia: A Review. Agricultural Reviews, 42(3): 335-341. doi:10.18805/ag.R-179. [Scholar Google]

Janice, D.A., Jemmy, F.T., 2014. Genetic characterisation of avocado (Persea americana Mill.) in two regions of Ghana. African Journal of Biotechnology, 13(51): 4620–4627. doi:10.5897/AJB2014.14023. [Scholar Google]

Juma, I., Geleta, M., Nyomora, A., Saripella, G.V., Hovmalm, H.P., Carlsson, A.S., Fatih, M., Ortiz, R. 2020. Genetic diversity of avocado from the southern highlands of Tanzania as revealed by microsatellite markers. Hereditas, 157(1): 40. doi:10.1186/s41065-020-00150-0. [Scholar Google]

Juma, I., Fors, H., Hovmalm, H. P., Nyomora, A., Fatih, M., Geleta, M., Carlsson, A. S., & Ortiz, R. O. 2019. Avocado Production and Local Trade in the Southern Highlands of Tanzania: A Case of an Emerging Trade Commodity from Horticulture. Agronomy, *9*(11): 749. doi.org/10.3390/agronomy9110749. [Scholar Google]

Letunic, I., Bork, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research, 47(1): 256–259. doi:10.1093/nar/gkz239. [Scholar Google]

Liu, Y., Ge, Y., Zhan, R., Lin, X., Zang, X., Li, Y., Yang, 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. doi:10.3390/agriculture10040102. [Scholar Google]

Nassiry, M., Javanmard, A., Tohidi, R. 2009. Application of Statistical Procedures for Analysis of Genetic Diversity in Domestic Animal Populations. American Journal of Animal and Veterinary Sciences, 4(4): 136–141. doi:10.3844/ajavsp.2009.136.141. [Scholar Google]

Pearse, D.E., Crandall, K.A. 2004. Beyond F_{ST}: Analysis of population genetic data for conservation. Conservation Genetics, 5(5): 585–602. doi:10.1007/s10592-003-1863-4. [Scholar Google]

Pritchard, J.K., Stephens, M., Donnelly, P. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics, 155(2): 945–959. doi:10.1093/genetics/155.2.945. [Scholar Google]

Rheindt, F.E., Edwards, S.V. 2011. Genetic Introgression: An Integral but neglected component of speciation in birds. The Auk, 128(4): 620–632. [Scholar Google]

Ruiz-Chutan, J.A., Berdúo-Sandoval, J.E., Kalousová, M., Fernández, E., Ziarovska, J., Sánchéz-Pérez, A., Lojka, B. 2022. SSRs markers reveal high genetic diversity and limited differentiation among populations of native Guatemalan avocado. J. Microbiol. Biotechnol. Food Sci., 12: e6134–e6134. [Scholar Google]

Schnell, R., Brown, J., Olano, C., Power, E., Krol, C., Kuhn, D., Motamayor, J. 2003. Evaluation of Avocado Germplasm Using Microsatellite Markers. Journal of the American Society for Horticultural Science, 128(6): 881–889. [Scholar Google]

Sexton, J.P., Montiel, J., Shay, J.E., Stephens, M.R., Slatyer, R.A. 2017. Evolution of Ecological Niche Breadth. Annual Review of Ecology, Evolution, and Systematics, 48: 183–206. URL: doi:10.1146/annurev-ecolsys-110316-023003. [Scholar Google]

Sharon, D., Cregan, P.B., Mhameed, S., Kusharska, M., Hillel, J., Lahav, E., Lavi, U. 1997. An integrated genetic linkage map of avocado: Theor Appl Genet., 95(5/6): 911–921. doi:10.1007/s001220050642. [Scholar Google]

Smith, J.S.C., Chin, E.C.L., Shu, H., Smith, O.S., Wall, S.J., Senior, M.L., Mitchell, S.E., Kresovich, S., Ziegle, J. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (Zea mays L.): comparisons with data from RFLPS and pedigree: Theoretical and Applied Genetics, 95: 163–173. doi:10.1007/s001220050544. [Scholar Google]

Peakall, R. and Smouse P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics, 28(19): 2537-2539. [Scholar Google]

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution, 30(12): 2725–2729. doi:10.1093/molbev/mst197. [Scholar Google]

Terheggen, A. 2019. Ethiopia's Potential Role in the Global Avocado Market URL: doi:10.13140/RG.2.2.31755.23848.

Yang, T., Cai, Y., Huang, T., Yang, D., Yang, X., Yin, X., Zhang, C., Yang, Y., Yang, Y. 2024. A telomere-to-telomere gap-free reference genome assembly of avocado provides useful resources for identifying genes related to fatty acid biosynthesis and disease resistance. Horticulture Research, 11(7): uhae119. doi:10.1093/hr/uhae119. [Scholar Google]

Yeh, F.C. and Boyle, T.J.B. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belgian Journal of Botany, 157-157.