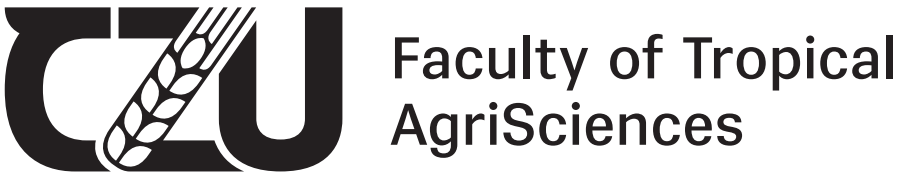
CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

FACULTY OF TROPICAL AGRISCIENCES



Assessment of morphological and genetic diversity of avocado (*Persea americana* Mill.) in Guatemala

Dissertation thesis

Study programme: Tropical Agrobiology and Bioresource Management

Author: José Alejandro Ruiz Chután, M.Sc.

Supervisor: prof. Ing. Bohdan Lojka, Ph.D.

Prague, 2024

DECLARATION OF AUTHORSHIP

I hereby declare that the content presented in this thesis, titled "Assessment of the morphological and genetic diversity of avocado (*Persea americana* Mill.) in Guatemala," submitted as a partial fulfillment of the requirements for the Ph.D. degree at the Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, is entirely my own work, unless explicitly listed in the reference section. Furthermore, I affirm that no part of this work has been submitted for any other degree at this university or any other institution.

Prague, 2024

……...

José Alejandro Ruiz Chután

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my supervisor, Bohdan Lojka, for his unwavering support, invaluable guidance, and boundless patience throughout the course of my thesis research.

I extend special thanks to my dear colleagues and friends, Julio Berdúo, Marie Kalousová, and Carlos Villanueva, who have been steadfast companions on this academic journey from its inception. Their unwavering support, guidance, and the camaraderie we shared have been instrumental at every turn.

My friends deserve heartfelt appreciation for their unwavering support, prayers, and blessings. Their encouragement provided me with the strength to persevere through challenges.

My family, my eternal source of inspiration, has offered unwavering support, prayers, and blessings throughout this journey. Their love and encouragement have been my guiding light.

Thanks to all my coauthors who have been instrumental in creating incredible scientific articles. I deeply admire each one of you for your dedication and contributions to our shared body of work.

Lastly, I express profound gratitude to God for being my ultimate companion during this academic challenge. I am thankful for the wisdom and self-belief granted, empowering me to reach this significant milestone.

Together, these remarkable individuals and forces have been instrumental in my academic achievement, and for that, I am truly thankful.

This thesis is based on the results that were published in five separate papers as follow:

Degu HD, **Ruiz-Chután JA**, Kalousová M. 2024. Genetic diversity and population structure analysis of avocados (*Persea americana* Mill) from southern Ethiopia and Guatemala using polymorphic SSR markers. Genetic Resources and Crop Evolution. https:// 10.1007/s10722-023-01831-1

Ruiz-Chután JA, Kalousová M, Maňourová A, Degu HD, Berdúo-Sandoval JE, Villanueva-González C E, Lojka B. 2023. Core collection formation in Guatemalan wild avocado germplasm with phenotypic and SSR data. Agronomy **13**:2385. https://doi.org/10.3390/agronomy13092385

Ruiz-Chután JA, Berdúo-Sandoval JE, Maňourová A, Kalousová M, Villanueva-González CE, Fernández E, Žiarovská J, Sánchez-Pérez A, Lojka B. 2023. Variability analysis of wild Guatemalan avocado germplasm based on agro-morphological traits. Tropical and Subtropical Agroecosystems **26**:52. https://doi.org/10.56369/tsaes.4663

Ruiz-Chután JA, Berdúo-Sandoval JE, Kalousová M, Fernández E, Žiarovská J, Sánchez-Pérez A, Lojka B. (2022). SSRs markers reveal high genetic diversity and limited differentiation among populations of native Guatemalan avocado. Journal of Microbiology, Biotechnology and Food Sciences **12**:e6134. https://doi.org/10.55251/jmbfs.6134

Ruiz-Chután JA, Berdúo-Sandoval JE, Kalousová M, Lojka B, Fernández E, Žiarovská J, Sánchez-Pérez A. 2020. Diversidad genética de materiales nativos de aguacate guatemalteco a través del marcador molecular AFLP. Ciencia, Tecnología y Salud **7**:155–169. https://doi.org/10.36829/63CTS.v7i2.746

CONTENTS

List of Tables

List of Figures

List of Abbreviations

AFLP	amplified fragment length polymorphism
A-NE	accession-to-nearest-entry distance
AMOVA	analysis of molecular variance
ANOVA	analysis of variance
AS	leaf anise smell
BIC	bayesian information criterion
BMPs	best management practices
CAPS	cleaved amplified polymorphic sequence
CC	core collection
CML	color mature leaf
CR	coincidence rate of range
CR	coincidence rate of range
CS	cotyledon surface
CTAB	cetyl trimethyl ammonium bromide
CV	coefficient of variation
CYT	color young twig
CZU	Czech University of Life Sciences Prague
DAPC	discriminant analysis of principal components
DArT	diversity arrays technology
E-NE	entry-to-nearest-entry distance
FAMD	factorial analysis of mixed data
FAO	Food and Agriculture Organization
F_{IS}	inbreeding coefficient
FL	fruit length
FS	fruit shape
FSS	fruit skin surface
F_{ST}	Wright's fixation index
FT	flesh texture
FTZ	Faculty of Tropical Agrisciences;
FW	fruit weight
GD	fene diversity
H’	Shannon diversity index
HCPC	hierarchical clustering on principal components
He	Nei's gene diversity (expected heterozygosity)
Ho	observed heterozygosity
Ht	total gene diversity
Hw	mean gene diversity within populations
HWE	Hardy-Weinberg equilibrium
IPGRI	international plant genetic resources institute
LL	leaf length
LS	leaf shape
LW	leaf width
MCMC	Markov chain Monte Carlo
MD	mean difference percentage
MFSC	mature fruit skin color
MSN	minimum spannin network
MSN	minimum Spanning Network
Na	number of alelles
Ne	effective number of alleles
NGS	next generation sequencing
Nm	gene flow
PAGE	polyacrylamide gel electrophoresis
PCA	principal component analysis
PCR	polymerase chain reaction
PL	pedicel length
PP	petal pubescent
PS	pedicel shape
QQ	quantile-quantile
RAPD	random amplified of polymorphic DNA
RFLP	restriction fragment length polymorphism
SD	standard deviation
SL	sepal length
SNP	single-nucleotide polymorphism
SRAP	sequence-related amplified polymorphism
SS	seed shape
SSR	simple sequence repeats
SW	seed weight
TC	trunk circumference
TS	trunk surface
UPGMA	unweighted pair group method arithmetic
USAC	Universidad de San Carlos de Guatemala
USDA	United States Department of Agriculture
VD	variance difference percentage
VD	variance difference percentage
VNTR	variable number of tandem repeats
VR	variable rate of coefficient of variance
VR	coefficient of variance
WGRC	Wondo Genet Research Centre

Abstract

Avocado (*Persea americana* Mill.) holds immense cultural, economic, and medicinal significance in Guatemala, one of the centers of domestication for the species. However, limited knowledge exists regarding the genetic diversity and population structure of native avocado populations in Guatemala. This study aimed to address this gap by examining the agro-morphological and genetic characteristics of native avocado germplasm providing valuable insights for the selection of genotypes in conservation and breeding programs. Furthermore, the comparison with Ethiopian avocados aims to analyze the influence of introduced Guatemalan germplasm on the formation of cultivated genotypes and to identify opportunities for more efficient utilization of the available germplasm for breeding purposes in Ethiopia. To explore the native germplasm, standardized descriptors for avocado (*Persea* spp.) were employed to evaluate agro-morphological traits. Additionally, molecular characterization was conducted using AFLP and SSR molecular markers. A total of 189 avocado trees, grown from seeds in eight geographical populations, were sampled for analysis. The agro-morphological assessment allowed for the exploration of various characteristics, indicating the presence of Mexican, Guatemalan, and West Indian avocado races in Guatemala. Fruit shape, skin color, flesh texture, and anise odor in the leaves were identified as key traits for grouping trees according to known races and can therefore be used in the characterization of cultivars and native trees whose ancestry is unknown. Furthermore, the molecular characterization using AFLP and SSR markers facilitated a deeper understanding of the genetic diversity and population structure of native avocado populations. The analysis revealed a high degree of genetic diversity (Ho = 0.53, He = 0.83), indicating the presence of a wide range of alleles (24 alleles per locus). However, low differentiation was observed between populations (F_{ST} = 0.018) suggesting a significant rate of migration (Nm = 12.25) and genetic mixing among the analyzed materials. The sampled individuals were classified into three main genetic clusters by the model-based STRUCTURE and discriminant analysis of principal components (DAPC). To preserve the genetic resource, the proposed core collection successfully captured the maximum genetic diversity present in the entire collection, as demonstrated by several indicators. These indicators include the desirable genetic distance, with a high variance difference percentage (VD) of 95.5%. Additionally, the coincidence rate of range (CR) was 92.06%, and the variable rate of coefficient of variance (VR) was 108.71%, all of which exceeded the threshold values of

VD (80%), CR (80%), and VR (100%) required for a good core collection. Comparing the core set to the entire germplasm, both the coefficient of variation and Shannon-Weaver diversity indices increased in the core set. In addition to the Guatemalan context, a fourth study explored the diversity and population structure of avocados in Ethiopia. It identified high genetic diversity and differentiation among populations across different regions. The study highlighted the importance of conserving original alleles and emphasized the significance of Wondo Genet population, which exhibited higher genetic diversity compared to other regions. Overall, these findings shed light on the agro-morphological and genetic diversity of avocado populations in Guatemala. The findings emphasize the importance of germplasm conservation, both for preserving cultural heritage and for future breeding programs aiming at developing avocado cultivars with desirable traits.

Key words: Native avocado, agro-morphological traits, molecular markers, core collection, factorial analysis of mixed data, hierarchical clustering on principal components.

1 Introduction

Avocado (*Persea americana* Mill.) is a subtropical species which belongs to the *Lauraceae* family, one of the oldest known flowering plant families (Renner 1999). Avocado holds significant economic importance as a subtropical/tropical fruit crop worldwide, originating from Central America and Mexico (Schaffer et al. 2012). This species displays remarkable variability, leading to the emergence of distinct ecological races (Galindo-Tovar et al. 2008).

Horticulturalists widely recognize three primary ecological races of avocado: Mexican (*P. americana* var. *drymifolia*), Guatemalan (*P. americana* var. *guatemalensis*), and West Indian (*P. americana* var. *americana*) (Schaffer et al. 2012). These races exhibit distinguishable traits in terms of morphology, horticulture, and physiology (Williams 1977; Bergh & Ellstrand 1986). The Mexican race is considered native to Central Mexico, adapted to colder climates, while the Guatemalan race is primarily found in the mid and high altitudes of Guatemala's mountains and displays some cold tolerance (Schaffer et al. 2012). The West Indian race, native to central and northern South America, was introduced to the West Indies after the Columbian era, showcasing adaptation to warm and humid tropical lowland conditions (Bergh et al. 1973).

Avocado follows a protogynous dichogamy pattern that encourages cross-pollination, and there are no sterility barriers among the three racial types (Schaffer et al. 2012; Stern et al. 2021). Consequently, most commercial avocado cultivars are interracial hybrids (Alcaraz & Hormaza 2007; Schaffer et al. 2012).

The limited availability of information and research regarding avocado, specifically in terms of characterization and accurate identification of native germplasm, hinders the genetic improvement efforts of avocado in Guatemala. Understanding the genetic diversity is crucial for the development of effective strategies concerning germplasm collection, management, conservation, domestication, and enhancement of the species' genetic resources (Salgotra & Chauhan 2023).

The study aimed to evaluate variability in native Guatemalan avocado germplasm through agro-morphological and genetic analysis, offering insights for genotype selection in conservation and breeding programs. The comparison with Ethiopian avocados aims to analyze the influence of introduced Guatemalan germplasm on the formation of cultivated genotypes and to identify opportunities for more efficient utilization of the available germplasm for breeding purposes in Ethiopia.

2 Literature review

2.1 Taxonomy

2.1.1 Classification and distribution of the *Persea* genus

The genus *Persea* (Clus.) Miller, a member of the family Lauraceae, has its origins in woody magnolian ancestors and represents a distinct group that has not given rise to any extant plant taxa (Renner 1999). Alongside the Annonaceae, Magnoliaceae, and Proteaceae, it is among the oldest recorded flowering plant families. Throughout history, species within this family have been utilized for various purposes, such as food, spices, medicine, cosmetics, industrial applications, timber, and ornamentals. The genus can be divided into two subgenera: *Eriodaphne*, primarily found in South America, and subgenus *Persea*, native to Mesoamerica and encompassing edible avocados (Van der Werff 2002). The genus originated in African Gondwanaland, and its ancestral species migrated to Asia, Europe, North America, and South America, likely during the Paleocene era. The convergence of the Americas in the late Neogene reunited the genus, while mountain formation in Central America created new habitats for speciation (Scora & Bergh 1990, 1992).

Subgenus *Persea* includes three recognized species: *P. schiedeana* Nees, *P. pallescens* (Mez) Lorea-Hern, and *P. americana* Mill (Cruz-Maya et al. 2018). *Persea americana* is polymorphic and comprises several taxa, considered botanical varieties or subspecies, known as horticultural races (Wolstenholme & Whiley 1999). The commercial avocados consist of the varieties *P. americana* var. *americana* Mill. (West Indian or Lowland avocado), var. *drymifolia* (Schlecht & Cham.) Blake (Mexican avocado), and var. *guatemalensis* Williams (Guatemalan avocado), are all recognized as geographical ecotypes (Wolstenholme & Whiley 1999; Lavi et al. 2003). Other varieties, including var. *nubigena* (Williams) Kopp, var. *steyermarkii* Allen, var. *zentmyerii* Schieber and Bergh, and var. *tolimanensis* Zentmyer and Schieber, are considered distinct varieties contributing to the ancestry of var. *guatemalensis* (Schieber & Bergh 1987). Another wild botanical variety is *floccosa* Mez.

As germplasm exploration continues, more taxa are expected to be described. An example is the endemic form of *P. americana* var. *americana* found in Costa Rica, where the typical vars. *drymifolia* and *guatemalensis* are nearly absent. This endemic variety, intermediate between Guatemalan and West Indian avocados, should be recognized as *P. americana* var. *costaricensis*, according to Ben-Ya'acov et al. (1995). An avocado with a

round, hard shell, endemic to Monte Verde, Costa Rica, has been proposed as *P. americana* var. *tilaranensis* Scora. Another avocado variant, characterized by a higher anethole content overshadowing the prominent estragole found in var. *drymifolia*, is known as "aguacate de anis" and may be recognized as a chemovar of var. *drymifolia* (Schieber & Bergh 1987). In Ecuador, there are unique endemic avocado types whose relationship to known varieties from Mexico and Central America remains uncertain. Further investigation is needed to clarify this connection (Ben-Ya'acov et al. 1992). Additionally, Ben-Ya'acov & Barrientos (2003) have recently reported the discovery of a pubescent-leaved primitive Guatemalan avocado from Chiapas (southern Mexico), suggesting that it may represent a new *Persea* species.

2.2 Origin and history

2.2.1 Theories on the center of origin for avocado species

Regarding the center of origin of the species, there have been several theories since Popenoe (1935) proposed the region from Mexico to the north of South America as the center of origin. Later Williams (1977) mentioned that the center of origin of the avocado is located in the highlands of central and central-eastern Mexico, as well as in the highlands of Guatemala. Storey et al. (1986) propose the Chiapas (southern Mexico) - Guatemala - Honduras area as the center of origin based on identifying possible wild avocados. Years later, Bergh (1992) mentions the south-central region of Mexico and Guatemala, where the species may have originated. The origin of the avocado has been related to areas with large populations or where individuals have been identified as wild (Storey et al. 1986; Bergh 1992).

However, Galindo-Tovar & Arzate-Fernández (2010a) mention that neither paleoclimatic data nor the relationship of the location of the fossils with the trajectory of biotic replacement and responses to environmental catastrophes has been considered, which, according to these authors, would place the center of origin of the species in a much more northerly region. In the process of reconstructing the history of the avocado, Scora & Bergh (1992) mention that the genus *Persea* was already in North America 56-35 million years ago, in agreement with the subtropical climate conditions of the Sierra Nevada (California) described for that period (Millar 1996) and with the avocado fossils reported by Schröeder (1968).

From this scenario, it is possible that at the time of the formation of the Sierra Nevada in the late Cenozoic (Liu & Shen 1998; Wakabayashi & Sawyer 2001), the first modern avocados originated from an adaptation process in this area and not further south in the Chiapas - Guatemala - Honduras area, where favorable climatic conditions did not exist during the Miocene - Pliocene as described by Galindo-Tovar and Arzate-Fernández (2010). These authors also mention that during the last glaciation in the Sierra Nevada, the climate changed to dry and cold, so the avocado trees moved southward and became extinct in the area of origin.

2.2.2 The three avocado races

The avocado, originally from Central America and Mexico, has been an integral part of the diet for over 9,000 years (Chen et al. 2009). Among *P. americana*, three distinct ecological races can be identified: Mexican, Guatemalan, and West Indian (or Antillean). Each race exhibits unique traits such as leaf morphology, fruit characteristics, and flowering period, which are summarized in **Table 1** (Paull & Duarte 2011).

Sterility barriers do not exist among the three races or any taxonomic category within *P. americana*, allowing for easy hybridization when trees of different races grow in close proximity. Consequently, commercial avocado cultivars are primarily interracial hybrids resulting from chance seedlings and varying degrees of hybridization (Alcaraz & Hormaza 2007).

These avocado varieties possess distinct morphological, ecological, and molecular characteristics, as described by some studies (Bergh & Ellstrand 1986; Knight & Campbell 1999; Schnell et al. 2003; Galindo-Tovar et al. 2008; Chanderbali et al. 2013):

a. *Mexican race* – Delicate skin, large and often loosely attached seed, and smaller fruit size than commercial preference. Suited for high elevations, exhibits the highest cold resistance, and offers a high-oil content with a rich nutty flavor. High-quality pure Mexican-race avocados are rare, they have contributed genes associated with early maturity and cold tolerance, among others.

b. *Guatemalan race* – Thicker skin, smaller and firmly embedded seed, with some variations in skin thickness. Notably, Guatemalan-race avocados excel in horticultural quality. They have a longer time to maturity, enabling a later harvesting season and facilitating year-round commercial fruit picking when crossed with earlier-maturing races. Well adapted to high elevations and capable

of withstanding cold weather. Guatemalan-race avocados are known to possess highly valuable horticultural genes, making them a dominant presence in the germplasm of subtropical avocado cultivars worldwide.

c. *West Indian race* – Greater tolerance to salt and chlorosis. Thrives in lowland tropical regions. Hybrids with Guatemalans-race avocados bridge the two harvesting seasons and combine Guatemalan-race quality with the West Indian adaptation to tropical climates. West Indian avocados are famous for their relatively low oil content but high sugar content, resulting in a distinctly less "nutty" flavor compared to subtropical cultivars.

Table 1. Comparison of the three horticultural races of avocado based on tree and fruit traits.

		Race		
	Trait	**Guatemalan (G)**	**Mexican (M)**	**West Indian (WI)**
Tree	Climate	Subtropical	Semitropical	Tropical
	Cold tolerance	Intermediate	Most	Least
	Salinity tolerance	Intermediate	Least	Most
	Leaf anise	Absent	Present	Absent
	Young leaf color	Green with red tinge	Green	Pale yellow
	Mature leaf color	Dark green	Dark green	Pale green
Fruit	Blooming season	March to April	January to February	February to March
	Bloom to fruit maturity	10 – 18 months	5 – 7 months	6 – 8 months
	Size	Small to large	Tiny to medium	Medium to very large
	Shape	Mostly round	Mostly elongate	Variable
	Color	Green	Often dark	Green or reddish
	Skin thickness	Thick	Very thin	Medium
	Skin surface	Rough	Waxy bloom	Shiny
	Skin peelability	Rigid	Membranous	Leathery
	Seed size	Small	Large	Variable
	Seed cavity	Tigh	Loose	Variable
	Seed surface	Smooth	Smooth	Rough
	Oil content	High	Highest	Low
	Pulp flavor	Rich	Anise-like, rich	Sweeter, milder

Based on the information from Paull & Duarte (2011),Galindo-Tovar & Arzate-Fernández (2010a), Chanderbali et al. (2013) and Ayala-Silva & Ledesma (2014)

2.3 Domestication process

Domestication involves the intentional cultivation and modification of plant species by humans through selective breeding, genetic manipulation, and cultivation techniques to enhance their utility (Mueller & Flachs 2022). This practice commenced as humans identified wild plants suitable for consumption or crafting materials, collected their seeds, and purposefully cultivated them. As the practice evolved, people started selecting seeds from cultivated plants with desirable traits such as taste or size to grow subsequent crops in the following years (Barrera-Redondo et al. 2020).

2.3.1 Ancient utilization of avocado by the Mayan culture

Determining the point of domestication of a plant is a challenging task. According to Pickersgill (2007), the domestication process of a plant starts with managing wild plants, then selecting plants for cultivation, and ends with human selection leading to morphological changes that differentiate the domesticated plant from its wild parent. In Mesoamerica, trees have been an integral part of the subsistence strategies of various cultures, and the avocado has played a significant role in the history of several Mesoamerican civilizations. The Mayan culture, for instance, used trees such as avocado, plum (*Prunus americana* Marsh.), and sapote (*Pouteria sapota* L.) at least 3,400 years ago (Colunga & Zizumboo 2004).

The avocado has been highly valued by the different human groups that inhabited Mesoamerica, and several pieces of evidence exist regarding its domestication process. However, it is necessary to integrate this information and clarify the domestication process through an interdisciplinary approach that includes historical and paleoecological data (Solares et al. 2023). Storey et al. (1986) reported that avocado consumption dates back to 6,500 years B.C. in Coaxcatlán, Puebla, and fruit selection had probably already begun. However, the selection results were not visible until much later, up to 900 years B.C., due to the avocado's long juvenile state and life cycle (Landon 2009).

2.3.2 Avocado domestication and the influence of Mayan culture

According to Gama & Gomez (1992), selecting avocado seeds based on their size is reasonable due to the wide variety of this tree, and domestication may have occurred independently and for various purposes determined by people and climatic conditions.

From this perspective, the first phase of avocado domestication possibly happened when people started collecting the best fruits in the forest, selecting trees with desirable traits, and harvesting them *in situ* at the forest agroecosystem level, leading to the first morphological changes in the tree (Galindo-Tovar et al. 2007, 2013).

The second phase of avocado domestication began when climate changes prompted human groups to sow the seeds of the best fruits near their homes to conserve the avocado when it became scarce (Macneish 1964). This unintentional modification of the biophysical environment led to a closer human-plant interaction. The third phase involved intentional cultivation and selection, whereby the most valuable trees were taken to the most favorable habitats, and the adaptation of the tree for specific uses was increased (Galindo-Tovar et al. 2013).

Colunga and Zizumbo (2004) reported that already domesticated avocados were brought to the Mayan lowlands of Yucatan, Belize, and Guatemala at least 3,400 years ago by the first human groups from central or south Mexico (Chiapas). Other inhabitants also established an important avocado domestication center through selection, indicating a second domestication continued by the Yucatecan Lowland Maya culture, which has been recognized as one of the main avocado domesticators (Gama & Gomez 1992).

The West Indian avocado is characterized by its adaptation to tropical conditions, altitudes of less than 1,000 m above sea level, and its resistance to salinity and chlorosis. It is suggested that these characteristics were acquired to adapt to the climate and calcareous soils characteristic of the Yucatan Plain. An essential part of the domestication process of this physiological race of avocado was carried out in Mayan backyard orchards (Galindo-Tovar & Arzate-Fernández 2010b).

Thus, the domestication of the avocado occurred in three phases, with the third phase involving intentional cultivation and selection. The Mayan lowlands of Yucatan, Belize, and Guatemala were an essential center for avocado domestication, and the West Indian avocado was adapted to the climate and soils of this region. Molecular approaches estimated divergence times which varied from ~40,000 years between the Lowland and Guatemalan groups to > 1.0 million years between the Mexican and the two other groups (Solares et al. 2023). The early divergence among groups may have been driven in part by ecological differences among regions, especially given the evidence that native germplasm in Mexico is genetically subdivided by elevation (Chen et al. 2009). Further interdisciplinary research that integrates historical and paleoecological data is needed to clarify the domestication process of the avocado (Fedick et al. 2023).

2.4 Avocado botanical description

The avocado tree is a tall, upright plant that can grow up to 9 to 18 m, with a bole diameter of 30 to 60 cm. The leaves come in various shapes, including lanceolate, elliptic, oval, ovate, or obovate, and may be alternate, dark green, glossy on the upper surface, and whitish on the underside (Morton 1987; Bost et al. 2013). The leaves vary in length, ranging from 7.5 to 40.0 cm long. The fruit is pear-shaped, often necked, oval, or nearly round, measuring 7.5 to 33.0 cm long and up to 15 cm wide. The fruit's skin can be yellow-green, deep-green, reddish-purple, or almost black, and may be speckled with tiny yellow dots, smooth or pebbled, glossy or dull, thin or leathery, up to 6 mm thick, pliable or granular and brittle (Figure 1). The flesh of the avocado is generally entirely pale to rich-yellow in color, but in some fruits, there is a thin layer of soft, bright-green flesh immediately beneath the skin. The avocado fruit has a single seed enclosed in two brown, thin, papery seed coats that often adhere to the flesh cavity. The seed may be oblate, round, conical, or ovoid in shape, hard and heavy, ivory in color, and 5.0 to 6.4 cm long (Morton 1987).

Figure 1. Native Guatemalan avocado tree (own source).

2.5 Biology

2.5.1 Flowering

Avocado trees typically exhibit a remarkable abundance of flowers during the flowering period, surpassing a million in number in adult trees. However, many of these flowers ultimately fall without developing into fruits, a typical feature of evolutionary old plant taxa (Davenport 1986). Avocado flowers possess both male and female reproductive organs, classifying them as bisexual. Notably, they demonstrate protogynous dichogamy, a unique characteristic where each bisexual flower opens twice with an intermediate closure (Stern et al. 2021). The avocado flower has both functional male and female organs in the same flower but opens and closes twice over a two-day period. The first day it functions as a female flower with a receptive stigma, and the next day it functions as a male flower with the stigma no longer receptive and dehisced anthers (Alcaraz & Hormaza 2021).

Moreover, one can distinguishing between so called A-type flowers, which function as females in the forenoon and as males in the afternoon, while B-type flowers behave as males in the forenoon and as females in the afternoon (Nirody 1922). This restricts self-pollination and close-pollination, promoting reciprocal cross-pollination between the two groups (Ashman et al. 2004; Borrone et al. 2008; Schnell et al. 2009). Although some attempts at artificial cross-pollination between A and B types have yielded negative results (Stout 1923) it is possible to do it following appropriate management approaches (Alcaraz & Hormaza 2014). Hence, avocado flowering behavior is a complex mechanism that restricts or even prevents effective self-pollination and close-pollination. Understanding the pollination process can contribute to improving fruit set and increasing avocado yields, considering that less than 1% of avocado flowers develop into fruits (Davenport 1986).

A study conducted in southern Spain focused on the transition from pollination to fertilization, shedding light on the reasons behind premature flower abscission and the persistence of others on the tree (Alcaraz et al. 2013; Alcaraz & Hormaza 2019). The flower undergoes distinct stages in its reproductive process. Initially, the female stage flower (**Figure 2**A) opens and remains open for a duration of two or three hours. After this period, it closes and remains closed throughout the remainder of the day and night. The following day, the flower opens once again; however, during this male stage (**Figure**

2B), the stigma is no longer receptive to pollen grains. Instead, the flower releases pollen and subsequently closes (Nirody 1922; Stout 1923; Alcaraz & Hormaza 2011).

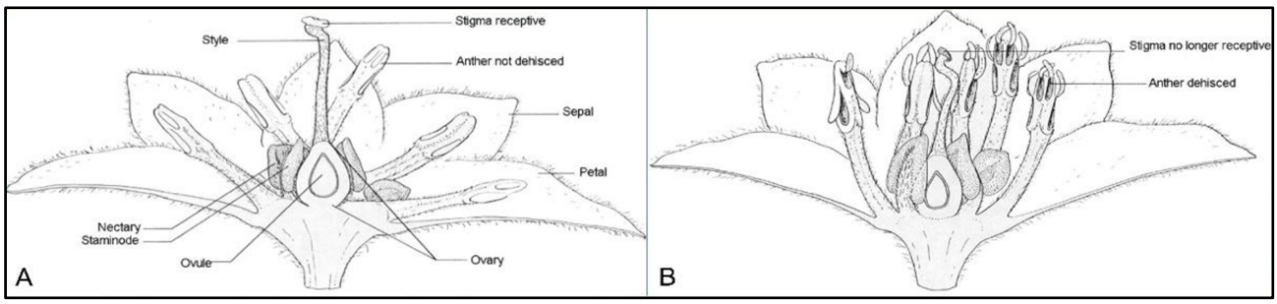


Figure 2. Bisexual flowers of the Avocado; A) Female stage; B) Male stage. McGregor (1976).

Pollination is a critical stage in ensuring fertilization and seed production, playing a vital role in the sexual reproduction of plants. However, many flowers do not receive an adequate amount of pollen when they close in the female stage. Some studies indicate that the number of pollen grains arriving on the stigma exceeds the amount required for ovule fertilization, yet insufficient pollination can result in low yields (Burd 1994; Larson & Barret 2000; Ashman et al. 2004). In the context of avocado production, studies have highlighted the deficient transfer of pollen to the stigma of female flowers as the primary limiting factor for fruit set (Alcaraz & Hormaza 2009). The complete description of the reproductive structure of *P. americana* is shown in the **Figure 3**.

2.5.2 Pollination

Avocado exhibits three modes of pollination. The first is cross-pollination, which occurs during warm weather conditions. Pollen is transferred from male flowers of A-type to female flowers of B-type, and vice versa (Ish-Am & Eisikowitch 1993). The effectiveness of cross-pollination depends on factors such as the distance between pollen donors and the pollinated trees, as well as the duration of overlap between male and female-stage flowers (Alcaraz & Hormaza 2014). The second mode is close pollination, which takes place when neighboring flowers on the same plant cross-pollinate during the overlapping period of male and female stage flowers. Pollen from male flowers lands on the stigmas of female flowers (Kämper et al. 2021). Self-pollination occurs when pollen grains reach the stigma within the same flower (Stout 1923; Ashman et al. 2004).

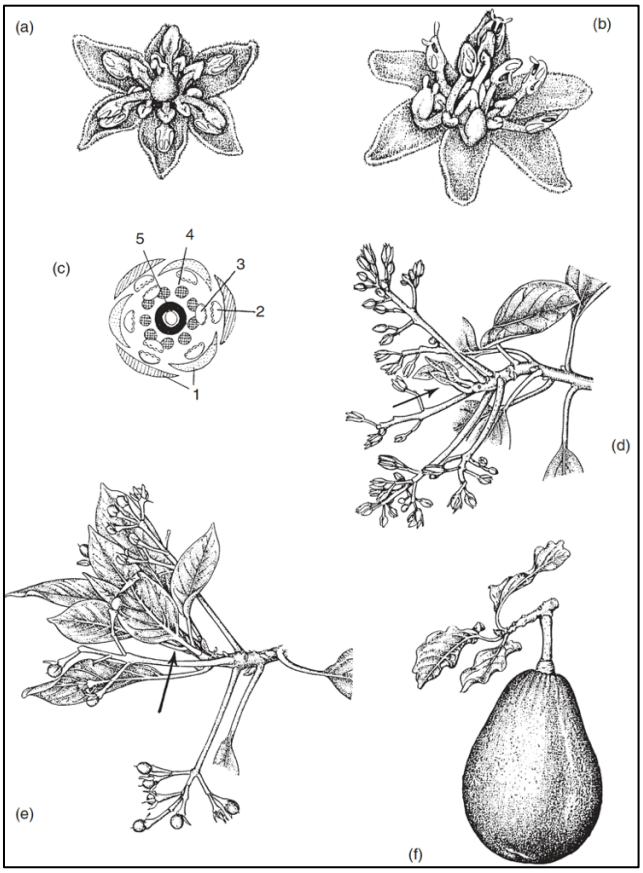


Figure 3. Reproductive structures of avocado (*Persea americana* Mill.). (a) Flower at female stage. (b) Flower at male stage (pollen releasing). (c) Floral diagram where 1 = sepal (outer) and petal (inner) or tepals, 2 = stamen of mid-whorl; 3 = stamen of inner whorl; 4 = glandular staminode; 5 = pistil. (d) Inflorescence with a terminal vegetative bud (arrow). (e) Fruit set with the renewal spring shoot growth (arrow). (f) An avocado fruit. (Drawings by P. Fawcett in Tomlinson (1980)).

2.5.3 Optimal growth conditions

For optimal fruit production, avocado plants require warm temperatures and sunny, windless locations. In terms of cold tolerance, Mexican avocados are the hardiest, able to withstand temperatures as low as -8°C. On the other hand, West Indian varieties cannot survive below freezing temperatures, and Guatemalan trees can endure temperatures down to -4°C. It is important to protect avocado trees from strong winds and freezing temperatures (Lundman 2018).

Studies have shown that avocados thrive in temperatures between 20 to 25°C, with nighttime temperatures above 10°C and daytime temperatures ranging from 20°C to 30°C during the flowering stage to ensure a good fruit set (Lahav & Trochoulias 1982).

However, avocado trees can tolerate higher temperatures if they receive sufficient water. Avocado trees exhibit a preference for regions with an annual rainfall ranging between 1000 to 1500 millimeters. This well-distributed precipitation pattern is particularly vital during critical growth stages such as flowering and fruit set. However, in areas where rainfall is insufficient or irregular, supplementary irrigation becomes indispensable. On average, an avocado tree requires around 800 to 1200 liters of water per week, depending on factors such as tree age, soil type, and climate conditions (Carr 2013). Properly managing water resources, whether from natural rainfall or controlled irrigation, plays a pivotal role in sustaining healthy avocado orchards and ensuring consistent fruit production of high quality (Corrales-García et al. 2019; Cervantes-Paz & Yahia 2021).

Avocado flowers are highly sensitive to environmental conditions. Observations from various regions such as California, Florida, Australia, Israel, and France have demonstrated that under optimal climatic conditions, flower openings occur uniformly and predictably, while cloudy days can disrupt the regular pattern of flower openings. In low-temperature conditions, both male and female flower openings in A-type cultivars may be delayed to the extent that their behavior resembles that of B-type cultivars (van Rooyen & Bower 2006; Arpaia et al. 2018).

2.5.4 Environmental and social footprint of avocado cultivation

Recent findings shed light on the adverse effects of expanding avocado cultivation in Michoacán, Mexico, the global epicenter of avocado production. Deforestation and forest fragmentation in this region have significantly impacted biodiversity, soil quality, and hydrological systems (Denvir et al. 2022).

Regarding the water footprint, a recent study reveals that avocado cultivation necessitates substantial water usage, potentially exacerbating water scarcity issues in regions where avocados are cultivated (Sommaruga & Eldridge 2021).

It is crucial to recognize the nuanced socioeconomic impacts associated with avocado production. While the industry has introduced economic benefits, such as increased employment opportunities and poverty reduction, the presence of inequity in the region hampers the positive socioeconomic transformations (Denvir et al. 2022).

2.6 Propagation methods

2.6.1 Propagation through sowing seeds

For successful seed germination, it is crucial to use seeds obtained from ripe fruits rather than immature or fallen fruits. Once the seeds are selected, they should be disinfected by subjecting them to a temperature of 50°C and then immediately transferred to cold water to prevent contamination by fungi or other organisms (Platt 1976). Individual seeds are then sown at intervals of 1 to 5 cm between lines, with the wider and flatter basal part of the seed placed downwards. The germination process typically occurs between 41 and 62 days after sowing (Lozi et al. 2018). It is important to avoid exposing the seeds to direct sunlight for extended periods to prevent dehydration. During the night, the seeds can be uncovered. Once the seedlings have two fully developed leaves, they are ready for transplantation into the soil (Storey et al. 1986).

To transplant the seedlings, a 20 cm hole should be prepared in the ground, and the new seedling can be placed in it. At this stage, it is crucial not to fertilize the soil since the seed contains sufficient nutrients for seedling growth. The newly transplanted seedling should be watered immediately after transplantation and again the following day to maintain moisture and prevent drying out. It is worth noting that this method is not recommended for large-scale commercial plantations due to the longer fruiting time (Adjei et al. 2011).

2.6.2 Propagation by grafting

Grafting is a well-established technique in horticulture that utilizes the natural wound healing mechanisms of plants to merge two distinct genotypes into a single plant. This process involves joining different plant varieties together, allowing them to grow as one entity (Loupit et al. 2023). The advantages of using the grafting method are as follows:
i. Shortening the juvenile period to enter production at an early age; ii. Reducing the size of the plant to better control pruning (low pruning); iii. Transfering resistance to pests and diseases by creating patterns. Among the drawbacks, incompatibility issues can emerge at various grafting stages. Furthermore, there is a heightened cost involved, including the use of double seeds, requiring additional greenhouse space for transplanting double seedlings, the need for certified seeds, employing skilled labor or offering extra training, and selecting indeterminate varieties over specific ones because of the shorter production time necessary to recover the investment (Reyes-Herrera et al. 2020; Cañas-Gutiérrez et

al. 2022). Although seedling rootstocks have been tradicionally used in avocado production, clonal rootstock with tolerance to soil fungi are increasingly being used worldwide.

The selection of an appropriate pattern is essential for successful grafting, and it should meet specific requirements (Reyes-Herrera et al. 2020). The prevalent technique for grafting avocados involves the cleft graft method. This traditional field grafting method includes creating a vertical split in the center of the rootstock and inserting one or two branches (scions) with two or three buds into the rootstock's cambium layer (Ahsan et al. 2019). Ideally, the pattern should originate from a healthy tree, measuring around 60cm in height and 1cm in width. To prepare the pattern, the lower leaves (within 35 cm of the ground) are carefully removed, leaving the stem bare. A beveled cut is then made, and a graft from a mature and fruitful tree is inserted into the bevel cut, typically around 10cm in length. To secure the graft, it is covered with plastic and further protected by enclosing it within a plastic bag to prevent water damage during watering (Haberman et al. 2020).

The saplings are typically ready for transplantation within four to six months after grafting. The spacing for planting frames is determined based on factors such as soil type, topography, variety or cultivar (due to growth habit), and prevailing environmental conditions. When transplanting the trees, the holes in the soil should generally have a depth of 60 centimeters (Mahbou et al. 2022). In soils with low fertility, it is recommended to add two kg of organic matter or substrate to the holes one week prior to transplantation, ensuring it is free from pathogens to avoid plant contamination. The plant should be removed from the bag and placed centrally in the 60-centimeter hole, with the hole then filled with soil. It is important to tamp the soil and ensure that the plant is level with the soil surface (Galang 1940; Mahbou et al. 2022).

In general, the distances between plants and rows range from 7m x 9m to 10m x 12m. A common spacing used is 10 meters between plants and 10 meters between rows, although in modern plantings spacing is often 6m x 6m, 6m x 5m or even lower. Various planting systems can be employed. For example, the square system can be 8 x 8 meters with 156 plants per hectare, 9 x 9 meters with 123 plants per hectare, or 10 x 10 meters with 100 plants per hectare. The staggered system can be 8 x 8 meters with 180 plants, 9 x 9 meters with 142 plants, or 10 x 10 meters with 115 plants (Razeto et al. 1992).

2.7 Best management practices for avocado production

Avocado, a versatile and nutritious fruit, has gained global popularity, leading to a surge in demand and production. However, avocado cultivation faces challenges that necessitate the implementation of best management practices (BMPs) for sustainable outcomes. A comprehensive guide from FAO underscores the significance of post-harvest operations in maintaining fresh fruit quality. This guide encompasses various aspects, including harvesting, handling, processing, packaging, storage, and marketing, providing analytical methods for assessing avocado composition and quality (FAO 2004).

Wangithi et al. (2022) indicate that integrated pest and pollinator management practices are pivotal for enhancing avocado production. Understanding ecological interactions among pests, pollinators, and the environment is crucial for developing effective management strategies, as highlighted in the study. Additionally, studies evaluating post-harvest treatments, such as hot water immersion and ethylene exposure, have demonstrated their potential in enhancing avocado quality. These treatments offer practical solutions to improve fruit quality and extend shelf life (Bill et al. 2014).

In California (Gustafson et al. 1979) and Mediterranean subtropical region (Cárceles Rodríguez et al. 2023), water management practices have been tailored for avocado orchards, emphasizing the importance of recognizing microsite variability within orchards. This understanding informs irrigation management strategies, ensuring optimal water usage. Furthermore, a comprehensive review addresses the human-environmental issues associated with the rapid expansion of avocado production in Michoacán, México, and increasingly in other subtropical regions worldwide. The review advocates for management and policy actions throughout the supply chain, aiming to establish a sustainable production framework for this essential commodity (Denvir et al. 2022).

2.8 World avocado production

2.8.1 Production trends and statistics

Over the past decade, global avocado production has witnessed a remarkable increase of over 30% (FAOstat 2021). The surge in avocado cultivation can be attributed to the global increase in consumption, notably in the USA. According to the USDA, per capita avocado consumption in the USA has skyrocketed by 443% over the past two decades, soaring from 1.6 pounds in 1995 to a peak of 7.1 pounds in 2015 (Darnton & Rickenbrode, 2017). This remarkable growth is driven by the escalating demand for both fresh and processed

avocado products, heightened consumer awareness of the nutritional advantages offered by avocados, and the extensive use of avocados in the cosmetic industry. In 2001, global avocado production stood at 2.8 million tons and surged to 8.69 million tons in 2021, reflecting an annual growth rate of 5% during the review period (FAOstat 2021) (**Figure 4**). Between 2014 and 2021, there was a notable increase of 7.6% in global avocado production, reaching a peak volume of 8.69 million tons in 2021. Concurrently, the total area devoted to avocado cultivation expanded in correlation with production growth, rising from 324,826 hectares in 2001 to 858,152 hectares in 2021, with an annual growth rate of 4.42% (FAOstat 2021). However, limited arable land for agriculture remains a global constraint on production. In 2018, approximately 35% of the total global avocado production, on average, was allocated for the export market, distinguishing avocados from other tropical fruits in terms of international trade (Sibuladi 2020).

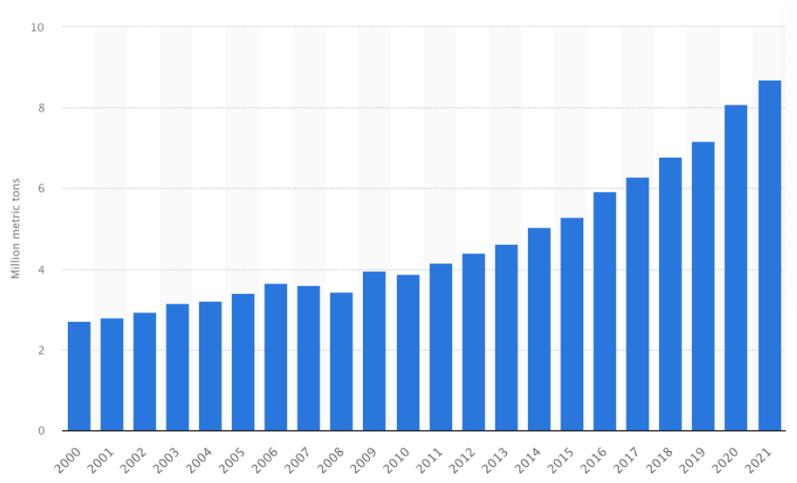


Figure 4. Avocado production worldwide from 2000 to 2021 (FAO 2021).

2.8.2 Leading avocado-producing countries

Avocado has gained increasing attention and market share on an international scale. Around 80% of the avocado cultivars grown, consumed, and traded worldwide encompass major varieties such as Hass and Fuerte (Rincon-Patino et al. 2018). Based on the statistical results of the Food and Agriculture Organization of the United Nations, the avocado sector has quickly grown, with a worldwide planting area of around 0.8 million hectares and an annual production of 8.8 million tons (FAOstat 2023). The leading avocado-producing countries globally included Mexico (34%), Dominican Republic (11%), Peru (8%), and Indonesia (6%), among others (Cruz-López et al. 2022; Nyakang'i

et al. 2023). Mexico, continuing to hold the top position, remaining the largest avocado producer globally, contributing approximately one-third of the total global production. In recent years, other countries have also entered the avocado industry, recognizing its status as a superfood, and have begun exporting to international markets. A notable example is China, which has experienced remarkable growth in avocado production. From having no commercial avocado production in 1991, China has established 20,266 hectares of avocado plantations and produced 124,110 tons of avocados in 2017 (Sibuladi 2020). As of 2022, Guatemala held the 13th position in global avocado production, contributing 1.71% of the total worldwide output. Over the last five years, the export volume has shown a consistent growth trend, reaching a current value of 10.8 million kg (**Figure 5**), equivalent to a value of 12.2 USD million (**Figure 6**) (FAOstat 2022).

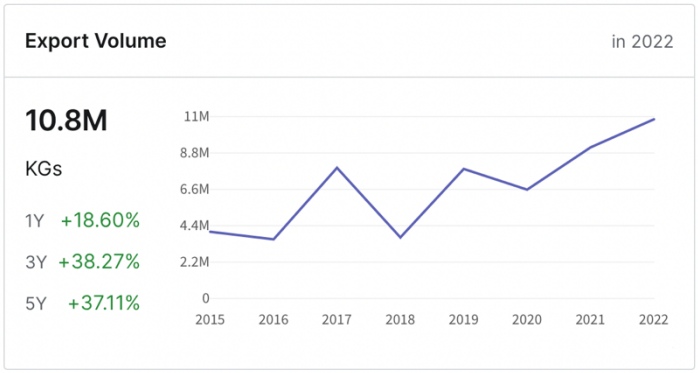


Figure 5. Volume of avocado exported from Guatemala (FAO 2022).

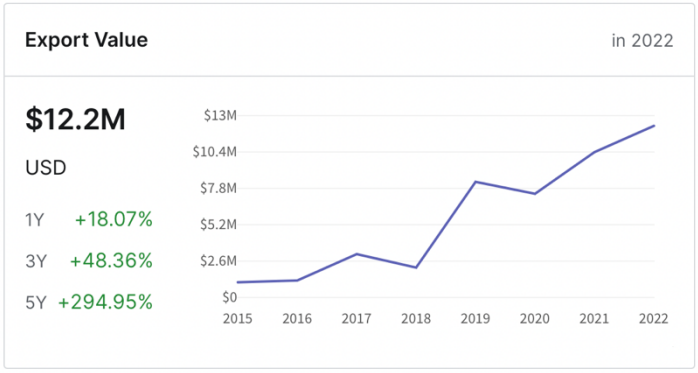


Figure 6. Value in millions of dollars generated by the export of avocado in Guatemala (FAO 2022).

2.9 Chemical composition and nutritional value

Avocado is commonly referred to as the "butter fruit" due to its high fat content. As the fruit ripens, the oil content in avocado increases (Ozdemir & Topuz 2004). Monounsaturated fatty acids, with oleic acid being prominent, are the primary types of fatty acids found in avocado. Although less abundant, avocado also contains linoleic acid (polyunsaturated) and palmitic acid (saturated) (Villa-Rodríguez et al. 2011). The high concentration of monounsaturated fatty acids in avocado has been associated with its various health benefits, making lipids one of the extensively studied chemical components of avocado.

Avocado stands out among fruits for its relatively high protein content, with levels around 2% compared to the average protein content of other fruits at approximately 1% (Paull & Duarte 2011). Additionally, avocado is a rich source of vitamins, particularly vitamins E and C, pigments such as anthocyanins, chlorophylls, and carotenoids (Gross et al. 1973; Ashton et al. 2006), sterols, phenolic compounds (Golukcu & Ozdemir 2010), and seven-carbon sugars and their related alcohols, such as D-mannoheptulose and perseitol (Meyer & Terry 2010) (**Table 2**). The comprehensive analysis of these metabolites has been made possible by advancements in metabolomics and analytical techniques, enabling the identification and quantification of complex compounds in avocado.

Table 2. Nutritional content of avocado.

Nutrient	Amount per 150g	Nutrient	Amount per 150g
Calories	240	Iron	0.83mg
Total Fat	22g	Potassium	762mg
Saturated Fat	3g	Vitamin A	98µg
Trans Fat	0g	Vitamin C	13mg
Cholesterol	0mg	Vitamin E	13mg
Sodium	10mg	Vitamin K	32µg
Total Carbohydrates	12g	Vitamin B6 (Pyridoxine)	0.44mg

Dietary Fiber	10g	Folate (Vitamin B9)	132mg
Sugars	0.2g	Magnesium	43mg
Protein	3g	Phosphorus	81mg
Vitamin D	0.75µg	Zinc	1mg
Calcium	18mg	Copper	0.26mg

Based on the information from Dreher et al. (2021) and Ford & Liu (2020)

The wide range of nutritive and non-nutritive components found in avocado contribute to its organoleptic properties and have the potential to enhance human health through their health-promoting effects. Extensive research has established a link between avocado consumption and various health benefits, including the maintenance of normal serum cholesterol levels, weight management, diabetes control, and cancer prevention (Ding et al. 2007; Devalaraja et al. 2011; Dreher & Davenport 2013). These effects are primarily attributed to the presence of fatty acids, dietary fiber, D-mannoheptulose and perseitol, potassium, magnesium, vitamins C, E, K, and B group, carotenoids, phenolics, phytosterols, and terpenoids in avocado (Dreher & Davenport 2013).

While avocado can bring positive health effects, it is important to note that no single food can provide all the necessary nutrients and bioactive compounds for optimal nutrition. A well-rounded diet should include a variety of foods from different groups such as fruits, vegetables, legumes and potatoes, fish, and meat to ensure comprehensive nutrition. Additionally, consuming a combination of different foods can enhance the bioavailability and absorption of specific nutrients and bioactive compounds. For example, studies have shown that consuming carotenoid-rich fruits or vegetables in conjunction with avocado or avocado oil can significantly increase the absorption of carotenoids, thereby augmenting their health effects (Unlu et al. 2005).

2.10 Medicinal properties

2.10.1 Use of avocado in traditional medicine

The avocado tree and its fruits have a long history of being used as herbal medicine in regions where it grows naturally. Avocado leaf tea has been traditionally employed to alleviate symptoms of diarrhea, bloating, and gas. It is also believed to assist in coughs and gout by eliminating uric acid from the body. The tea is reputed to possess liver-cleansing properties and can help lower blood pressure (Oboh et al. 2016). In Mexico,

avocado leaf herbal teas have been used for centuries to treat menstrual disorders and as a contraceptive, as they are believed to induce menstruation. Laboratory tests have shown that extracts from avocado leaves efficiently inhibit herpes simplex virus types I and II, which cause cold sores and genital herpes (Wiradona 2017).

Avocado seeds are known for their antibacterial and antifungal properties and have traditionally been utilized to treat diarrhea and dysentery. The peel of the fruit is occasionally employed to remedy intestinal worms, while the pulp is believed to have aphrodisiac effects (Bangar et al. 2022). Animal studies have demonstrated the blood pressure-lowering, antiviral, and anti-inflammatory benefits of leaf extracts, but further research is needed to systematically examine the medicinal applications of leaves, bark, and seeds using human subjects (Kulkarni et al. 2010).

In a small-scale scientific study involving 16 men with ages ranging from 27 to 72, different quantities of avocado were administered, ranging from 12 to 112 fruits per day. Interestingly, half of the participants experienced a notable reduction in cholesterol levels, while none of the remaining participants showed an increase. This suggests that avocados could be a favorable choice for individuals with elevated levels of cholesterol or triglycerides (Wang et al. 2020).

2.10.2 Treatment for atherosclerosis, angina pectoris, and Alzheimer’s disease

The avocado fruit contains alpha-carotenes, which possess antioxidant properties that can potentially safeguard against the oxidation of LDL-cholesterol, also known as the "bad" cholesterol, thereby reducing the risk of atherosclerosis (Wang et al. 2020). Moreover, consuming avocados may provide benefits for individuals with angina pectoris related to atherosclerosis. It is worth noting that antioxidant levels in the bloodstream might play a role in the progression of Alzheimer's disease. Research indicates that individuals with Alzheimer's disease tend to exhibit significantly lower levels of alpha-carotenes in their blood compared to healthy individuals (Ashton et al. 2006; Wang et al. 2020).

2.10.3 Remedy for digestion and blood sugar balance

Due to its alkaline properties and the protective effects of its fat on the mucous membranes, the avocado fruit is believed to be beneficial for individuals with ulcers or gastritis. Research has shown that consuming avocados can aid diabetics in managing

their blood sugar levels effectively, indicating the potential helpfulness of the fruit for individuals with diabetes (Dabas et al. 2013).

2.11 Exploring avocado germplasm diversity: Guatemala and Ethiopia

Avocados have been known in Guatemala since ancient times and have a rich diversity of cultivated and native landraces. On the other hand, avocados were introduced to Ethiopia in 1938 by a missionary who planted the avocado seed in Wondo Genet (Kamaraj et al. 2020). These avocados are still there and provide fruit to the community. Since then, avocado has been introduced to the food culture of Ethiopia.

To improve avocados, the use of locally adapted landraces is a crucial step. Information on the genetic diversity of landraces is the primary (Savolainen et al. 2013; Tiffin & Ross-Ibarra 2014) step for developing appropriate strategies to maximize productivity. The Guatemalan avocado landraces have spread to form the genetic base for local avocados elsewhere (Galindo-Tovar et al. 2008). There is high genetic diversity between avocado landraces in Guatemala due to the large geographical area, the varied climatic conditions, and the occurrence of many ecological types (Alcaraz & Hormaza 2007). The presence of open pollination and cross-compatibility among the races of avocados has facilitated natural diversity by enabling the exchange of genetic material, fostering the development of a wide range of traits and adaptations. (Reyes-Herrera et al. 2020).

Since its introduction to Ethiopia, avocado became a popular tree crop, cultivated in Southern Ethiopia in a region well known for coffee cultivation. As a result, avocado trees become the perfect shade tree for the coffee plantation (Asfaw & Ågren 2007). In addition to its use as a shed plant, avocados have a high market their fruit and oil production (Tesfaye et al. 2022). However, there is little information about the genetic diversity of avocados in Ethiopia. By comparing the continental genetic diversity of avocados with Ethiopia, the genetic diversity and population structure can be traced, and the changes observed between the origin and the avocadoes introduced to Ethiopia can be uncovered. The accurate determination of the avocado races will facilitate breeding and avocado improvement.

2.12 Genetic markers in plant sciences

Genetic markers play a crucial role in plant breeding advancements (Kebriyaee et al. 2012). These markers are specific genes or DNA sequences located on chromosomes that control particular traits. They serve as indicators or flags closely associated with the target gene (Collard et al. 2005). Genetic markers can be categorized into classical markers, such as morphological, biochemical and cytological markers, and DNA/molecular markers, which include techniques like restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSRs), single-nucleotide polymorphism (SNP), and diversity arrays technology (DArT) markers (Jiang 2013).

2.12.1 Morphological markers

Morphological markers are valuable tools for distinguishing visual traits such as seed structure, flower color, growth habit, and other important agronomic characteristics. These markers are user-friendly and do not require specialized instruments, biochemical, or molecular techniques. Breeders have effectively utilized morphological markers in crop breeding programs. However, there are limitations to consider, including their limited number, susceptibility to plant growth stages, and environmental influences (Eagles et al. 2001). Throughout history, humans have employed various morphological markers to study variations for application in plant breeding (Karaköy et al. 2014).

Morphological markers typically pertain to external characteristics of plants. These markers aid in identifying, classifying, and understanding the genetic evolution of various species or populations. However, a plant's phenotype is influenced by both genetic factors and environmental conditions. Assessing plants genetic resources using morphological markers relies on subjective judgments and descriptions, often leading to incomplete accuracy. Additionally, measuring and identifying morphological traits is time-consuming, and mitigating the impact of environmental factors is challenging. Consequently, the utility of morphological markers is limited in evaluating quantitative traits. Nevertheless, they remain effective for assessing qualitative traits, wherein phenotypic differences can be readily observed and measured (Floate et al. 1994; Yang et al. 2013).

Avocado is a fruit of great economic and agricultural importance, known for its diverse and intricate range of morphological traits. Understanding the morphological

variability within avocado cultivars, native trees and wild relatives is fundamental for classification, breeding, and conservation efforts. The evaluation of variability is a crucial step in describing and characterizing germplasm. Morphological markers offer a practical approach to assessing the extent of variation. These observable traits play a significant role in conservation and sustainable utilization strategies (Ranjitha et al. 2021). Studies conducted worldwide, including Mexico (Rincón-Hernández et al. 2011), Colombia (López-Galé et al. 2022), Tanzania (Juma et al. 2020b), and Ghana (Abraham et al. 2018), have employed morphological characterization to assess the genetic diversity of avocados, demonstrating its validity and utility.

The fruit is a primary focus of morphological studies due to its significance in the avocado industry. Variability exists in fruit size, shape, skin texture, and color. The diversity in these attributes has led to the classification of avocado cultivars based on fruit morphology, providing valuable insights for growers and consumers alike (Bost et al. 2013; Pereira et al. 2013; Espinosa-Alonso et al. 2017). Avocado trees exhibit variations in leaf shape, tree habit, and flowering patterns. Morphological traits such as leaf size and shape play a role in the identification of different avocado varieties and can be pivotal in selecting rootstocks for grafting (Scora & Bergh 1992; Abraham et al. 2018; Acosta-Díaz et al. 2020; Juma et al. 2020b; López-Galé et al. 2022).

Understanding morphological characteristics related to flowering, such as flower size, structure, and pollen viability, is essential for breeding programs aiming to enhance avocado production and ensure pollination success (Alcaraz et al. 2013; Alcaraz & Hormaza 2021). Exploration of morphological variability extends beyond cultivated avocado varieties to include native trees. The study of native avocado germplasm has revealed valuable traits, including disease resistance and adaptation to diverse environmental conditions (Rincón-Hernández et al. 2011; Abraham et al. 2018; Juma et al. 2020b; López-Galé et al. 2022).

2.12.2 Cytological markers

Cytological markers, a category of molecular markers extensively employed in plant research, have played a pivotal role over the years. These markers are instrumental in identifying linkage groups and conducting physical mapping studies based on the observation of chromosome morphology and behavior during cell division (Adhikari et al. 2017).

Among the frequently utilized cytological markers is the karyotype, a visual representation of a cell's chromosomes. Karyotypes prove invaluable in pinpointing chromosomal abnormalities like deletions, duplications, and translocations, providing crucial insights into genetic variations (Cao & Deng 2020). Another essential cytological marker involves chromosomal banding patterns, discernible through various stains. These patterns facilitate the identification of specific chromosomal regions and serve as a window into studying chromosomal evolution (Jiang 2013).

While cytological markers have historically been indispensable in plant research, the landscape has transformed with the advent of genomics and transcriptomics. Researchers now have access to a diverse array of molecular markers that significantly enhance our understanding of plant genomes. These advanced molecular markers serve as powerful tools, enabling in-depth exploration of plant genome organization and evolution (Figueroa & Bass 2010).

2.12.3 Biochemical markers

Biochemical markers are used to detect variation at the gene product level, such as changes in proteins and amino acids (Poczai et al. 2013). In plant science, molecular markers are more commonly used to detect variation at the DNA level, such as nucleotide changes 1. However, there are some biochemical markers that are used in plant sciences. These markers are based on the variation in specific banding patterns of the proteins separated in Polyacrylamide Gel Electrophoresis (PAGE) (Ramesh et al. 2020). Biochemical markers are known for their near genetic neutrality which can be used for genetic variability studies and linkage map construction (Ramesh et al. 2020).

Biochemical markers have been used to study the genetic diversity of plant species, to identify genetic markers linked to specific traits, and to study the evolutionary relationships between different plant species (Figueroa & Bass 2010). They have also been used to study the effects of environmental factors on plant growth and development (Ražná et al. 2019). Biochemical markers have the advantage of being relatively easy to use and interpret, and they can be used to study a wide range of plant species. However, they also have some limitations, such as the fact that they are not always highly informative and may not be suitable for all types of studies (Ernst 1999).

In essence, biochemical markers have been fundamental in plant research for numerous years, leveraging specific banding patterns of proteins in PAGE to discern

amino acid sequence variances. These markers, heralded as the earliest molecular markers, have significantly contributed to understanding genetic diversity, population dynamics, and other intricate aspects of plant biology.

2.12.4 DNA-based molecular markers

The development and application of molecular markers began in the 1980s, marking a significant advancement in plant genomic research. PCR-based DNA markers were later achieved, revolutionizing plant molecular breeding and genomics (Nadeem et al. 2018). These markers, which are nucleic acids, play a crucial role in plant improvement research by exhibiting polymorphism among individuals or populations. They can be amplified using specific oligonucleotide primers in the PCR technique, leading to the generation of distinct fragments. Polymorphism can arise from point mutations in the priming sites of oligonucleotides or from changes in the distance between terminal sequences caused by insertion or deletion mutations (Collard et al. 2005). Overall, molecular markers have become invaluable tools in plant research, facilitating the study of genetic variation and aiding in the improvement of plants (Amiteye 2021).

Restriction Fragment Length Polymorphism (RFLP). This technique involves digesting DNA samples with specific endonucleases, resulting in a distinct profile of fragments with varying lengths that are characteristic of each species. RFLPs offer the advantage of medium polymorphic variability and can be used without prior knowledge of the genome sequence being analyzed. However, they also come with drawbacks, including high development and running costs, as well as the need for high-quality and abundant DNA samples (Chaudhary & Maurya 2019).

Random Amplified of Polymorphic DNA (RAPD). Random Amplification of Polymorphic DNA (RAPD) was independently developed in 1990 by the teams of Welsh and McClelland (Welsh & McClelland 1990) and Williams and colleagues (Williams et al. 1990). RAPD utilizes short random PCR primers (8–15 nucleotides) that bind to multiple genomic regions, resulting in distinctive PCR profiles specific to each species (Bardakci 2001). RAPD is based on highly polymorphic molecular markers and requires a moderate amount of DNA. It falls into an intermediate range in terms of technical

development and running costs. However, one of its drawbacks is the relatively low reproducibility of the obtained results (Bardakci 2001).

Amplified fragment length polymorphism (AFLP). This technique involves digesting DNA samples with two restriction enzymes following the annealing of adapters, which generate cut boundaries serving as primer binding sites for PCR amplification. Polymorphisms are identified based on the presence or absence of DNA fragments observed through analysis on polyacrylamide gels (Vos et al. 1995; Blears et al. 1998) These markers exhibit high levels of polymorphism and are plentiful within the genome. However, the technique is associated with certain limitations, including high development and running costs, a need for high-quality and -quantity DNA, prerequisite knowledge of the DNA sequence, and intermediate reproducibility with limited automation capability (Blears et al. 1998; Fry et al. 2009).

Sequence-related amplified polymorphism (SRAP). Li and Quiros (Li & Quiros 2001) developed a marker system called SRAPs, which amplifies open reading frames (ORFs). It uses two primers, 17-18 nucleotides long, with specific sequences. PCR is performed with varying annealing temperatures. The resulting amplified products are separated on a gel, and the presence or absence of DNA bands determines polymorphisms. SRAPs are widely used in map construction, genomic and cDNA fingerprinting (Salazar et al. 2014), offering a simple and efficient approach to investigate genetic variations in different taxa (Uzun et al. 2009).

Cleaved Amplified Polymorphic Sequence (CAPS). CAPS markers, originally known as PCR-RFLP markers, combine the techniques of RFLP and PCR (Maeda et al. 1990). In this method, the target DNA is first amplified through PCR and subsequently digested using restriction enzymes (Jarvis et al. 1994; Michaels & Amasino 1998).The resulting CAPS products are visualized on agarose or acrylamide gels. Primers used in CAPS are designed based on sequence information obtained from genomics databanks, cloned RAPD bands, or cDNA sequences. CAPS markers are versatile and can be enhanced by combining them with other techniques like single-strand conformational polymorphism (SSCP), SCAR, AFLP, or RAPD to increase the detection of DNA polymorphisms (Agarwal et al. 2008). CAPS markers are co-dominant and have found applications in

genotyping, map-based cloning, and molecular identification studies (Weiland & Yu 2003; Spaniolas et al. 2006).

Single nucleotide polymorphisms (SNPs). Single-nucleotide polymorphisms (SNPs), introduced by Lander in 1996 (Lander 1996), are genetic variations resulting from single nucleotide changes in the DNA sequence. They are more common in intergenic and non-coding regions compared to coding regions (Zhao et al. 2003). SNPs located in non-coding regions can still have functional or regulatory significance, as they may be linked to important genomic sites, indicating selection signatures (Van der Heyden et al. 2014). While SNPs are mostly bi-allelic, a larger number of polymorphic loci is needed to match the power of multi-allelic SSR loci (Guichoux et al. 2011). Next-generation sequencing has greatly facilitated SNP detection and genotyping, while targeting polymorphic sites in conserved single-copy genes offers an alternative strategy with comparable genetic resolution to multi-allelic SSRs (Dutech et al. 2007; Kaiser et al. 2017). Although SNPs are typically biallelic, rare cases of triallelism exist for specific positions (Oliveira & Azevedo 2022).

Variable Number of Tandem Repeats (VNTR). VNTR markers encompass minisatellites and microsatellites. Minisatellites, discovered by Alec Jeffreys and his team in 1990, consist of repeat motifs typically 9 to 30 bp in length (Jeffreys et al. 1985, 1990). Microsatellites, also known as Simple Sequence Repeats (SSRs) or Short Tandem Repeats (STRs), were first described by Litt and Luty in 1989. They consist of repeat motifs mostly 2 to 4 bp long, with tandem repeats of mono-, di-, tri-, tetra-, or pentanucleotide units scattered throughout the genome (Litt & Luty 1989).

Microsatellites are abundant and polymorphic in non-transcribed genomic regions, making them selectively neutral markers. However, SSR loci can also occur in regions associated with transcription, translation, chromatin organization, or recombination (Li et al. 2002). SSRs mutate at much higher rates, ranging from 10 to 100 thousand times more frequently per generation than single-nucleotide substitutions, due to replication slippage (Guichoux et al. 2011). Their high mutation rates and neutral evolution lead to the accumulation of population-specific alleles, which reveal hidden population structures. The multi-allelic nature of SSRs increases the likelihood of detecting heterozygosity compared to bi-allelic markers. However, the exceptional variability of SSRs in relation

to other genomic regions may not necessarily reflect genome-wide genetic diversity patterns (Fischer et al. 2017; Tsykun et al. 2017; Zimmerman et al. 2020).

Moreover, the rapid mutation rates of SSRs can introduce confounding signals of population structuring and divergence. For example, frequent forward and backward mutations can generate identical alleles in unrelated or genetically isolated populations (homoplasy). Increasing the number of polymorphic SSR loci used can compensate for this effect, but the level of genetic differentiation in populations that diverged long ago may still be underestimated (Estoup et al. 2002).

Minisatellites and microsatellites exhibit high levels of polymorphism and genomic abundance. They have low requirements in terms of DNA quality and quantity, and offer high reproducibility (Oliveira & Azevedo 2022).

Avocado is renowned for its genetic diversity, a valuable resource for breeding programs, conservation efforts, and understanding the species' evolutionary history. Given the growing interest in avocado production, numerous studies have investigated the genetic diversity of global avocado resources using various molecular markers. These markers include RAPD (Fiedler et al. 1998), RFLP (Davis et al. 1998), AFLP (Cañas-Gutiérrez et al. 2015), ISSRs (Cuiris-Pérez et al. 2009; López-Guzmán et al. 2021), SSRs (Ashworth et al. 2004; Gross-German & Viruel 2013; Boza et al. 2018; Cañas-Gutiérrez et al. 2019; Juma et al. 2020a), and SNP (Chen et al. 2009). More recently, Next Generation Sequencing (NGS) has been employed to assess the genetic diversity of avocado germplasm with greater precision, making significant contributions to breeding programs (Ge et al. 2019; Rubinstein et al. 2019; Talavera et al. 2019; Kämper et al. 2021). These markers enable researchers to characterize and differentiate avocado cultivars, assess their relatedness, and identify valuable alleles for breeding.

The native avocado trees, primarily in regions of Mexico and Central America, house a vast reservoir of genetic diversity. These native avocado populations exhibit unique traits, such as disease resistance, fruit quality, and adaptability to changing climates (Chen et al. 2009; Landon 2009; Avendaño-Arrazate et al. 2019). Understanding and harnessing the genetic diversity within these native avocado trees is pivotal for uncovering traits with the potential to enhance cultivated varieties. The geographic distribution of avocado exerts a substantial influence on its genetic richness. Climatic variables, such as temperature and rainfall, have significantly molded the genetic composition of avocado populations (Popenoe 1935; Williams 1977; Storey et al. 1986; Bergh 1992). Notably, the Chiapas - Guatemala - Honduras region, recognized as the center of origin for

avocados, stands out as a hotspot for genetic diversity. This region is of paramount importance for avocado breeding programs, holding the promise of valuable genetic resources for future advancements. Genetic diversity assessments provide breeders with insights into potential sources of beneficial alleles (Govindaraj et al. 2015; Sharma et al. 2021).

Considerate both morphological and genetic diversity is crucial for the conservation and sustainable utilization of avocado genetic resources. Conservation efforts involve maintaining a representative collection of diverse avocado accessions, including native trees. This inherent diversity represents an invaluable reservoir of genetic resources, offering immense potential for breeding programs. These programs are directed towards the creation of enhanced cultivars that exhibit superior characteristics, including heightened disease resistance, improved fruit quality, and adaptability to evolving climatic conditions (Swarup et al. 2021; Salgotra & Chauhan 2023).

The study of avocado diversity continues to evolve with advancements in genomics and phenotyping techniques. Next-generation sequencing and genome-wide association studies hold promise for uncovering the genetic basis of important traits (Alseekh et al. 2021; Karikari et al. 2023). Furthermore, integrating morphological and genetic data will provide a more comprehensive understanding of avocado diversity (Sunil et al. 2011; Sartie et al. 2012; Vinu et al. 2013; de Andrade et al. 2017). The morphological and genetic diversity of avocado is a critical aspect of its biology and cultivation. This thesis underscores the significance of characterizing and conserving this diversity for the sustainable development of avocado as a globally important crop. Understanding avocado's rich genetic heritage is essential for addressing challenges in production and ensuring the availability of diverse and resilient avocado varieties for future generations.

2.13 Native avocado trees

A native avocado tree refers to a tree that grows naturally and without direct human intervention in its native habitat. It is not cultivated, managed, or intentionally planted by humans. These trees have evolved and adapted over time to their specific ecological conditions, and they exhibit traits and characteristics that are distinct from cultivated or introduced avocado varieties. Native avocado trees are part of the natural ecosystem and contribute to the biodiversity of the region in which they are found (Wolstenholme & Whiley 1999; Ben-ya et al. 2003).

3 Objectives and Hypotheses

The study aimed to evaluate variability in native Guatemalan avocado germplasm through agro-morphological and genetic analysis, offering insights for genotype selection in conservation and breeding programs. Furthermore, the comparison with Ethiopian avocados aims to analyze the influence of introduced Guatemalan germplasm on the formation of cultivated genotypes and to identify opportunities for more efficient utilization of the available germplasm for breeding purposes in Ethiopia.

The specific objectives were:

i. to characterize the genetic resources of native Guatemalan avocado germplasm (*Persea americana* Mill.) based on agro-morphogical traits for genetic diversity analysis.
ii. to investigate the genetic variation, relatedness and population structure of native Guatemalan avocado genotypes using AFLPs and SSRs markers.
iii. to select a core collection of native Guatemalan avocado genotypes for long term conservation based on agro-morphological traits and molecular markers.
iv. to compare genetic diversity in Guatemalan and Ethiopian avocado germplasm to highlight valuable resources in Ethiopian avocado that can enhance global cultivation, resilience, and diversity conservation.

Hypotheses

i. The native Guatemalan avocado germplasm has undergone natural genetic diversification, resulting in a significant diversity of agro-morphological traits within the population.
ii. Native Guatemalan avocado genotypes will exhibit significant genetic variation based on AFLP and SSR markers, reflecting diverse alleles and genotypic profiles. Additionally, molecular data will indicate a correlation between genetic relatedness and geographic distribution, highlighting the influence of geography on the population's genetic structure.
iii. The core collection of avocados selected based on a combination of agro-morphological traits and molecular markers will consist of unique and rare genotypes, thereby ensuring the preservation of the genetic diversity of avocados.
iv. Guatemalan and Ethiopian avocado germplasm will exhibit distinct genetic variations, with Ethiopian avocado harboring unique traits valuable for improving global avocado cultivation, resilience, and diversity conservation.

4 Methodology

4.1 Study site and sampling in Guatemala

Using information from the Guatemalan atlas of wild relatives of cultivated plants (Azurdia et al. 2011) and in collaboration with the staff of Rafael Landívar University Herbarium, local experts and communities, a field survey was conducted to identify native avocado trees.

A total of 189 distinct avocado trees were sampled, representing eight geographic populations across three physiographic regions: Sacatepéquez, Chimaltenango, Sololá, Totonicapán, Quiché, Huehuetenango, Alta Verapaz, and Baja Verapaz departments in central, western, and northern regions (**Figure 7**). Native avocado tree samples were collected from diverse land use systems, including forests, homegardens, scattered trees in pastures, protected natural areas, riparian zones, and rural or indigenous lands. Ecological characteristics of each department examined are provided in **Table 3**. The number of individuals per population varied (8 to 36) depending on accessibility and availability. The latitude, longitude, and elevation of each tree site were recorded. For molecular analysis three fresh leaves from each individual tree were collected, 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.

Table 3. Ecological characteristics of the studied locations.

Location	Code	n	Lat	Lon	Region	T	R	masl
Chimaltenango	Chi	23	90.68	14.59	Central	9.2 °C	1,779 to 2,573	2,097 to 3,962
Sacatepéquez-Chimaltenango	Sac-Chi	32	90.81	14.59				
Sacatepéquez	Sac	36	91.06	14.80				
Sololá	Sol	8	91.16	14.83	Western	10.4 °C	1,141 to 2,056	1,801 to 2,990
Totonicapán-Quiché	To-Qui	36	91.41	15.02				
Huehuetenango	Hue	23	91.13	15.82				
Baja Verapaz	BV	11	89.94	15.20	Northern	15.8 °C	1,850 to 3,410	784 to 1,877
Alta Verapaz	AV	20	90.51	15.42				

n: number of sampled trees, T: mean annual temperature, R: annual rainfall range in mm yr^{-1}, masl: altitudinal range in meters above sea level.

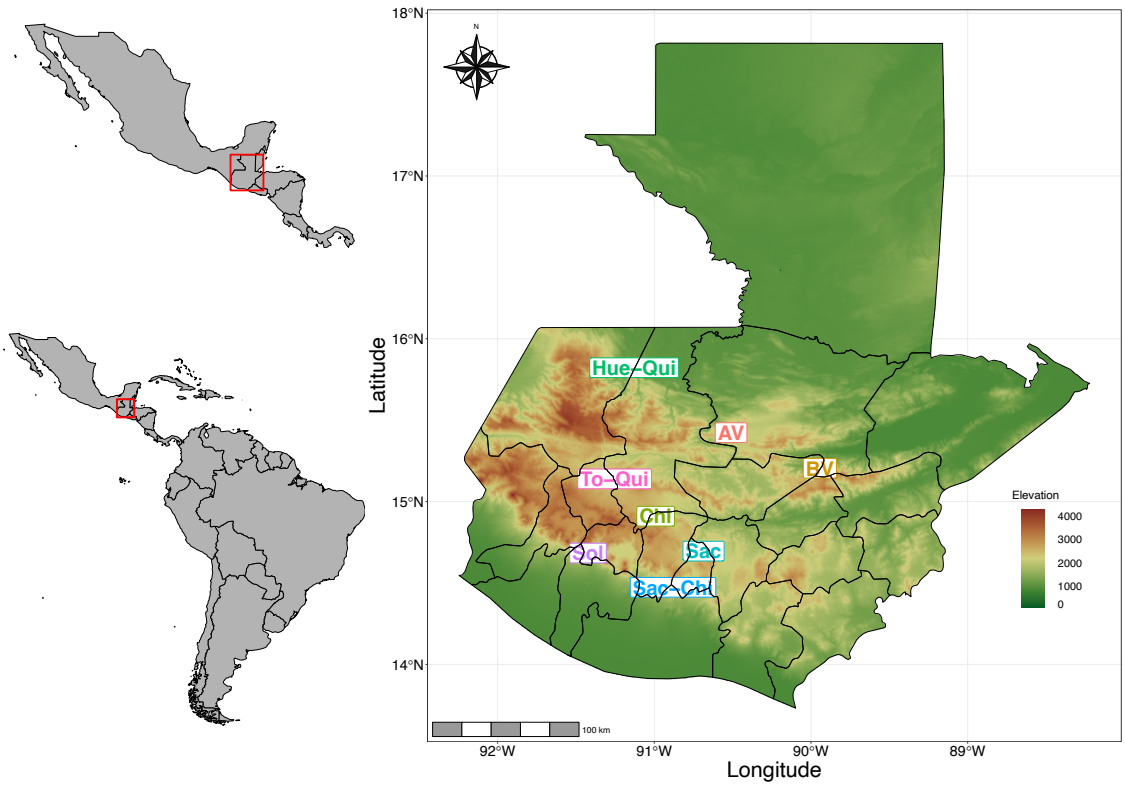


Figure 7. Map of Guatemala, displaying the geographical location of sampled avocado populations.

4.2 Variability analysis of native Guatemalan avocado germplasm based on agro-morphological traits

4.2.1 Measurement of qualitative and quantitative morphological traits

To characterize the selected avocado trees, the IPGRI field guide for avocado crops (IPGRI, 1995) was utilized, employing 21 plant descriptors for the assessment of the tree trunk, young twigs, leaves of comparable age, flowers, ripe fruits, and seeds. Following the methodology outlined by Juma et al. (2020), each tree was evaluated using five twigs, three to five leaves, two to four fruits, two to three flowers, and three seeds. These descriptors enable easy and rapid differentiation between phenotypes, exhibiting high heritability and consistent manifestation across different environments (IPGRI, 1995). The study focused on utilizing descriptors that provided maximum discrimination (**Table 4**).

Table 4. List of quantitative and qualitative traits assessed and their alternative variants.

Tree part	Quantitative character	Abbreviation	Measurement unit
Overall tree	Trunk circumference	TC	cm
	Leaf length	LL	mm
	Leaf width	LW	mm
Flower	Sepal length	SL	mm
	Fruit weight	FW	g
Fruit	Fruit length	FL	cm
	Seed weight	SW	g
	Qualitative character		**Level**
Overall tree	Trunk surface	TS	even, rugged, very rugged
	Color young twig	CYT	yellow, green, coopery, maroon, red
	Color mature leaf	CML	green, dark green
	Leaf shape	LS	ovate, narrowly, obovate, oval, roundish, cordiform, lanceolate, oblong, oblong-lanceolate
	Leaf anise smell	AS	absent, intense
Flower	Petal pubescent	PP	scarce, intermediate, dense
	Pedicel shape	PS	cylindrical, conical, rounded
Fruit	Fruit skin surface	FSS	even, intermediate, rugged
	Mature fruit skin color	MFSC	clear green, green, dark green, yellow, red, purple, black
	Fruit shape	FS	oblate, spheroid, high spheroid, ellipsoid, narrowly obovate, obovate, pyriform, clavate, rhomboidal
	Flesh texture	FT	watery, buttery, doughy, granular
Seed	Seed shape	SS	oblate, spheroid, ellipsoid, ovate, broadly ovate, cordiform, base flattened apex rounded, base flattened apex conical
	Cotyledon surface	CS	smooth, intermediate, rough

The evaluation of avocado traits involved sensory assessments, direct measurements, and visual observations. The tree characteristics involved assessing the smoothness or roughness of the trunk's surface through tactile examination, alongside the measurement of trunk circumference using a tape measure. Detailed measurements of leaves included recording leaf length, width, and sepal length using a vernier caliper. Visual observations were made to note the young twig color, leaf shape and mature leaf color. The presence

of an anise odor, achieved by crushing and smelling the leaves. The petal pubescence and pedicel shape in flowers was observed visually. Evaluation of fruits incorporated sensory testing to determine the texture of the fruit's flesh. The weight of the fruit was measured using a portable semi-analytical balance. Shape and mature fruit skin color were compared to reference pictures from a field guide. The weight of the seeds was measured using the same portable semi-analytical balance. Cotyledon surface texture was determined by tactile examination.

4.2.2 Data analysis

The statistical analysis was conducted using the *compareGroups* package (Subirana et al. 2014) in R software v.4.2.0 (R Core Team 2022). One-way analysis of variance (ANOVA) and Tukey's tests were performed to assess the significance of the location factor on the measured morphological values. The coefficient of variation was calculated to determine the variability across populations for each quantitative attribute. The *GGally* package (Schloerke et al. 2021) was utilized to estimate the correlation matrix. For the qualitative morphological attributes, a cross-tabulation statistical approach was employed to examine the frequency distribution among populations. The Pearson Chi-square ($\chi 2$) test was used to determine the relationship between cross-tabulation variables. The *ggstatsplot* package (Patil 2021) in R was utilized for the cross-tabulation and Chi-square tests. Significance threshold of 0.05 was used for all statistical analysis.

The morphological relationships among the 189 avocado trees and the association between quantitative and qualitative variables were explored using factor analysis of mixed data (FAMD). FAMD combines the principles of principal component analysis (PCA) and multiple correspondence analysis (MCA) and is specifically designed for analyzing datasets that contain both quantitative and qualitative variables, balancing their influence in the analysis (Pages 2004). To ensure equal contribution of variables measured at different scales in the FAMD analysis, standardization of variables was performed prior to analysis (Kenkel 2006), optimizing the variance explained in each dimension. The FAMD was performed and visualized using the *FactoMineR* (Lê et al. 2008) and *factoextra* packages (Kassambara & Mundt 2020), respectively.

Hierarchical Clustering on Principal Components (HCPC), based on the FAMD analysis, was applied to identify clustering patterns and population structure among the sampled trees. HCPC combines principal component methods, hierarchical clustering,

and partitioning clustering, including the k-means method. The *factoextra* package was used to visualize the HCPC results through a dendrogram and factor map. The dendrogram was exported in Newick tree format to the Interactive *Tree Of Life* v6 (iTOL) (Letunic & Bork 2019) for customization and visualization. Test values (Morineau 1984) were computed to rank the most distinctive quantitative and qualitative categories within each cluster identified by HCPC.

4.3 Genetic diversity and population structure of native Guatemalan avocado using AFLP and SSRs markers

4.3.1 DNA isolation

The cetrimonium bromide (CTAB) technique was used to extract DNA (Doyle & Doyle 1987). A Nanodrop^{tm} (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}.

4.3.2 AFLP protocol

The AFLP molecular marker (Vos et al. 1995) was used due to its high capacity in identifying polymorphic regions in avocado genetic diversity studies (Ramírez et al. 2005; Gutiérrez-Díez et al. 2009; Nerdo et al. 2009; Cañas-Gutiérrez et al. 2015; Cerda-Hurtado et al. 2015). The analysis was performed using the AFLP Analysis System I Kit from Invitrogen® following the manufacturer's instructions (Zabeau & Vos 1993).

Selective amplification was performed using the M-CAA + E-AAC primers from the kit, which produced higher polymorphism. The PCR products were visualized on a 5% acrylamide gel to observe the amplified bands. The gel staining was carried out with silver nitrate using the following procedure: immersion in a fixing solution of 10% v/v glacial acetic acid for 35 minutes, distilled water for 20 minutes, staining solution (0.15% w/v silver nitrate, 0.15% v/v formaldehyde) for 40 minutes, distilled water for 10 seconds, revelation solution for 5 to 6 minutes according to the appearance of bands (6% w/v sodium carbonate, 0.3% v/v formaldehyde, 5 ppm sodium thiosulfate), fixing solution to stop the process for approximately 5 minutes, and finally washed in distilled water. The gel was allowed to dry for 2 days, and then the band profile was read (Karam et al. 2006).

4.3.3 SSR protocol

Twelve microsatellite primer pairs earlier designed for *P. americana* (Sharon et al. 1997; Ashworth et al. 2004) were utilized to amplify all DNA samples. We used four different colors to mark forward primers for detection fluorescently. Three multiplex PCRs were performed with varying annealing temperatures for each one and optimized concentrations for each primer (**Table A1**). 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 **Table A1**, and Multiplex PCR Plus (1 X) (QIAGEN®, DUS, DE). The Thermal Cycler T 100 (BIO-RAD, HER, 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, FSTC, CA, USA). A 1 µL aliquot of PCR products was combined with 0.2 µL of GeneScan-500 LIZ (Applied Biosystems, FSTC, CA, USA) and 12 µL of Hi-Di formamide (Applied Biosystems, FSTC, CA, USA). GeneMarker v.2.4.0 was used to score the microsatellite alleles (Softgenetics, CS, PA, USA).

4.3.4 Data analysis

Genetic diversity analysis. With the band profile generated with AFLP molecular, a binary matrix of presence (1) and absence (0) was generated for each amplified locus, from which the analysis of genetic diversity and population genetic structure was conducted using the *AFLP-Surv* 1.0 program (Vekemans et al. 2002). For this analysis, the number of polymorphic loci (# loc_P), percentage of polymorphic loci (PLP), expected heterozygosity under Hardy-Weinberg equilibrium (Hj), average genetic diversity within populations (Hw), average genetic diversity among populations (Hb), total genetic diversity (Ht), Wright's fixation index (F_{ST}), and genetic distance matrix with Nei's index were estimated.

With the SSR data *GenoDive* v.3.05 (Meirmans 2020) was used to compute basic statistics including 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}) in order to evaluate genetic diversity at each locus and population. *PopGenReport* v.3.0.7 package (Adamack & Gruber 2014) was used to

estimate the null allele frequency for each SSR locus. *FSTAT* software was employed to calculate allelic richness (Ar) using a rarefaction method (Goudet 1995). The formula for determining gene flow (Nm) was $Nm = (1 - G'_{ST(Nei)}) / (4 * G'_{ST(Nei)})$ (Slatkin & Barton 1989). Shannon's information index (I) was calculated using *GenAlEX* v.6.5 (Peakall & Smouse 2012). Tests for Hardy-Weinberg equilibrium and Analysis of Molecular Variance (AMOVA) were conducted using *GenoDive*.

Genetic differentiation analysis. Genetic differentiation of populations was determined through AMOVA implemented in *poppr* library (Kamvar et al. 2014). Covariance components were used to calculate fixation indices and determine gene flow and differentiation between populations. A randomization test with 1,000 permutations determined significance. To assess isolation by distance, the Mantel test was performed in *ade4* package v.1.7 (Bougeard & Dray 2018) on the genetic and geographic distance matrices with 10,000 permutations. Pairwise population differentiation was analyzed using the F_{ST} index with 1,000 permutations in the *mmod* package.

Population structure analysis. The genetic structure of the avocado samples was explored using a hierarchical cluster analysis. The *poppr* package was used to calculate Nei's genetic distance and apply a hierarchical Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering. To compare the results of the UPGMA analysis, a Minimum Spanning Network (MSN) was constructed, as it is an excellent method for visualizing the genetic relationships among individuals based on a genetic distance matrix. The MSN was constructed and visualized using the *igraph* and *poppr* packages, respectively. To further examine the structure of the populations, a Bayesian clustering method was run using *STRUCTURE* v2.3.4 (Pritchard et al. 2000) with 10,000 steps and 100,000 MCMC iterations for cluster numbers from 1 to 10. 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 (Wang 2017). The optimal value of K was estimated with the methods of Evanno (Evanno *et al.* 2005) and Puechmaille (Puechmaille 2016). The results were analyzed and visualized through *StructureSelector* (Li & Liu 2018) web server and the *pophelper* library (Francis 2017).

STRUCTURE population assignment was confirmed with a DAPC analysis (Jombart *et al*. 2010) implemente in *adegenet* library (Jombart 2008). 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 the "*optim.a.score"* function.

Joint analysis of phenotypic and molecular data. To examine the relationship among the morphological and genetic data a tanglegram analysis was employed. This relatively recently introduced approach visually compares two dendrograms with the same terminal vertices, presenting a side-by-side representation of both dendrograms. Matching objects are linked by straight-line segments, referred to as inter-tree edges (Wotzlaw et al. 2012). This approach allowed to assess the correspondence and relationships between the two-clustering generated from both kind of data. To conduct tanglegram analysis, first genetic distances were calculated with SSR data set. The pairwise distances were then hierarchically clustered using Ward's method with the *ape* package v.5.6 (Paradis & Schliep 2019) in R, and the results were visualized through a dendrogram. For morphological data, first the factor analysis of mixed data (FAMD) was used as implemented in *FactoMineR* package. Hierarchical Clustering on Principal Components (HCPC), based on FAMD results, was applied to create a dendrogram using Ward's method and identify clustering among the sampled trees.

The tanglegram was carried out using the *dendextend* package v.1.15.2 (Galili 2015) in R, taking both dendrograms as input. The entanglement between the two dendrograms was computed. Entanglement is a measure with value between 1 (fully mismatched labels) and 0 (fully aligned labels). Additionally, the cophenetic correlation coefficient was used to estimate the correlation between the dendrograms. The value can range between -1 to 1 with near 0 values meaning that the two trees are not statistically similar.

Furthermore, genetic groups were established by integrating phenotypic trait-based and genetic distance matrices. The joint matrix was created by summing both matrices using the *sidier* R package v.4.1.0 (Muñoz-Pajares 2013) and Ward's method were used to create a hierarchical cluster dendrogram. Finally, correlation between

genetic, morphological, and joint distance matrices was computed with the Mantel test at 10,000 permutations in *ade4* package.

4.4 Selection of a core collection of avocado genotypes for long term conservation based on agro-morphological traits and molecular markers

4.4.1 Development of the core collection

Initially, seven distinct core collections were generated, which included one core collection developed using the Sequential Backward Selection as subsetting strategy in the R package *GeneticSubsetter* v.0.8 (Graebner & Cuesta-Marcos 2016) using SSR data. With the joint distance matrix previously described, another core collection was constructed applying the accession nearest entry method and expected heterozygosity criteria using the *CoreCollection* package v.0.9.5 (Brouwer & Blok 2022) implemented in R. Furthermore, a combined chdata object was constructed by incorporating the phenotypic, molecular, and joint distance matrices. Subsequently, the *corehunter* package v.3.2.2 (De Beukelaer et al. 2018; De Beukelaer & G 2023) was utilized to generate five core collections (CC) based on the combined data, following optimization of average genetic distance-based criteria, as described in Odong et al. (2013). The methods encompassed the optimization of average genetic distances between each accession and the nearest entry in the core (A-NE) and the average distance between each entry and its closest neighboring entry (E-NE) as suggested by Kaur et al. (2022).

- maximizing E-NE distances (CC 01)
- maximizing A-NE distance (CC 02)
- maximizing both E-NE and A-NE with equal weightage of 1:1 (CC 03)
- E-NE and A-NE with unequal weightage of 0.3:0.7 (CC 04)
- E-NE and A-NE with equal weightage of 0.7:0.3 (CC 05).

The core set size was determined to be approximately 20% of the entire collection based on the neutral allele theory (Brown 1989).

4.4.2 Assessment of the core collections

A comprehensive comparison of the seven core sets was conducted, utilizing genetic distance criteria as outlined by Odong et al. (2013). Various statistical parameters, including mean difference percentage (MD%), variance difference percentage (VD%),

variable rate of coefficient of variance (VR%), and coincidence rate of range (CR%) for quantitative traits (Hu et al. 2000), were calculated. For qualitative traits, the coverage criteria were applied (Kim et al. 2007). To evaluate the correlation between the trait correlation matrices of the core collection and the entire collection, the Mantel test (Mantel 1967) was performed.

The MD% should not exceed 20%, signifying minimal differences in trait means between the core and primary collections. An optimal CR% should surpass 80%, indicating substantial overlap in trait ranges between the core and primary collections. Moreover, a robust core collection exhibits lower VD values and higher VR values, reflecting effective capture of diversity compared to the primary collection. Meeting these criteria ensures the core collection's efficiency in preserving primary collection diversity.

After the identification of the optimal core collection with maximal diversity and representativeness, a comparative analysis of quantitative trait means between the selected core set and the entire collection was conducted. This analysis involved the utilization of the Newman–Keuls test (Newman 1939; Keuls 1952) and t-test. We assessed the homogeneity of variances for quantitative traits in both the entire germplasm and the selected core collection using Levene's test (Levene 1960). Furthermore, the Wilcoxon rank test (Wilcoxon 1945) was employed to evaluate differences in frequency distribution. To provide a visual comparison of frequency distribution between the entire germplasm and the core collection, boxplots were generated.

To provide a comprehensive comparison of the distribution patterns of continuous traits between the core set and the entire collection, we generated quantile-quantile (QQ) plots (Wilk & Gnanadesikan 1968) and computed Kullback–Leibler distances (Kullback & Leibler 1951). The assessment of phenotypic diversity included the calculation of the Shannon–Weaver diversity index (H') and evenness using the frequencies of qualitative traits (Shannon 1948). Additionally, we analyzed the interrelationships between various quantitative and qualitative traits in both the entire germplasm and the core collection through Pearson correlation coefficients. To unravel trait relationships and their contributions to multivariate variation, we applied Principal Component Analysis (PCA). All statistical analyses related to the core collection were conducted using the R package *EvaluateCore*.

4.5 Comparison of the genetic diversity of Guatemalan and Ethiopian avocado germplasm

4.5.1 Study site and sampling in Ethiopia

A total of 109 avocado accessions, originating from three Ethiopian regions (Sidama, Gamo, Wolaita), and the Wondo Genet Research Centre (WGRC) were included in this study. All individual samples are landraces that started as chance seedlings, except a few recently created hybrids. The young leaves were collected in a falcon tube (15 ml) filled with the color indicator Silica Gel Desiccant. The samples were shipped to the Faculty of Tropical AgriSciences, Czech University of Life Sciences, Prague, for molecular analysis. In addition to the Ethiopian accessions, native avocado samples collected from Central, West, and North Guatemala were also included in the comparison (**Figure 8**).

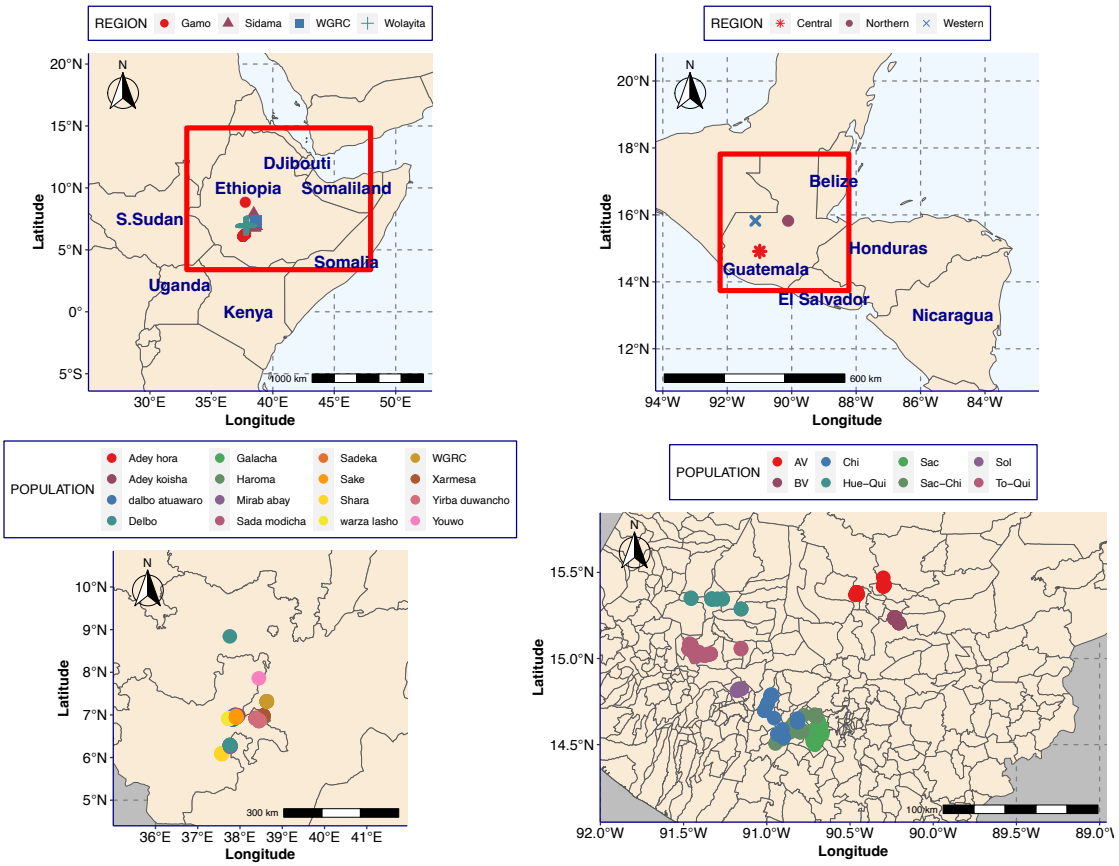


Figure 8. Map depicting the geographical locations of the sampled avocado regions in Ethiopia and Guatemala. The regions from Ethiopia include Sidama, Gamo, Wolaita, and the Wondo Genet Research Centre (WGRC). Additionally, the map displays the regions in Central, West, and North Guatemala where native avocado samples were collected.

4.5.2 DNA isolation

The cetrimonium bromide (CTAB) technique was used to extract DNA (Doyle & Doyle 1987). A Nanodrop^{tm} (Thermofisher Scientific, WA, 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}.

4.5.3 SSR markers and PCR amplification

Twelve microsatellite primer pairs, originally designed for *P. americana* (Sharon et al. 1997; Ashworth et al. 2004), were used to amplify all DNA samples. Those loci are the same used for the Guatemalan samples. The same method described in section 4.3.3 was applied.

4.5.4 Data analysis

The data was counted by peak feature and fragment size of the corresponding peaks, which were inserted into the GenAlEx v.6.5.03 program (Peakall & Smouse 2012) for further analysis. The GenAlEx program was used to estimate locus-based diversity indices such as the number of alleles (Na), the effective number of alleles (Ne), observed heterozygosity (Ho), expected heterozygosity, (He), and gene diversity (GD). Hardy-Weinberg Equilibrium (HWE), inter-population diversity was calculated with Popgene32 software v.1.32 (Labate 2000). The Polymorphic Information Content by locus was estimated using *PopGenUtils* v.0.1.8 R package (Tourvas 2023). AMOVA and the average gene diversity at the loci for each population using the GenAlEx version program. The GenAlEx program was further used to compute fixation indices (F_{ST}, F_{IT}, F_{IS}, F_{CT}, and F_{SC}) and pairwise comparisons between populations.

Nei's genetic distance (G) was computed in *GenAlEx* and imported into *MEGAX* (Kumar et al. 2018), where the dendrogram in the Newick format was produced using the unweighted pair group method with arithmetic mean UPGMA (Sneath & Sokal 1973). The dendrogram was visualized and customized in the Interactive Tree of Life (*iTOL*) v4 (Letunic & Bork 2019).

The population structure was constructed using a Bayesian algorithm implemented in the *STRUCTURE* software, v.2.3.4 (Evanno et al. 2005). The analysis was based on an admixture ancestral model with correlated allele frequency. To determine

the number of population groups (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 10 iterations for each K. The optimal value of K was estimated through *StructureSelector* web server and the *pophelper* library. The *adegenet* package (Jombart 2008) and the Bayesian Information Criterion (BIC) value were used to choose the optimal number of gene clusters. Clusters were visualized with a scatter plot and bar graph. They were also confirmed with a DAPC using a priori information and cross-validation.

To find variance components, the *poppr* package (Kamvar et al. 2014) with 10,000 permutations was used to analyze the molecular genetic variation between and within populations. Based on the population groups identified by the *STRUCTURE* analysis, AMOVA was carried out. AMOVA-derived genetic differentiation values (F_{ST}) between populations were calculated and compared to conventional F statistics. The degree of diversity within the population was measured using Shannon's information index (I).

5 Results

5.1 Variability analysis of native Guatemalan avocado germplasm based on agro-morphological traits

5.1.1 Quantitative traits

The description of quantitative traits of each of eight geographic avocado populations is displayed in **Table 5**. Among all populations, the highest trunk circumference (TC) was observed in Sac-Chi, measuring 118.33 cm, while the Chi population exhibited the lowest TC value of 83 cm. The coefficient of variation (CV) for TC ranged from 10.6% (Sac-Chi) to 33.4% (BV), with a mean of 22.2%.

Fruit weight and length ranged from 164.8 g (AV) to 393.8 g (Hue-Qui) and 9.9 cm (To-Qui) to 13.1 cm (BV), respectively. For both fruit traits, the values are consistent with those reported from Mexico (López-Guzmán et al. 2015) and Colombia (López-Galé et al. 2022). For seed traits, the weight varied from 75 g (AV) to 96.02 g (Sac), with an overall mean of 86.6 g and the highest and lowest CV of 22.98% (Chi) to 14.26% (Sac-Chi). With the analysis of variance, excluding leaf width and petal length, all quantitative traits exhibited statistically significant ($p < 0.05$) differences when compared between populations. After the correlation analysis (**Figure 9**) the variables of fruit weight and seed weight demonstrated a moderate positive correlation ($r = .49, p < 0.0001$). The remaining variables were not correlated (**Figure 9**).

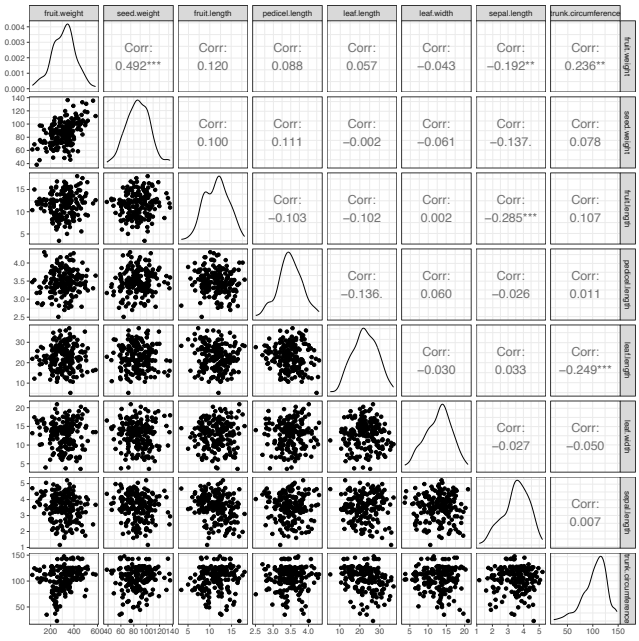


Figure 9. Correlation matrix showing Pearson's correlation coefficient between each quantitative variable.

Table 5. Description of quantitative traits among geographical populations. Mean values, standard deviation (SD) and coefficient of variation (CV) complemented by Tukey post-hoc test.

L	A	E	TC (cm)	LL (cm)	LW (cm)	SL (mm)	FW (g)	FL (cm)	PL (mm)	SW (g)
Sac	1,725	mean	118.22 ^{a}	20.61^{b}	12.82	3.51^{ab}	364.83 ^{a}	11.29 ^{ab}	3.55	96.02 ^{a}
		SD	13.042	5.49	3.873	0.636	85.133	2.745	0.281	18.5
		CV	11.03	26.66	30.21	18.13	23.34	24.32	7.93	19.27
Chi	2,054	mean	118.33 ^{a}	20.34 ^{b}	12.06	3.57 ^{ab}	260.04 ^{b}	13.04 ^{a}	3.5	82.68^{ab}
		SD	12.478	4.479	3.161	0.701	69.417	1.95	0.276	11.791
		CV	10.55	22.02	26.21	19.64	26.69	14.96	7.91	14.26
Sac-Chi	2,305	mean	82.59 ^{c}	24.68 ^{ab}	13.8	3.43 ^{ab}	305.09 ^{b}	10.64 ^{ab}	3.55	84.05^{ab}
		SD	25.555	6.283	3.39	0.838	75.323	2.933	0.332	19.32
		CV	30.94	25.46	24.57	24.44	24.69	27.57	9.35	22.98
Sol	3,315	mean	117.99 ^{ab}	20.31^{b}	13.36	3.07 ^{b}	349.51^{ab}	11.13 ^{ab}	3.59	88.24^{ab}
		SD	13.071	7.203	3.048	0.659	57.75	2.345	0.267	13.283
		CV	11.08	35.47	22.82	21.46	16.52	21.06	7.45	15.05
Hue-Qui	1,469	mean	93.2 ^{bc}	22.51^{ab}	13.68	4.03 ^{a}	295.04 ^{b}	9.85 ^{b}	3.37	79.69 ^{b}
		SD	23.295	7.566	2.532	0.708	54.382	2.948	0.42	17.684
		CV	25	33.6	18.5	17.7	18.4	29.9	12.3	22.2
To-Qui	2820	mean	104.64 ^{ab}	24.83 ^{a}	12.45	3.26 ^{b}	393.82 ^{a}	12.08 ^{a}	3.46	95.28 ^{a}
		SD	18.145	5.387	4.043	0.93	74.263	2.877	0.29	15.161
		CV	17.34	21.7	32.47	28.53	18.86	23.82	8.34	15.91
BV	810	mean	103.21^{abc}	26.91^{a}	10.11	3.17 ^{b}	295.53 ^{b}	13.13 ^{a}	3.31	76.5 ^{b}
		SD	34.498	5.548	3.289	1.095	46.472	2.377	0.426	12.77
		CV	33.4	20.6	32.5	34.5	15.7	18.1	12.9	16.7
AV	1294	mean	105.79 ^{ab}	20.66^{ab}	12.9	3.88^{ab}	164.8 ^{c}	12.05 ^{ab}	3.41	75.04 ^{b}
		SD	13.731	4.885	3.74	0.636	91.705	1.501	0.39	16.99
		CV	13	23.6	29	16.4	55.6	12.5	11.4	22.6
Total	--	mean	104.37	22.66	12.8	3.52	313.14	11.55	3.48	86.56
		SD	23.14	6.11	3.57	0.81	99.39	2.75	0.33	18.05
		CV	22.2	27	27.9	23.14	31.7	23.8	9.59	20.9

L: location, A: altitude in meters above sea level, E: estimator, N: number of sampled trees, TC: trunk circumference, LL: leaf length, LW: leaf width, SL: sepal length, FW: fruit weight, FL: fruit length, PL: pedicel length, SW: seed weight. Different letters indicate significant differences ($p < 0.05$)

5.1.2 Qualitative traits

The analysis of qualitative traits revealed three variations in trunk surface among the studied trees: even, rugged, and very rugged, accounting for 54.5%, 24.9%, and 20.6% of all samples, respectively. A significant association between trunk surface and populations was supported by statistical analysis ($\chi 2 = 29.21$, df = 14, $p < 0.001$) (**Figure 10**A).

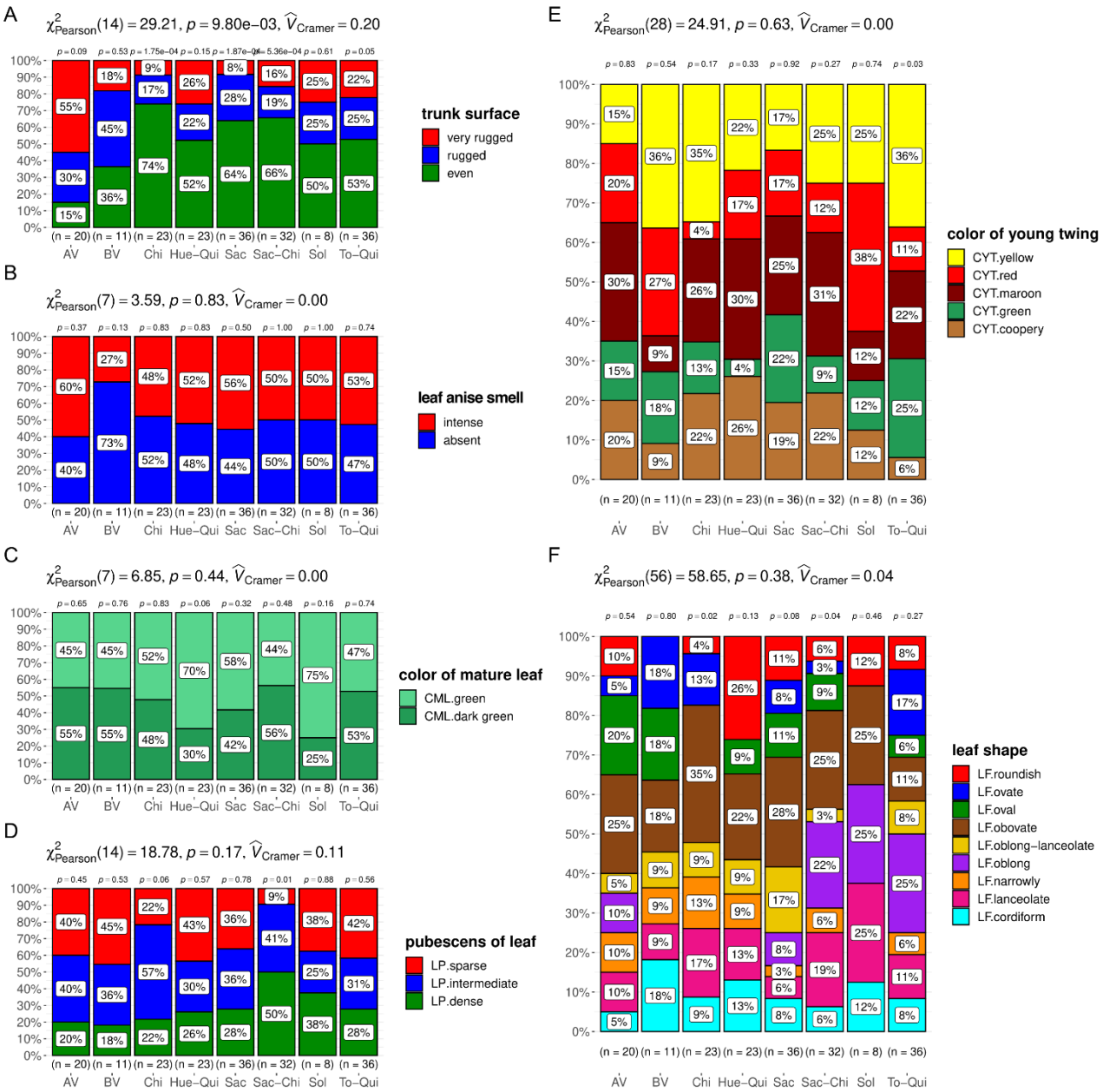


Figure 10. Description of frequency and chi-squared test of the qualitative traits of the overall tree. A: trunk surface. B: leaf anise smell. C: color of the mature leaf. D: pubescence of leaf. E: color of the young twig. F: leaf shape

Regarding the presence of anise aroma in leaves, it was found that 48% of the sampled trees did not exhibit this aroma, while the remaining 52% had the aroma. Statistical analysis did not reveal any association between anise aroma and populations (**Figure 10**B). In terms of leaf pubescence, it was observed to be sparse, intermediate,

and dense in 33%, 38%, and 30 % of the samples, respectively. No significant association was found between leaf pubescence and populations ($p > 0.05$) (**Figure 10**D). Nine different leaf shapes were identified in avocado characterization from Guatemala, and there was no statistical association between this trait and geographical populations (**Figure 10**F). The color of mature leaf (**Figure 10**C) and the color of mature twing did not exhibit significant association with the geographical populations (**Figure 10**E).

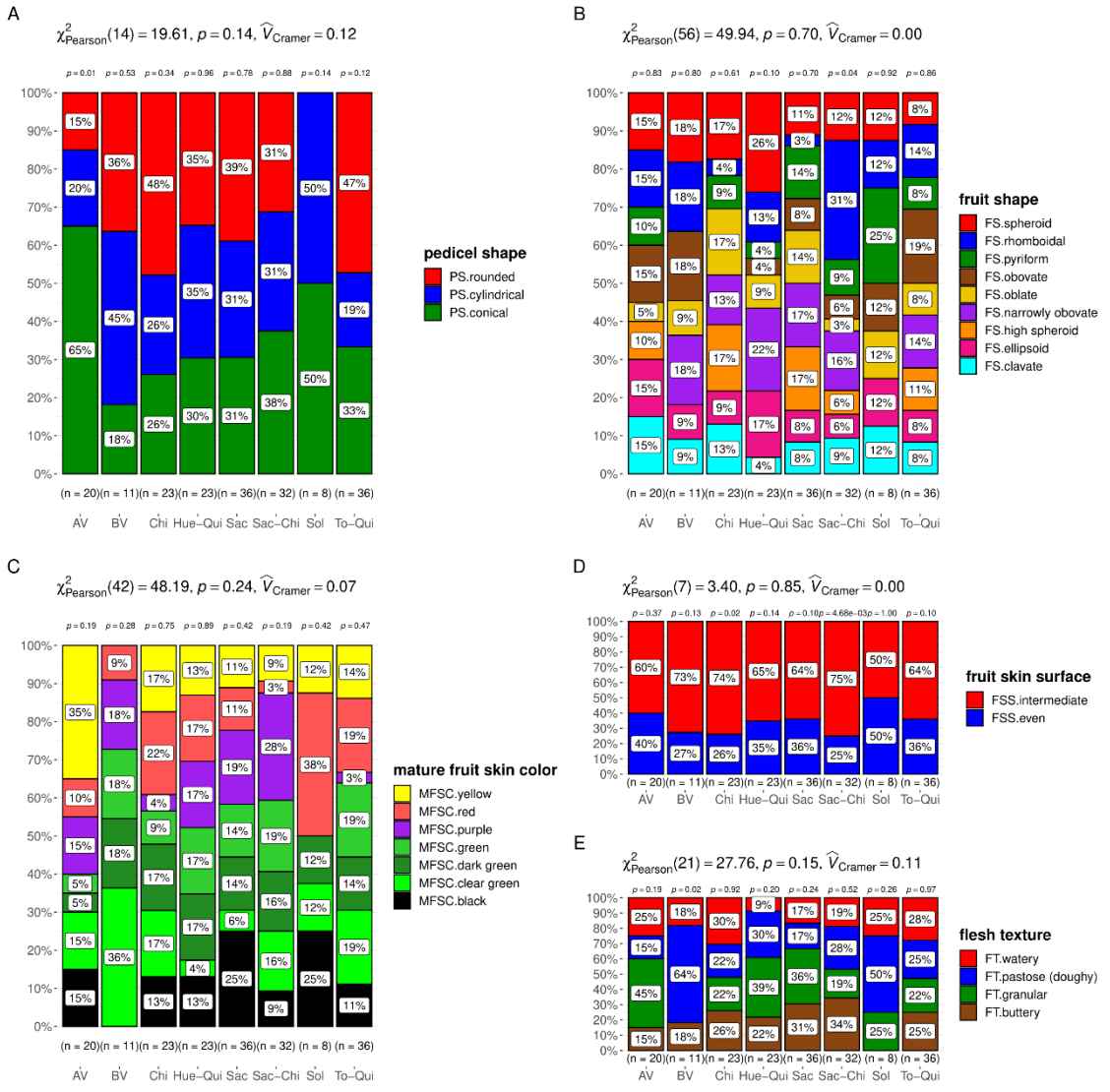


Figure 11. Description of frequency and chi-squared test of the fruit's qualitative traits. A: pedicel shape. B: fruit shape. C: mature fruit skin color. D: fruit skin surface. E: flesh texture.

The avocado fruit exhibited three distinct pedicel shapes and nine different fruit shapes in this study (Figure 11A, B). Four different fruit flesh textures were described:

buttery (35%), watery (21%), pastose (26%), and granular (18%) (Figure 11E). The fruit shape and skin surface did not show significant association with the geographical populations (**Figure 11**C, D) A total of eight seed shapes were observed in the avocado fruits, and the distribution of these shapes is presented in **Figure 12**A. The Chi-squared test did not yield significant results, except for BV, which showed an association with oblate and cordiform seed shapes. Regarding the surface texture of the seed cotyledon, smooth, rough, and intermediate textures were recorded, with proportions of 37%, 36%, and 28%, respectively, among the eight populations. However, no statistically significant association was found between the cotyledon surface and the populations (**Figure 12**B).

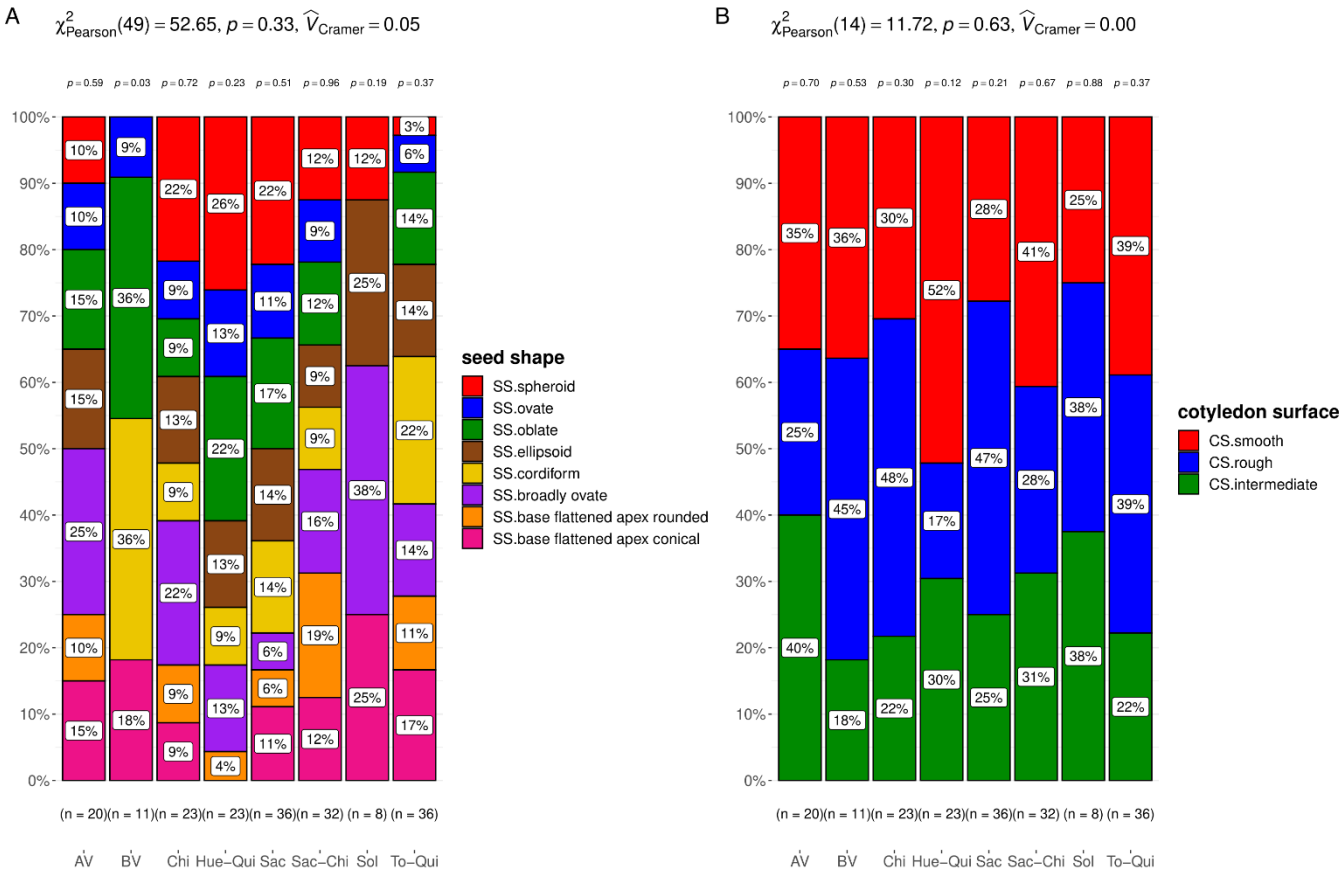


Figure 12. Description of frequency and chi-squared test of the avocado seed qualitative traits. A panel: seed shape. B panel: cotyledon surface.

5.1.3 Factorial Analysis of Mixed Data

The factorial analysis of mixed data (FAMD) was conducted to examine the relationship between variables, revealing that dimensions 1, 2, and 3 accounted for 21.1%, 15.6%, and

14.9% of the total variance, respectively. Additionally, by considering 5 dimensions, a cumulative variance of 75% was achieved. Notably, fruit shape, skin color, seed shape, flesh texture, and anise odor in the leaves contributed 13.9%, 13.7%, 13.6%, 12.6%, and 5.0% to the first two dimensions, respectively (**Figure 13**).

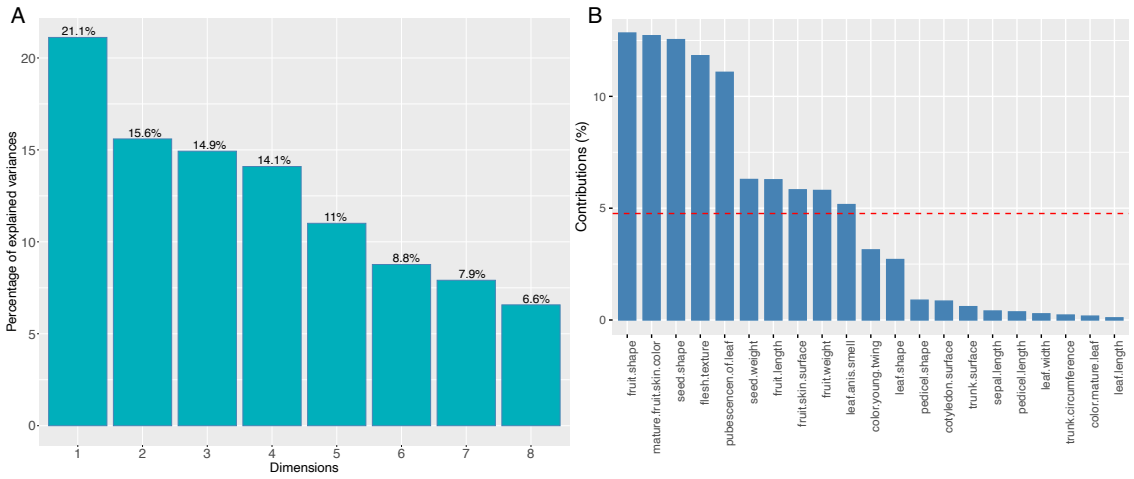


Figure 13. Scree plot showing the percentage of variance explained by each of the first 8 dimensions (A). Contribution of avocado variables to Dimension 1 and 2 (B). The red dashed line on the graph above indicates the expected average value, if the contributions were uniform.

The correlation between variables and the primary two dimensions was assessed using coordinates, squared cosine (cos2), and contribution. The depiction of variables on the factor map was determined by the angle between the variable point and the axis, represented by cos2. Additionally, contribution was calculated based on the separation between the perpendicularly projected point of the variable and the associated dimension axis. Based on the positive or negative contribution, variables were placed in different quadrants of the two dimensions. In the quantitative plot (**Figure 14**), variables such as fruit weight and seed weight were highlighted in a redder color. Similarly, in the qualitative plot (**Figure 15**), variables including FT. buttery, FSS.even, and SS.spheroid were also emphasized in a redder color.

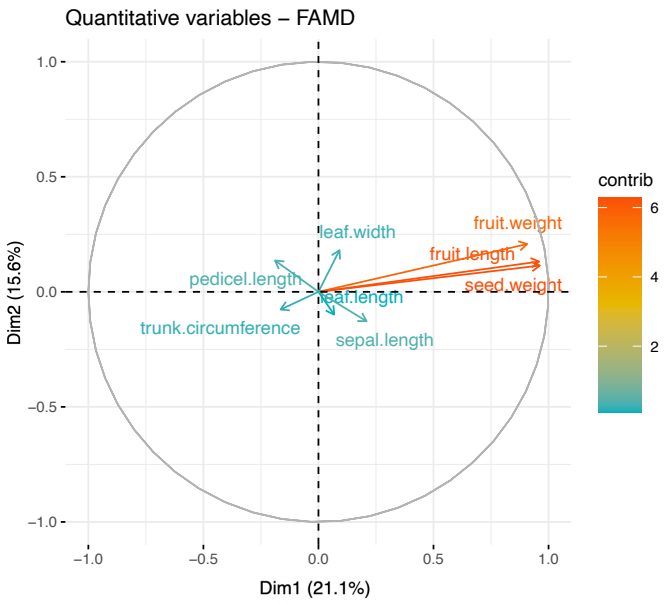


Figure 14. Correlation between quantitative variables from the FAMD analysis.

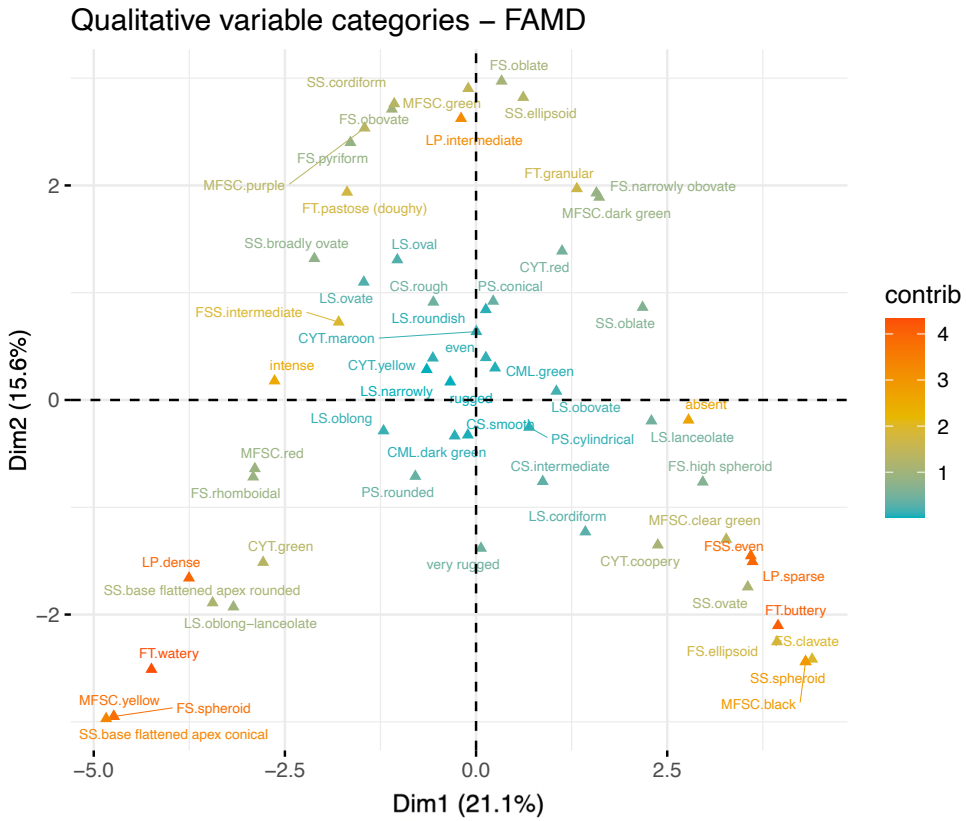


Figure 15. Correlation between qualitative variables from the FAMD analysis.

TS: trunk surface; CYT: color young twig; CML: color mature leaf; LS: leaf shape; AS: leaf anise smell; PP: petal pubescent; PS: pedicel shape; FSS: fruit skin surface; MFSC: mature fruit skin color; FS: fruit shape; FT: flesh texture; SS: seed shape; CS: cotyledon surface.

An association between all quantitative and qualitative variables of the analyzed avocado germplasm and their usefulness for profiling among sampled trees was revealed by the FAMD (**Figure 16**).

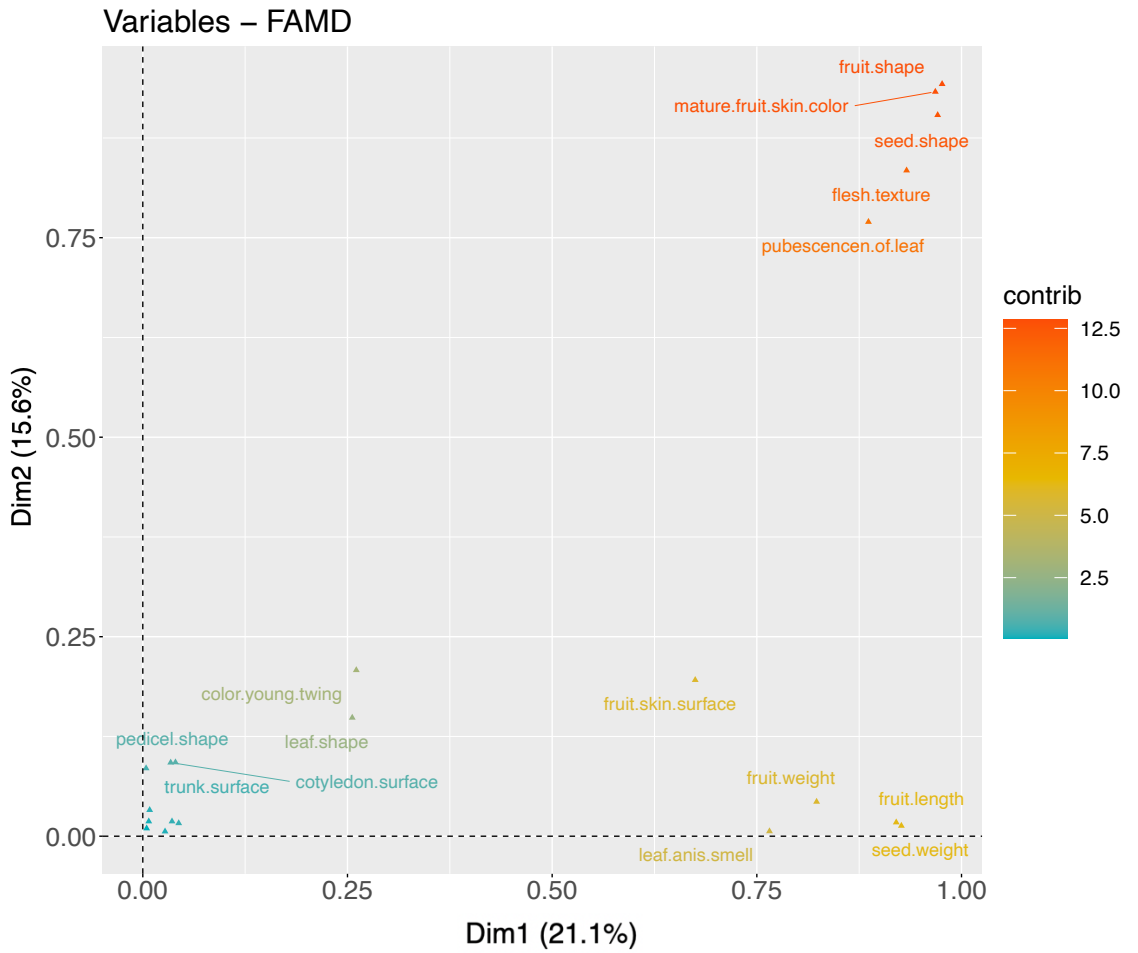


Figure 16. Association between all quantitative and qualitative variables from the FAMD analysis.

5.1.4 Population structure

The visualization of individual data points in the new feature space created by the first three dimensions, which contained the most informative information, was achieved through FAMD analysis (**Figure 17**). The FAMD individuals plot, representing the qualitative and quantitative morphological attributes of avocados, did not reveal distinct sample groupings. Instead, significant variance was observed among the examined trees at the population level, indicating a high level of phenotypic divergence within each area and an absence of differentiation among populations.

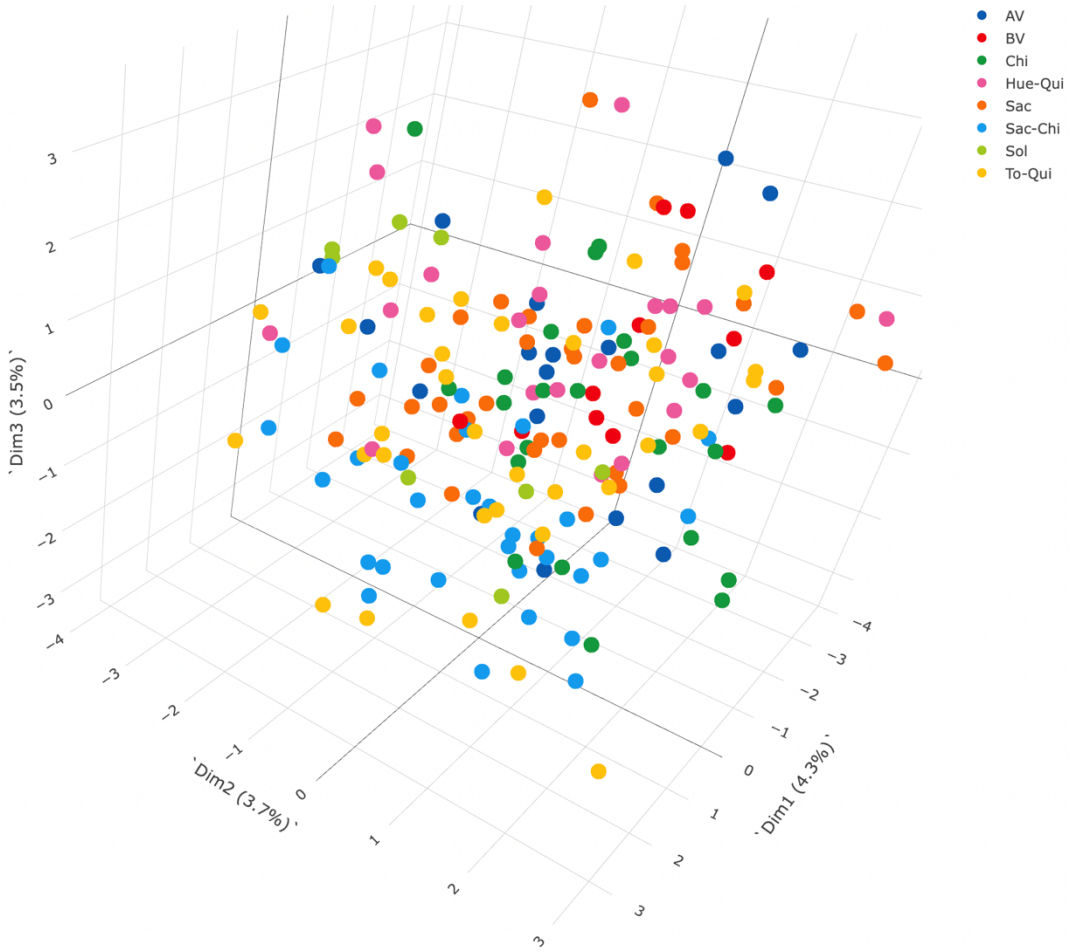


Figure 17. Avocado individuals' visualization-based in the first three dimensions based on FAMD.

5.1.5 Hierarchical Cluster of Principal Components

The hierarchical clustering algorithm on principal components (HCPC) was employed to analyze the data, resulting in a factor map and dendrogram. The HCPC analysis identified three distinct clusters among the 189 avocado trees examined. The factor map visualization (**Figure 18**) demonstrated that each cluster consisted of trees from all eight populations. Similarly, the dendrogram generated from the hierarchical cluster analysis revealed three separate groupings, represented by green, blue, and red branches, respectively. Interestingly, these groups contained samples from different populations, indicating a weak genetic structure among the populations studied. It is noteworthy that the largest group was represented by the green branches, followed by the blue branches, and finally, the smallest group was depicted by the red branches (**Figure 19**). These findings suggest a lack of distinct population differentiation and highlight the presence of genetic admixture among the avocado trees analyzed.

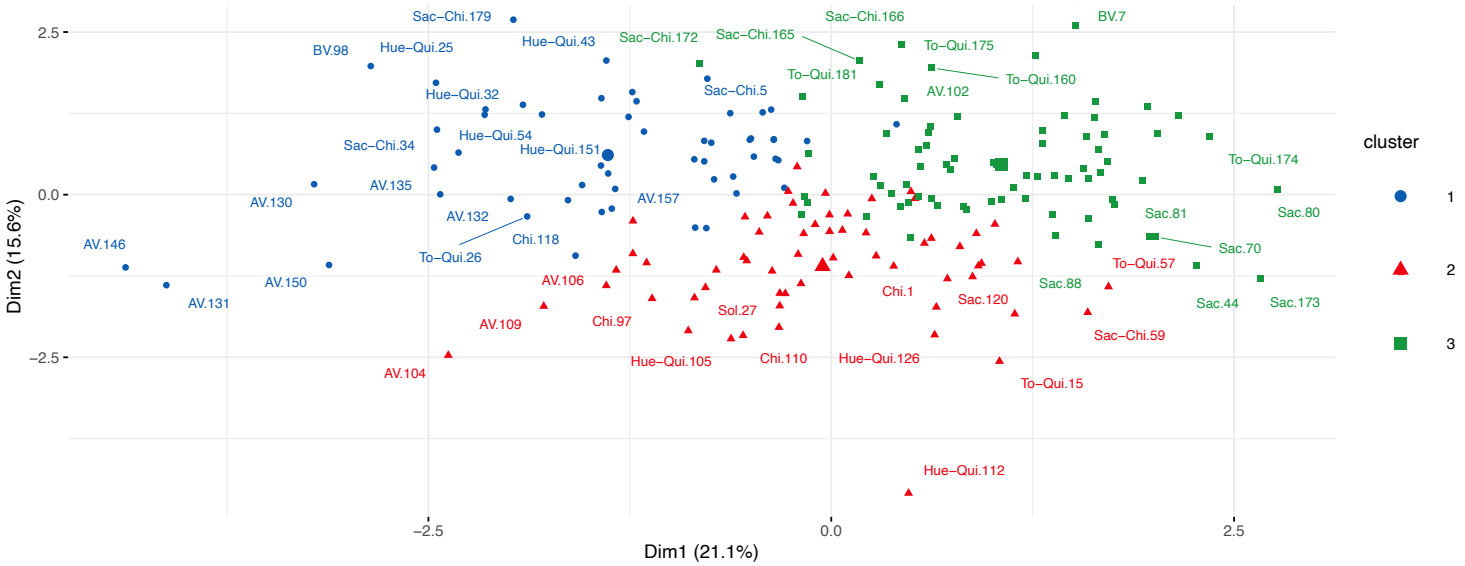


Figure 18. Avocado individuals' visualization-based in a factor map constructed from the HCPC algorithm.

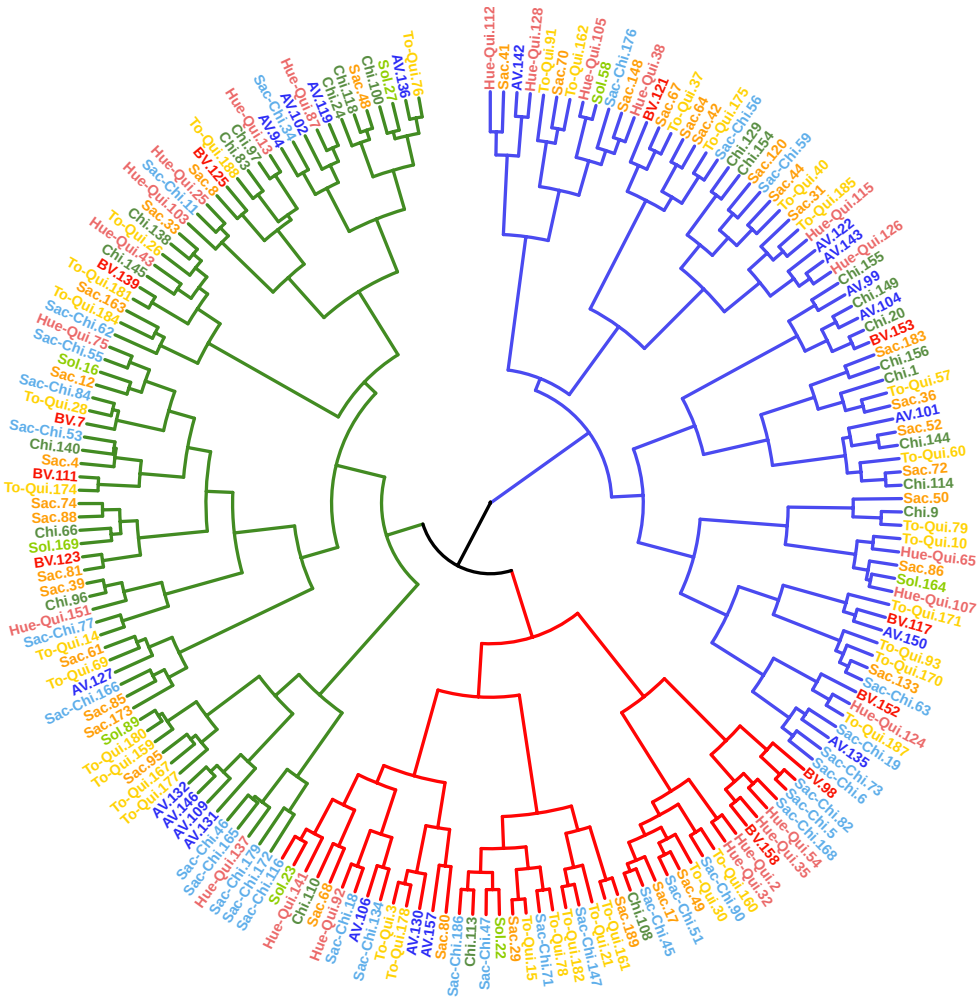


Figure 19. Avocado individuals' visualization-based in a dendrogram from the hierarchical cluster analysis.

The HCPC analysis demonstrated strong statistical associations between the clusters' partition and all quantitative characters ($p < 0.0001$), with the highest explained variances observed for fruit weight (FW), seed weight (SW), and fruit length (FL) (FW: eta2 = 0.16, $p < 0.0001$; SW: $eta^2 = 0.14$, $p < 0.0001$; FL: $eta^2 = 0.11$, $p < 0.001$). Additionally, the chi-square test revealed significant links between the clusters and the qualitative characters ($p < 0.0001$).

In the first cluster, a notable proportion of individuals exhibited a rugged fruit skin surface (90%) and a smooth cotyledon surface (61%). However, the presence of anise odor in the leaves was not highly represented within this cluster (v.test -3.51). Furthermore, the lowest seed weight (SW) was found in this cluster (v.test = -3.30) (**Table 6**). These results indicate the strong relationships between the analyzed traits and the clustering patterns, emphasizing the importance of these variables in distinguishing and characterizing the avocado populations.

Within the second cluster, a substantial proportion of individuals (92%) displayed a buttery flesh texture, while 69% exhibited anise aroma in the leaves. Additionally, 60.17% of the individuals had an even fruit skin surface. These findings indicate distinct characteristics associated with this particular cluster. Moreover, individuals within the second cluster demonstrated the lowest fruit weight (v.test -2.98) compared to the other clusters. Conversely, they exhibited significantly higher seed weight (v.test = 2.58) (**Table 6**), further distinguishing them from the rest of the clusters. These quantitative measurements underscore the unique attributes present within this cluster.

In the third cluster, a significant majority of individuals displayed distinct characteristics. Specifically, 78 % of the individuals exhibited a rounded fruit shape, 71 % had petal pubescence, and 63% showed a rough cotyledon surface. Interestingly, this cluster also had the highest values of FW and FL compared to the other clusters (**Table 6**). These findings indicate a clear morphological differentiation within this cluster, suggesting a unique genetic profile associated with these traits.

Table 6. Description of each cluster by the quantitative and qualitative character categories based on Hierarchical Clustering on Principal Components (HCPC).

Cluster 01									
Qn.T	**Mean in category**	**Overall mean**	**v.test**	**p.value**	**Ql.T**	**Cla/ Mod**	**Mod/ Cla**	**v.test**	**p.value**
FW	303.14	289.62	2.58	> 0.001	FSS rugged	90	34.18	5.9	> 0.001
FL	12.59	11.54	2.23	> 0.001	MFSC clear green	88.89	20.25	4.23	> 0.001
SW	81.09	86.55	-3.30	> 0.001	CS smooth	61.19	51.9	3.96	> 0.001
					FS pyriform	87.5	17.72	3.82	> 0.001
					AS present	9.9	39.73	-3.51	0.001
Cluster 02									
SW	90.56	86.55	2.58	> 0.001	FT buttery	91.67	30.14	5.71	> 0.001
FW	282.62	313.14	-2.98	> 0.001	AS intense	68.97	27.4	3.54	> 0.001
FL	22.66	25.65	-5.33	> 0.001	FSS even	60.17	68.49	2.72	> 0.001
					SS oblate	47.83	60.27	2.51	0.012
					LS lanceolate	14.81	5.48	-2.53	0.012
Cluster 03									
FW	313.05	289.62	3.78	> 0.001	PS rounded	77.92	32.16	5.33	> 0.001
FL	13.11	11.54	2.57	> 0.001	PP scarse	70.59	32.43	4.76	> 0.001
LL	20.84	22.66	3.24	> 0.001	FSS intermediate	37.31	67.57	4.41	> 0.001
					CS rough	62.54	27.03	3.89	> 0.001
					CYT maroon	7.46	13.51	-3.21	> 0.001

Qn.T: Quantitative trait; Ql.T: Qualitative trait; TC: trunk circumference; FW: fruit weight; FL: fruit length; LL: leaf length; SW: seed weight. The sign of the v.test indicates if the mean of the cluster is under or over-expressed for the category. FSS: fruit skin surface; MFSC: mature fruit skin color; CS: cotyledon surface; FS: fruit shape; AS: anise smell; CYT: color young twig; PP: petal pubescent; PS: pedicel shape; FT: flesh texture; SS: seed shape; LS: leaf shape; TS: trunk surface; Cla/Mod: proportion (expressed as percentages) of individuals with specific qualitative character category in the cluster; Mod/Cla: proportion (expressed as percentages) of individuals within the cluster with the specific qualitative character category.

5.2 Genetic diversity and population structure of native Guatemalan avocado using AFLP and SSRs markers

5.2.1 Population genetic diversity

The genotypic resolution of the AFLP and SSR markers are represented by the genotype accumulation curve in **Figure 20** and **Figure 21**, respectively. The genotype accumulation curve revealed that 50 AFLP and 4 SSR loci are needed to discriminate between all (100%) of individuals.

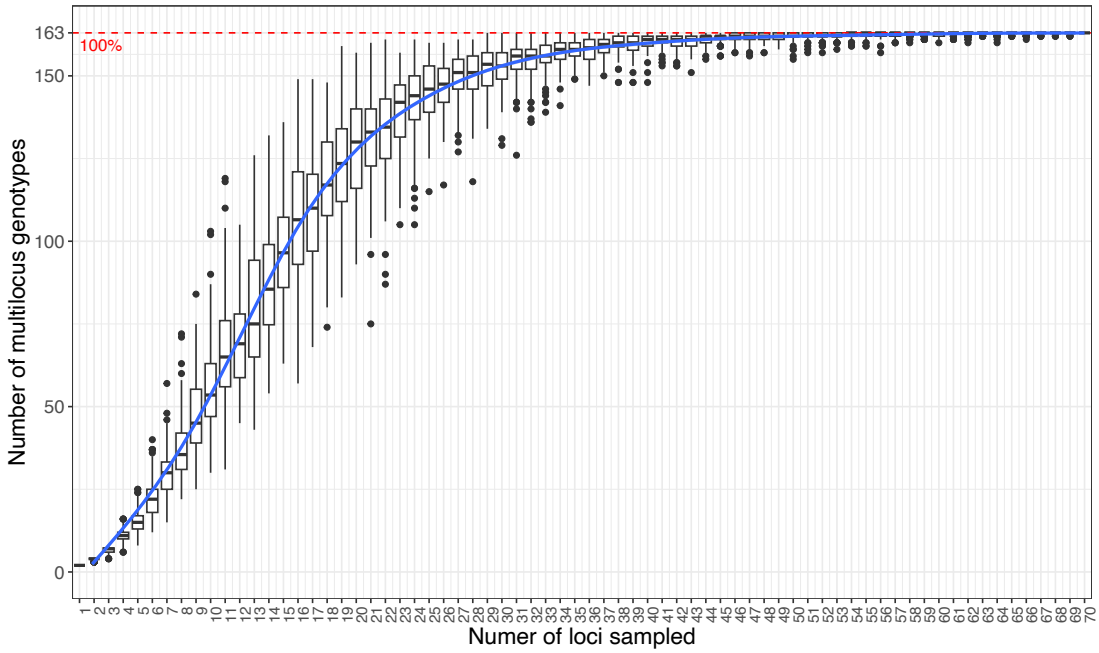


Figure 20. Genotype accumulation curve to assess avocado genotype differentiation using increasing cumulative AFLP markers.

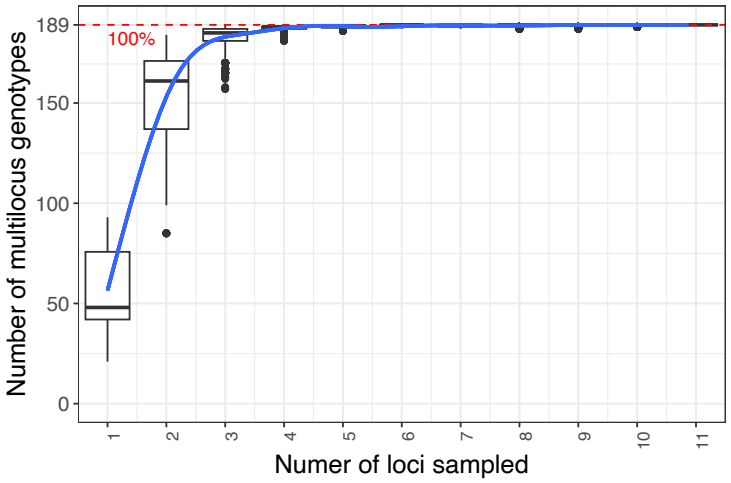


Figure 21. Genotype accumulation curve to assess avocado genotype differentiation using increasing cumulative SSR markers.

The genetic diversity within the avocado populations was assessed using several parameters. A total of 72 loci were amplified, with the total percentage of polymorphic loci reaching 98.6% (**Table 7**). The percentage of polymorphic loci (PLP) ranged from 45.8% (Sol) to 70.8% (Chi), indicating a relatively high level of genetic variation within the populations. Nei's gene diversity (Hexp) ranged from 0.135 (Sol) to 0.263 (Chi), highlighting the genetic variability present in the avocado germplasm analyzed.

Table 7. Genetic diversity measures for 8 populations of native Guatemalan avocado based on AFLP marker.

Population	n	#loc_P	PLP (%)	Hexp
Sac	36	36	50.0	0.166
Sac-Chi	32	42	58.3	0.182
Chi	16	51	70.8	0.263
Sol	6	33	45.8	0.135
To-Qui	28	43	59.7	0.185
Hue-Qui	18	39	54.2	0.201
BV	9	43	59.7	0.172
AV	18	39	54.2	0.173
Total	163	71	98.6	0.185

n: number of individuals; #loc_P: number of polymorphic loci; PLP: percentage of polymorphic loci; Hexp: Nei's gene diversity (expected heterozygosity).

Table 8 presents the characterization of 12 SSR loci in avocado from eight populations in Guatemala. The observed number of alleles (Na) ranged from 9 (AVT436) to 32 (AUCR418), while the allelic richness (Ar) ranged from 3.93 (AVAG07) to 8.66 (AVD001), indicating a diverse range of alleles across the loci. Expected heterozygosity (He) ranged from 0.60 to 0.92, indicating moderate to high levels of genetic diversity within the populations. The inbreeding coefficient (G_{IS}) ranged from 0.15 to 0.51, indicating varying levels of inbreeding within the populations. The gene flow (Nm) ranged from 1.74 to 124.75, suggesting varying degrees of gene flow among the populations. Most of the loci showed significance in the Hardy-Weinberg equilibrium test (**Figure A1**). Overall, the SSR analysis revealed high genetic diversity within the avocado populations studied, with varying levels of genetic differentiation and gene flow among the populations.

Table 8. Characterization of 12 SSR 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.

The overall gene diversity was moderate (Ht = 0.1933), with most diversity found within populations (Hw = 0.1872). Genetic differentiation among populations was low (Hb = 0.0061), indicating some genetic similarity. The Wright's fixation index (F_{ST}) indicated a small but significant level of population differentiation (F_{ST} = 0.0313) (**Table 9**). These findings highlight the presence of genetic variation and gene flow among avocado populations.

Table 9. Population genetic structure of 8 populations of avocado based on AFLP marker.

n	Ht	Hw	Hb	F_{ST}
7	0.1933	0.187	0.006	0.0313
S.E		0.015	0.002	0.312
Var		<0.001	<0.001	0.09

Ht: total gene diversity; Hw: mean gene diversity within populations; Hb: genetic differentiation among populations; F_{ST}: Wright's fixation index.

Table 10 provides an overview of the genetic diversity in eight avocado populations from different regions in Guatemala. The analysis revealed varying levels of genetic diversity among populations. The observed number of alleles (Na) ranged from 7.2 (BV) to 14.2 (To-Qui). The populations showed moderate to high levels of observed heterozygosity (Ho) (0.47 - 0.53) and expected heterozygosity (He) (0.78 – 0.89). The inbreeding coefficients (G_{IS}) ranged from 0.16 to 0.49, suggesting varying levels of inbreeding. The Shannon information index (I) (1.69 – 2.15) indicated high genetic diversity within populations. All populations exhibited a significant deviation from Hardy-Weinberg equilibrium. The null allele frequency ranged from 0.10 to 0.25 (**Table A2**).

Table 10. Genetic diversity of 189 avocado trees representing eight populations in Guatemala as detected by allele sizes at 12 SSR 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.

5.2.2 Population divergence

AMOVA based on AFLP markers revealed significant variation at different levels. Among regions, 1% of the total variation (Phi_{RT} = 0.01) was observed, indicating limited differentiation between regions. Among populations, 2.5% of the total variation (Phi_{PR} = 0.06) was attributed to population differences, indicating a moderate level of genetic variation among populations. The majority of the variation, 96.5% (Phi_{PT} = 0.03), was found within populations, highlighting the high level of genetic diversity within each population (**Table 11**). These results suggest that the genetic variation in avocado is primarily driven by differences within populations rather than between regions or populations.

Table 11. Analysis of molecular variance based on AFLP marker.

Source of Variation	Sum of squares	Variance component	Percentage of variation	F-stat	p-value
Among regions	19.13	0.061	1.0%	Phi_{RT} = 0.01	1.00
Among populations	93.78	0.528	2.5%	Phi_{PR} = 0.06	0.01
Within populations	1354.59	8.739	96.5%	Phi_{PT}= 0.03	0.01

The AMOVA analysis based on SSR loci revealed that most of the genetic diversity was observed within individuals (67.4%), while a considerable proportion was found within populations (30.7%). Only a small percentage of the variation was attributed to differences among populations (1.8%) (**Table 12**). Similarly, when the populations were grouped into regions, the analysis showed that the genetic diversity primarily resided within individuals (67.1%) and within populations (30.6%). Among regions, a modest proportion of the variation was observed (1.8%), and populations within regions exhibited a minimal level of differentiation (0.6%) (**Table 12**).

Table 12. Analysis of molecular variance based on SSR marker.

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	1.540	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	1.540	30.6	F_{IS} = 0.313	< 0.001
Among populations within regions	38.8	0.030	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 among the eight avocado populations was assessed using pairwise Phi_{PT} (**Figure 22**A) and $G'_{ST(Nei)}$ (**Figure 22**B) comparisons. Out of the 28 pairs of populations, only seven (AFLP) and five (SSR) did not exhibit significant differentiation ($p > 0.05$). The Phi_{PT} values ranged from 0 (Sac-Chi and Sac) to 0.06 (Hue-Qui and BV). The $G'_{ST(Nei)}$ values varied from 0 (AV and BV) to 0.06 (Sac and BV), indicating varying degrees of genetic differentiation. Additionally, the Mantel test demonstrated a significant correlation between genetic and geographical distance for AFLP ($r = 0.115$, $p < 0.05$) and SSR data ($r = 0.420$, $p < 0.05$) (**Figure 22**C-D).

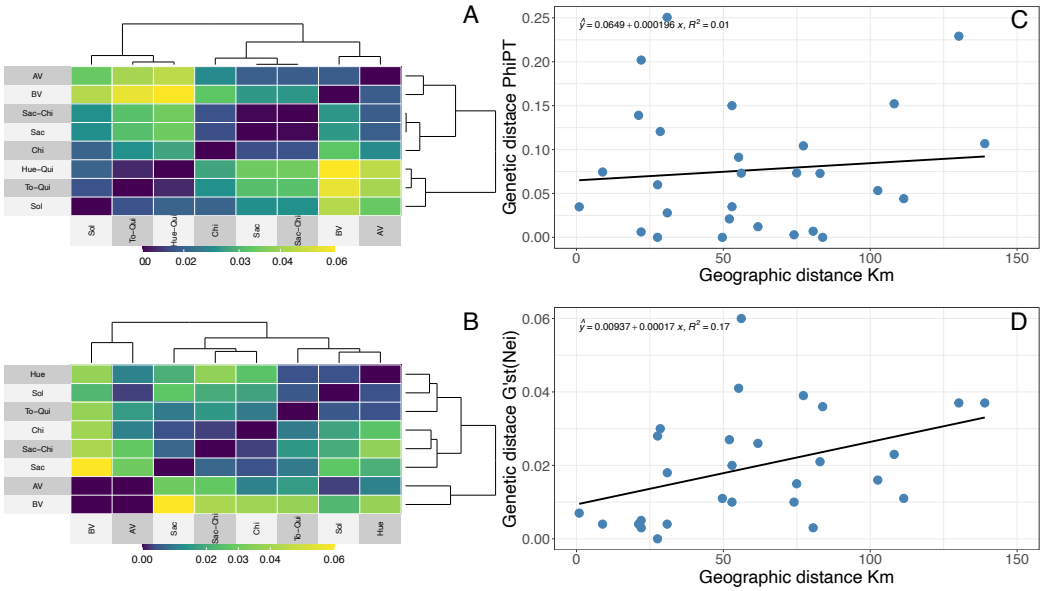


Figure 22. Pairwise heatmap and dendrogram based on A) Phi_{PT} values from AFLP B) $G'_{ST(Nei)}$ from SSR loci loci among the eight sampling sites. The Phi_{PT} and $G'_{ST(Nei)}$ matrix is represented by the heatmap color code, which includes discrete Phi_{PT} and $G'_{ST(Nei)}$ bins ranging from low to high genetic differentiation. Mantel test demonstrating a link between geographic and genetic distances between eight sampling sites from C) AFLP loci and D) SSR loci.

5.2.3 Population genetic structure

The UPGMA-based clustering analysis exhibited three discernible clusters, with each cluster comprising samples from different populations. This finding indicates a limited genetic structure among the populations, suggesting a relatively weak genetic differentiation (**Figure 23**).

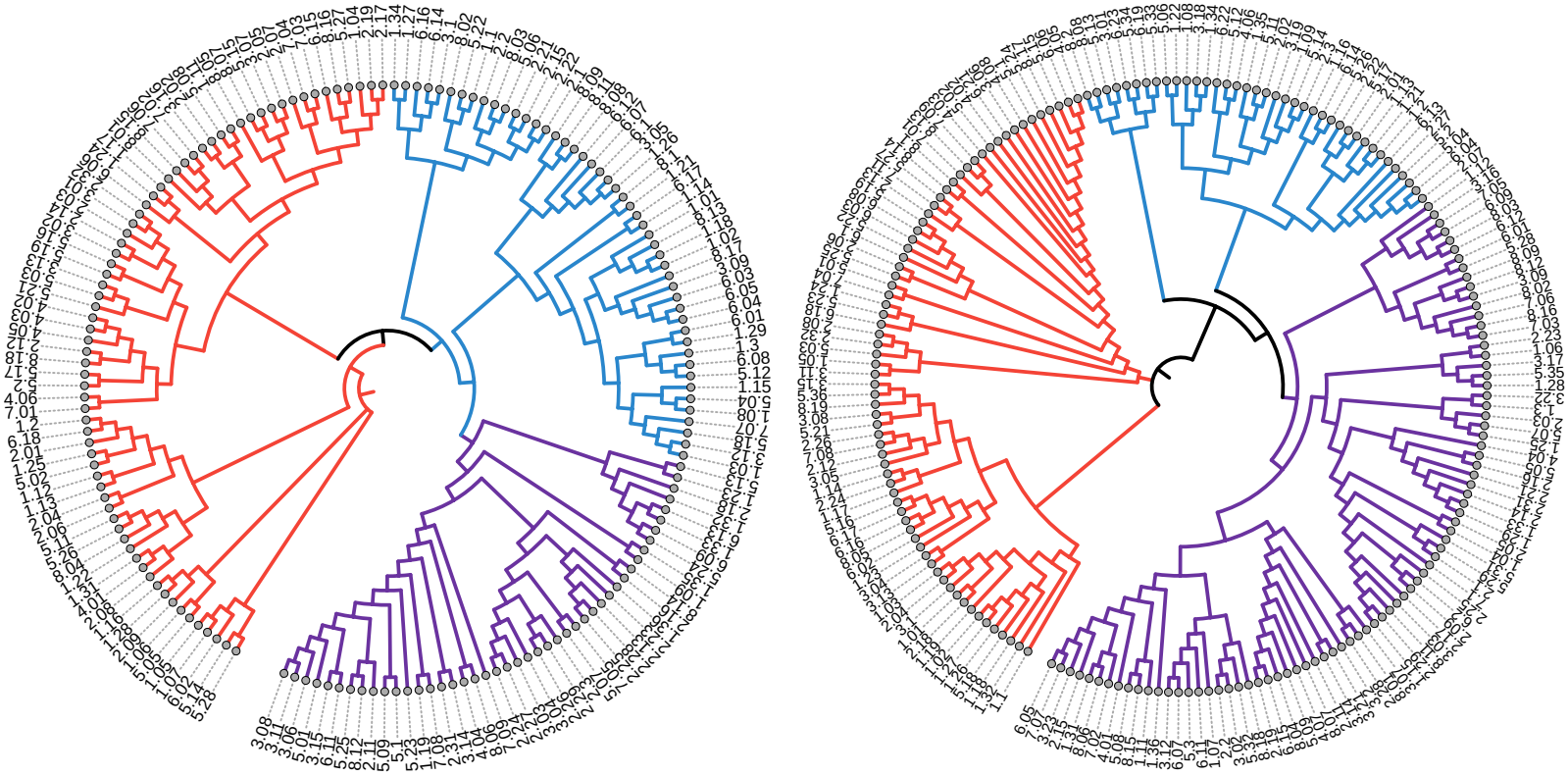


Figure 23. Dendrogram constructed with UPGMA showing the genetic relationship of the avocado trees from the eight analyzed populations in Guatemala based on a dataset of A) 72 AFLP loci and B) twelve microsatellite loci. The color-coded branches represent the three primary clusters identified on the dendrogram, indicating significant genetic admixture between samples from various populations.

The MSN was consistent with the UPGMA analysis, as it did not reveal a geographic pattern in the dispersion of individuals across the network, suggesting a weak population structure (**Figure 24**).

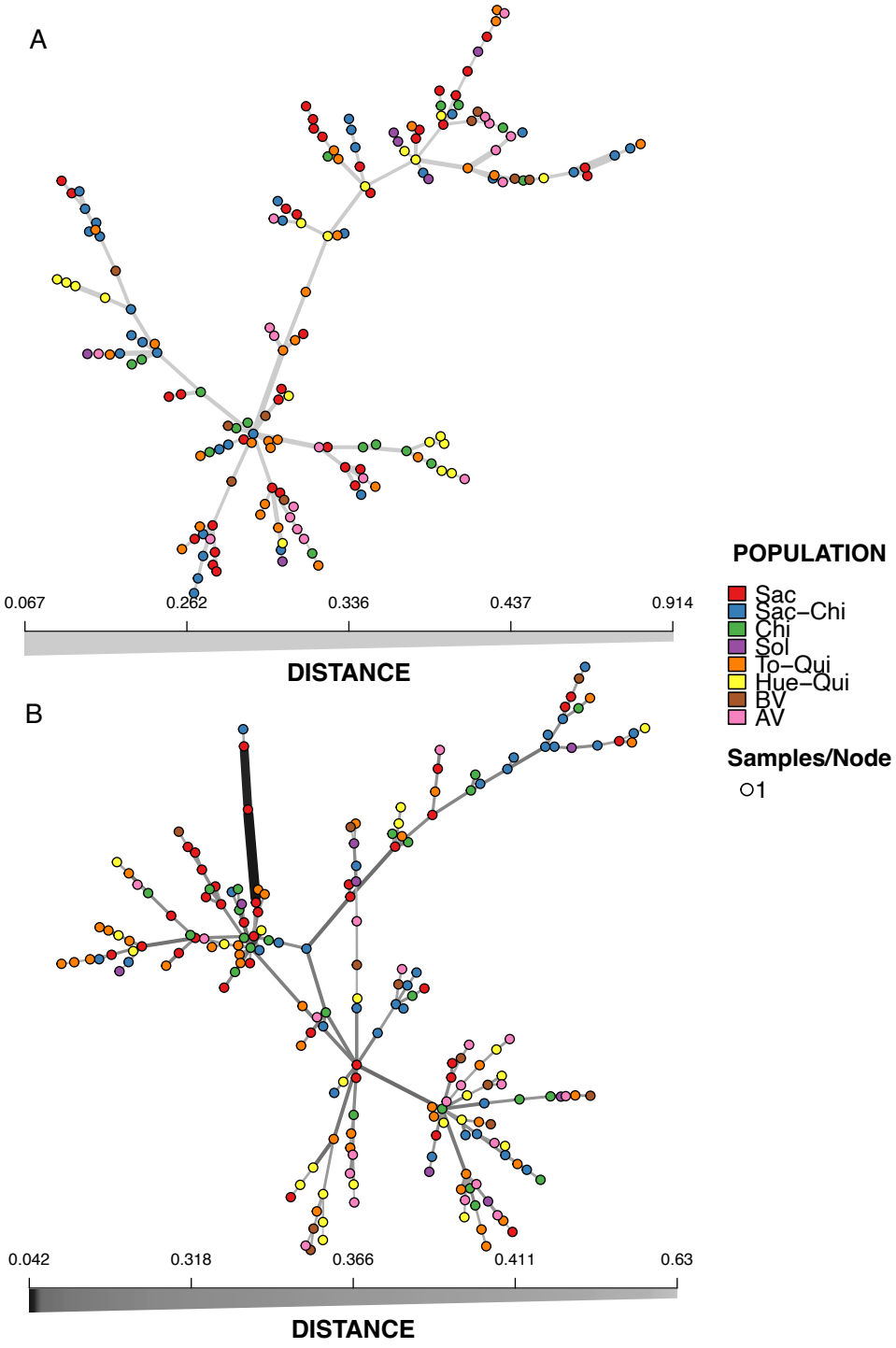


Figure 24. Minimum spanning network based on a dissimilarity matrix using A) Nei's for AFLP data and B) Bruvo's distance for SSR loci as calculated in poppr. Node colors represent population membership. Edge (line) thickness and shading represent relatedness between Multi Locus Genotypes.

The analysis of population structure using discriminant analysis of principal components (DAPC) indicated a weak population structure, as revealed by the similar structures obtained from both AFLP and SSR loci (**Figure 25**). DAPC, incorporating a priori sampling information, consistently identified the population structure, highlighting the genetic relationships among individuals. These findings suggest a lack of pronounced genetic differentiation among the studied populations, indicating potential gene flow and genetic admixture.

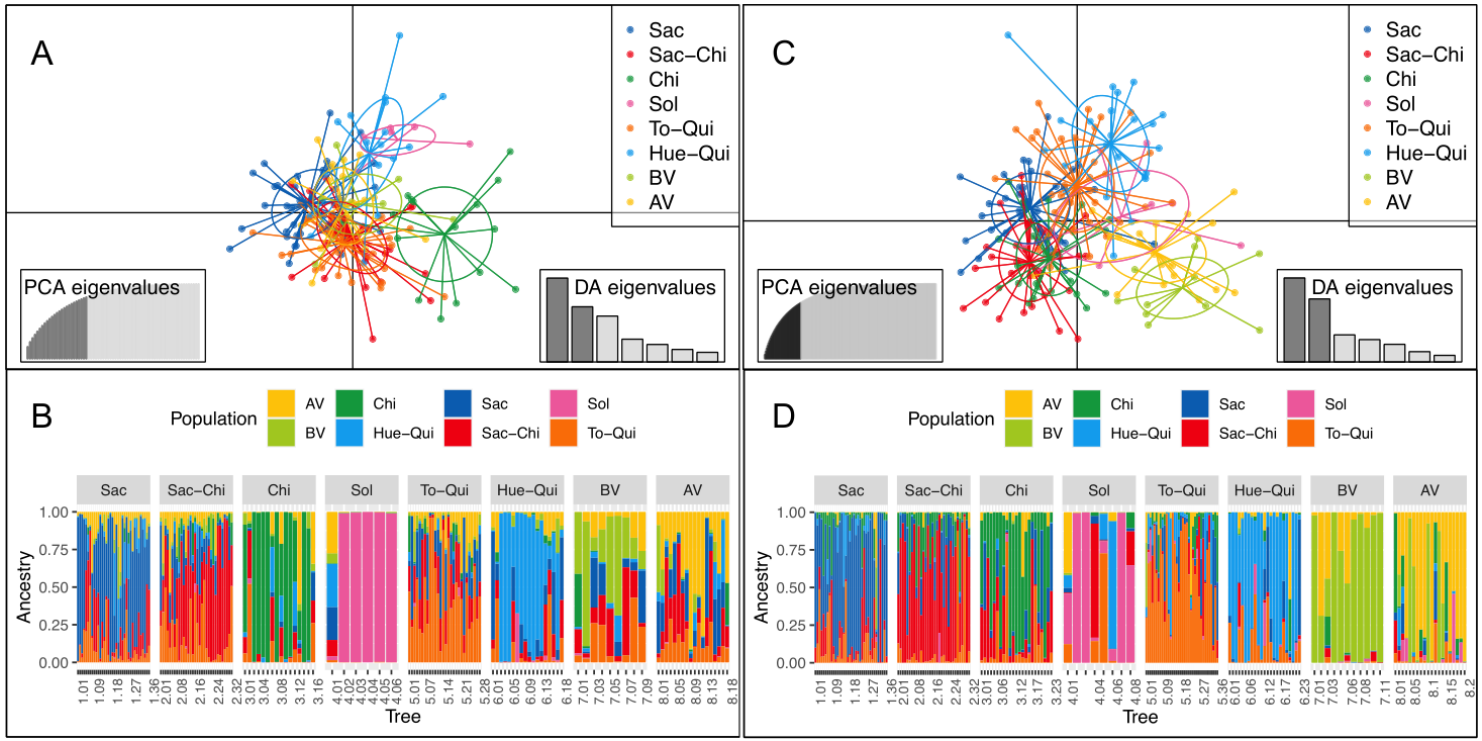


Figure 25. Population genetic structure was assessed using DAPC analysis on eight native avocado populations. AFLP loci analysis showed a scatterplot of the first two discriminant functions (A) and a bar graph representing individuals with vertical-colored lines (B). Similarly, SSR loci analysis also displayed a scatterplot (C) and a bar graph (D). Consistent color within different individuals indicated their group affiliation.

Based on the analysis using Puechmaille estimators and Evanno method obtained from StructureSelector, it was determined that the optimal number of clusters (K) among the avocado tree samples was found to be two (**Figure 26**).

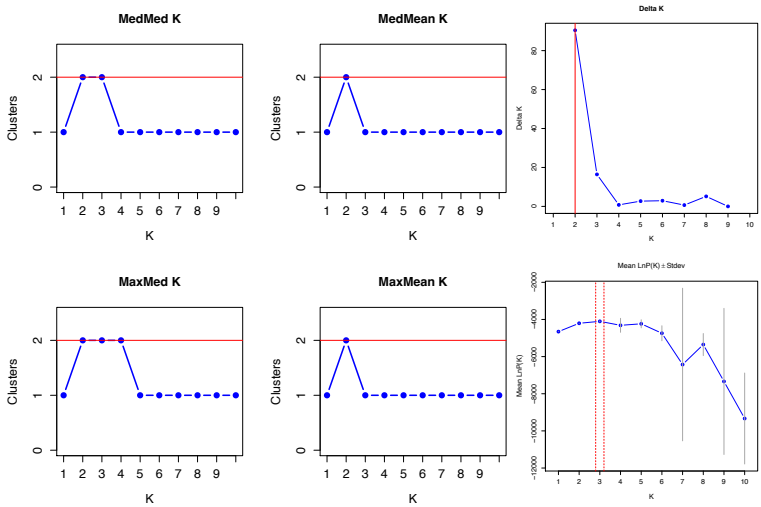


Figure 26. STRUCTURE analysis results for native avocado trees based on data from 12 microsatellites. Cluster number (K) estimation using Puechmaille estimators and Evanno method.

These findings indicate that the samples can be classified into two discrete groups or clusters (**Figure 27**). However, the analysis of AFLP loci reveals a limited population structure as no populations exhibit exclusive ancestral compositions associated with any of the identified clusters.

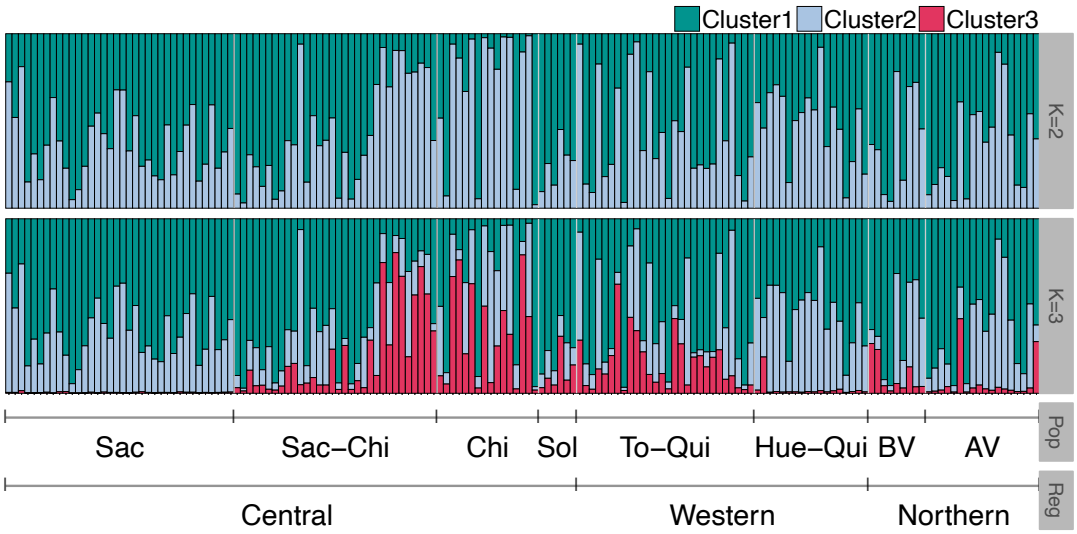


Figure 27. Estimated genetic structure of the eigth populations based on STRUCTURE analysis with cluster number (K) of two and three. Each stack bar, represents a different share of cluster per individual.

For SSR data the Puechmaille estimators and Evanno method suggested the true number of clusters was K=3 (**Figure 28**). When K = 3, the genetic structure analysis reveals a notable resemblance among the Chi, Sac-Chi, and Sac populations, as indicated by the predominant red color. Individuals from the Sol population exhibit genetic characteristics like those observed in certain individuals from other populations, such as BV and AV. The To-Qui and Hue-Qui populations display a predominant ancestry represented by the sky-blue color, with a lesser contribution from the black color (**Figure 29**).

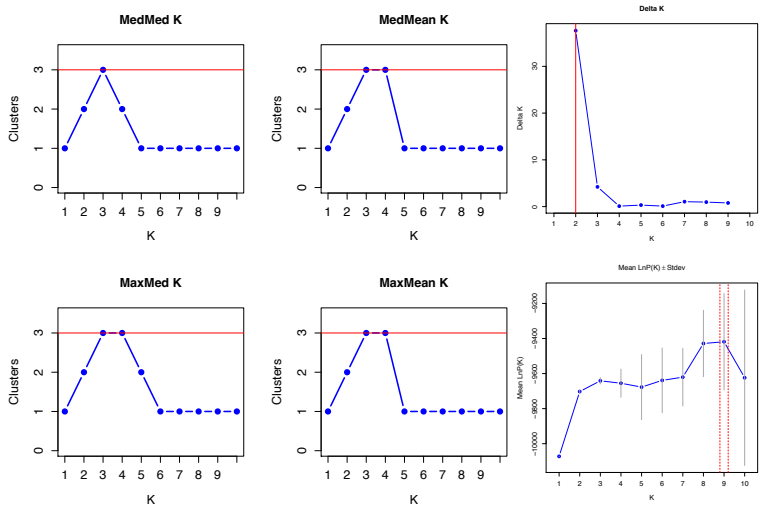


Figure 28. STRUCTURE analysis results for native avocado trees based on data from 72 AFLP loci. Cluster number (K) estimation using Puechmaille estimators and Evanno method.

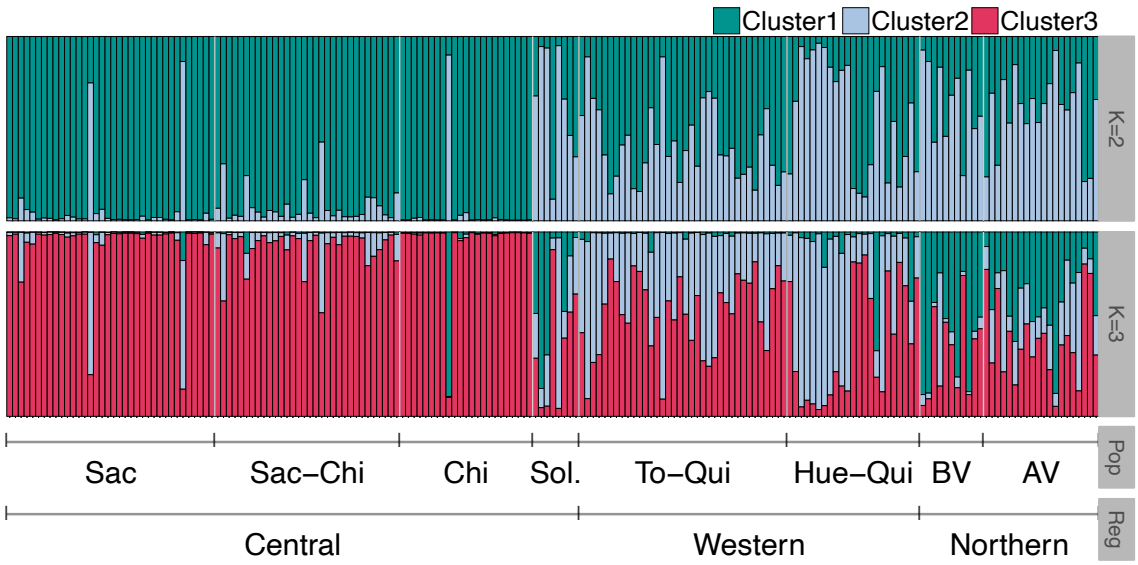


Figure 29. Estimated genetic structure of the eigth populations based on STRUCTURE analysis with cluster number (K) of two and three. Each stack bar, represents a different share of cluster per individual.

The optimal number of principal components (PC) in the PCA step of DAPC was 20 based on a-score value (**Figure A2**). The find.cluster function identified three clusters based on AFLP data, as shown in **Figure 30**A. The DAPC plot (**Figure 30**B) displayed three distinct clusters, with clusters 2 and 3 positioned to the right and cluster 1 to the left. The first discriminant functions contributed to the separation between cluster 1 from 2 and 3. Cluster 3 had the largest number of individuals (82), followed by cluster 2 (43) and cluster 1 (38). However, there was no clear separation of populations when assigning ancestry for K = 3 (**Figure 30**C). The DAPC-based clustering revealed a weak structure among the clusters, consistent with the results from STRUCTURE analysis and the UPGMA dendrogram.

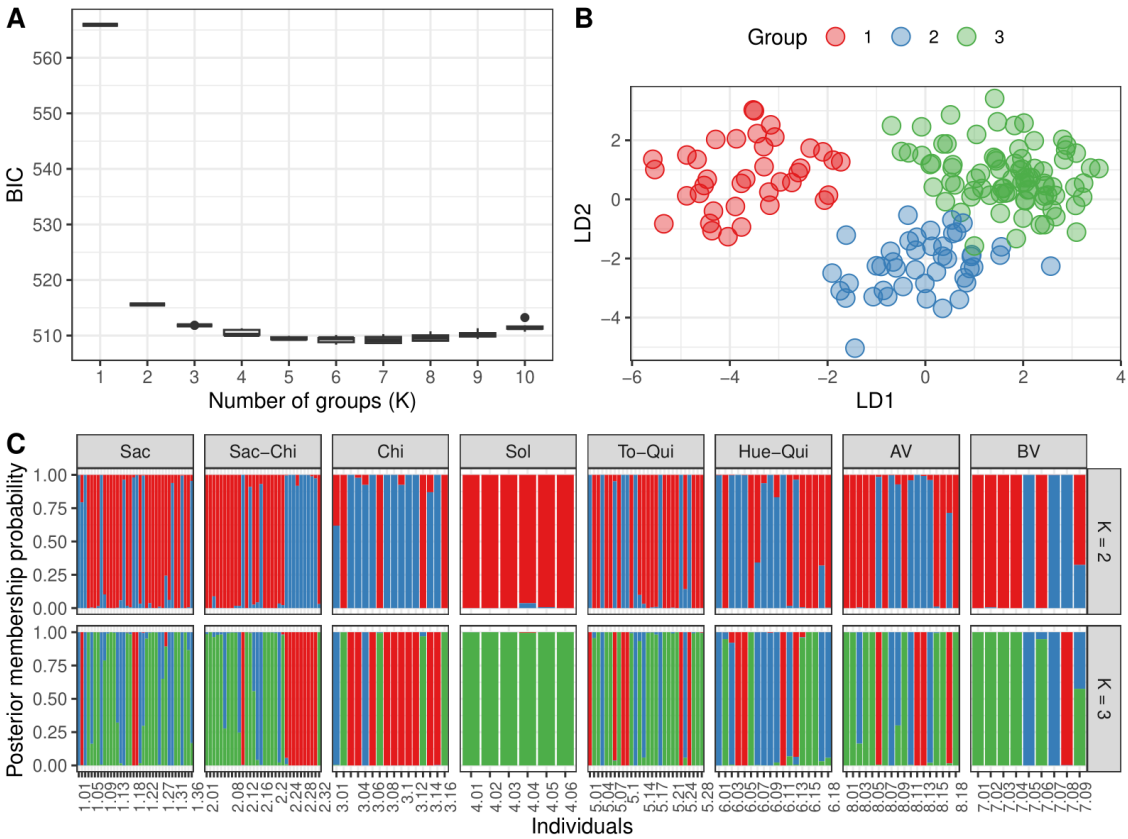


Figure 30. Population structure based on 72 AFLP loci. A) BIC to infer the most probable number of genetic groups (K = 3). B) DAPC scatterplot of the avocado trees grouped into 3 genetic groups. C) Barplot representation of the DAPC results. The probabilities of assignment to each genetic group are presented with different colors representing the genetic groups. Assignment probabilities at K = 2 and K = 3 are shown.

The optimal number PCs in the PCA step of DAPC was 21 based on a-score value (**Figure A3**). The find.cluster function applied to SSR data identified three clusters based on the BIC value (**Figure 31**A). The DAPC plot (**Figure 31**B) displayed three distinct clusters, with clusters 1 and 3 positioned to the right and cluster 2 to the left, separated by the first discriminant function. Discrimination between cluster 1 and 3 was mainly attributed to the second discriminant function. Cluster 1 comprised the largest number of individuals (76), followed by cluster 3 (59) and cluster 2 (54). When assigning ancestry for K = 3, no clear separation of populations was evident (**Figure 31**C). The DAPC-based clustering exhibited a weak structure among the clusters, consistent with the findings from STRUCTURE and the UPGMA dendrogram.

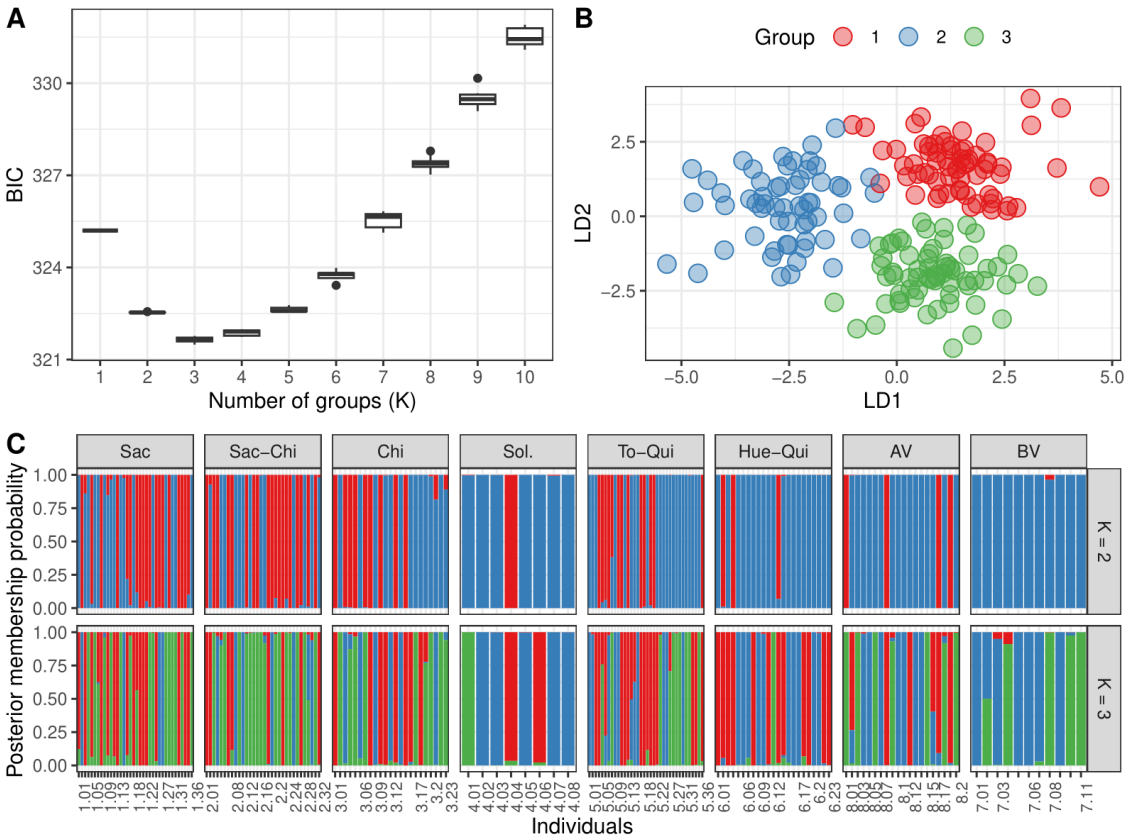


Figure 31. Population structure based on 12 SSR loci. A) BIC to infer the most probable number of genetic groups (K = 3). B) DAPC scatterplot of the avocado trees grouped into 3 genetic groups. C) Barplot representation of the DAPC results. The probabilities of assignment to each genetic group are presented with different colors representing the genetic groups. Assignment probabilities at K = 2 and K = 3 are shown.

5.2.4 Joint analysis of phenotypic and molecular data

The cophenetic correlation coefficient of 81.45% indicated a strong correspondence between the distance matrix and the dendrogram, validating the clustering of the germplasm based on phenotypic evaluations. Notably, three distinct groups were observed in the dendrogram, suggesting significant genetic differentiation among the evaluated individuals (**Figure 32**). Using SSRs markers to assess genetic diversity among native avocado genotypes, we identified the presence of three distinct groups (Figure 32). The cophenetic correlation coefficient of 92.39%, based on SSR data, confirmed the robustness and reliability of the formed clusters, highlighting the integrity of the clustering analysis.

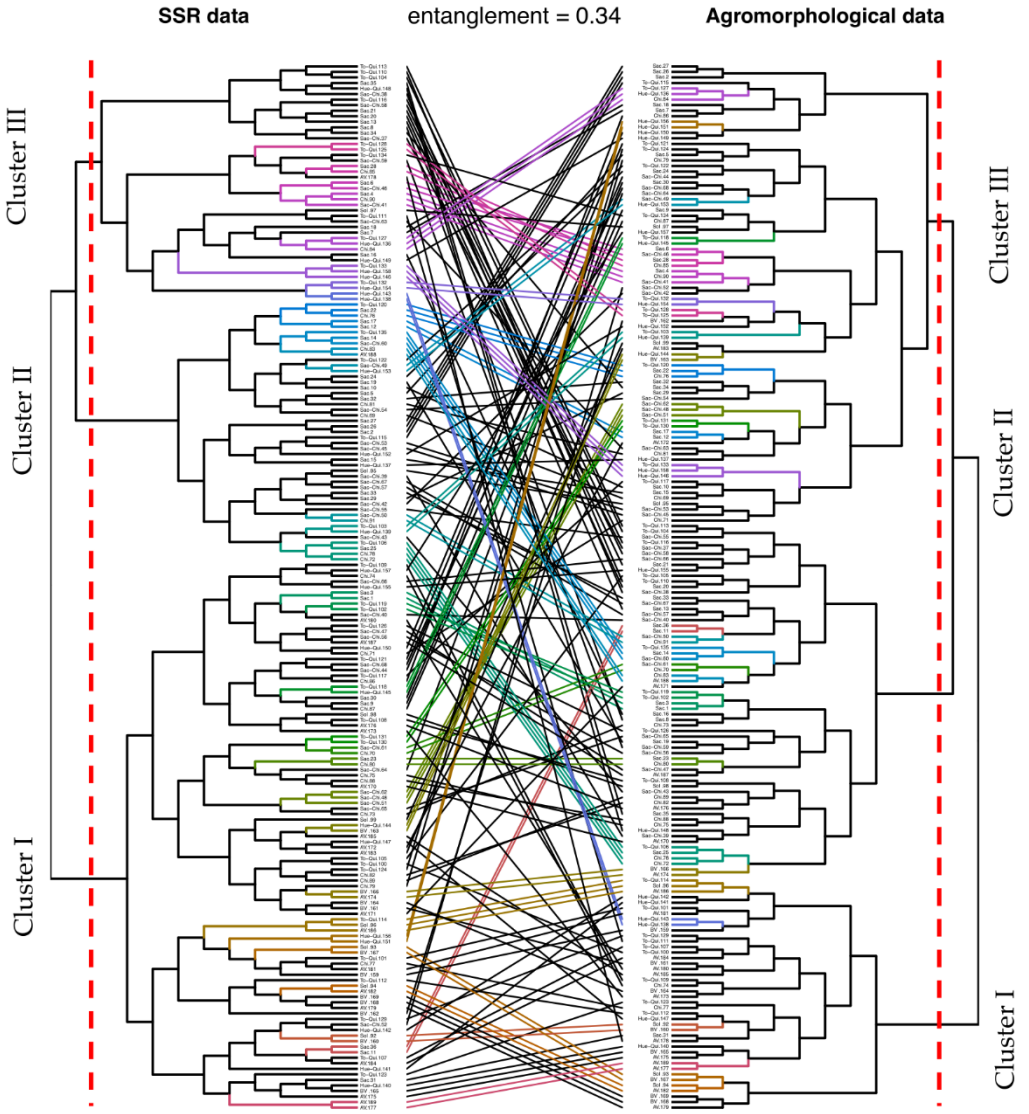


Figure 32. Tanglegram comparison of hierarchical clusters of 189 native avocado trees based on SSR (left) and phenotypic (right) data. Colored lines connect the subtrees with identical topology in both trees.

The joint matrix revealed three similarly sized clusters among the genotypes (**Figure 33**). The hierarchical dendrogram (**Figure 33**) and DAPC method (**Figure 31**) produced highly similar genotype assignments, with only 10 genotypes showing discordance between the two methods. This consistency in clustering results indicates the reliability and robustness of the analysis using the joint matrix, providing valuable insights into the genetic relationships among the genotypes.

The combination of morphological and molecular characterization yielded three distinct groups. However, when jointly analyzing the genotypes, the arrangement differed. The entanglement value of 0.34 indicated a noticeable divergence in genotype distribution between the two dendrograms (**Figure 32**), while the cophenetic coefficient was 0.65. This discrepancy suggests potential variations in the relationships among genotypes based on the different sets of data used for the analysis. Additionally, the phenotype and genotype dissimilarity matrices showed a very low correlation (r = 0.09) according to the Mantel test. In contrast, the molecular and phenotypic distance matrices each displayed strong correlations of r = 0.52 and r = 0.89 with the joint matrix, respectively. These results indicate that the relationships between phenotypic and genotypic characteristics were weak, while both genotype and phenotype were highly related to the joint analysis, suggesting a more robust association when considering both aspects together.

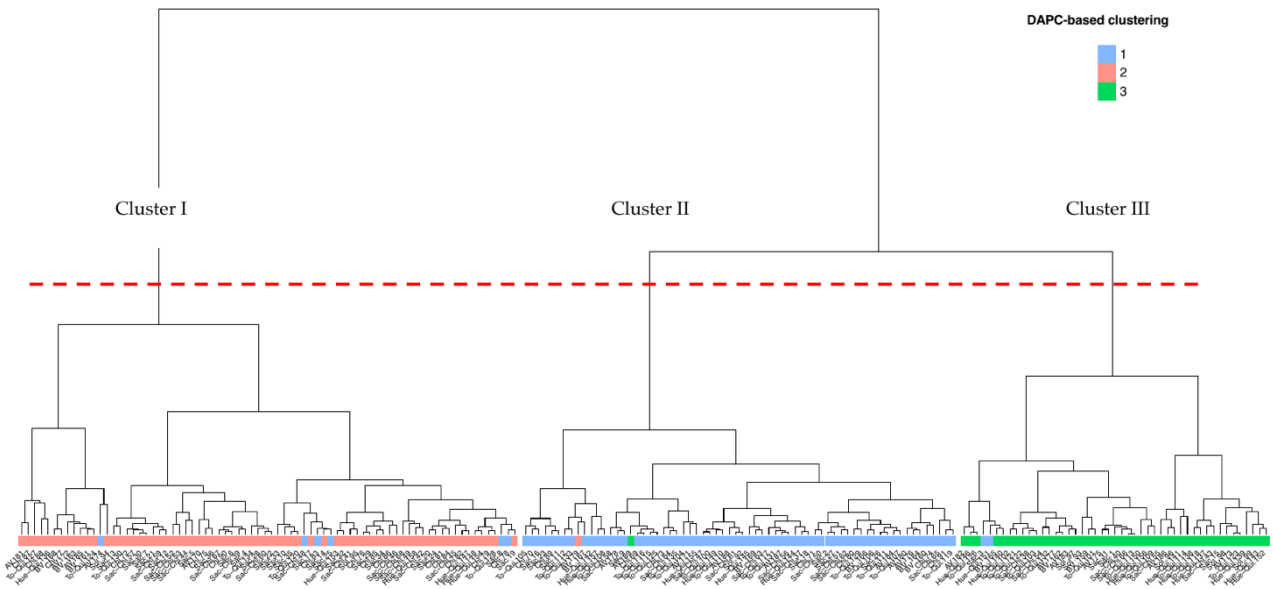


Figure 33. Hierarchical cluster analysis showing relatedness among the 189 native avocado genotypes based on phenotypic and molecular joint matrix.

5.3 Selection of a core collection of avocado genotypes for long term conservation based on agro-morphological traits and molecular markers

5.3.1 Assembly and quality evaluation of the core collections

The core collections generated by both *coreCollection* and *GeneticSubsetter* methods exhibited the lowest values of genetic distances E-NE, E-E, Shannon-Weaver diversity index (H'), as well as other indices based on mean and variance, such as MD% and VD% (**Table 13**). These results indicate that these core sets captured less diversity compared to other methods. Avocado core collection CC 03 obtained through the use of *CoreHunter* package, demonstrated the optimal values for all three genetic distances, with maximum E-NE and E-E and minimized A-NE. Additionally, CC 03 exhibited VD (96%), CR (92%), and VR (109%) values that exceeded the threshold CR (80%), and VR (100%), as well as a high H' (**Table 13**), which are essential for a robust core collection.

The inter-relationships between traits were preserved in all analyzed core sets, as indicated by the Mantel correlation, when compared to the whole collection. Additionally, CC 03 demonstrated a higher Ho value (0.576, **Table 13**) compared to the complete germplasm sample (Ho = 0.530, **Table 10**).

Table 13. Comparison of different core collections developed based on core quality evaluation indices.

Criterion	coreCollection	GeneticSubsetter	CC 01	CC 02	CC 03	CC 04	CC 05
A-NE	0.045	0.057	0.05	0.055	0.093	0.071	0.011
E-NE	0.221	0.255	0.244	0.233	0.237	0.229	0.238
E-E	0.121	0.125	0.132	0.129	0.124	0.122	0.126
MD%	46.34	55.26	37.56	25.13	22.95	37.69	22.37
VD%	75.34	63.93	82.4	63.04	95.56	77.67	91.04
CR%	84.92	72.06	78.76	69.51	92.06	89.45	85.05
VR%	104.05	115.52	93.98	109.24	108.71	117.36	101.06
H’	1.04	0.98	0.92	1.45	1.33	1.46	1.31
Mantel	0.914**	0.874**	0.793**	0.804**	0.818**	0.904**	0.748**
Ho	0.542	0.505	0.584	0.544	0.576	0.515	0.551

A-NE: the average distance between each accession and the nearest entry; E-EN: the average distance between each entry and nearest neighboring entry; E-E: the average genetic distance between entries; MD%: Mean difference percentage; VD%: variance difference percentage; CR%: coincidence rate of range; VR%: variable rate of range; H’, Shannon diversity index, Ho: observed heterozygosity ** indicates significance at $p \leq 0.01$.

This core set includes samples from the three genetic clusters revealed by the DAPC analysis (**Table A3**). Considering the various evaluation indices mentioned above, the CC 03 demonstrated the highest capture of prevalent diversity and representativeness from the entire germplasm. Therefore, CC 03 core collection was chosen for further use, as it captured the total diversity of the native Guatemalan avocado germplasm and will be used for comparative analysis with the entire germplasm collection.

5.3.2 Comparative evaluation of the core collection with the entire native Guatemalan avocado germplasm collection

The descriptive statistics, including means, ranges, coefficient of variation, interquartile range, and frequency distribution, were analyzed for various quantitative traits in both the selected core set and the entire collection (**Table 14**). The CC 03 exhibited a higher CV for all traits compared to the entire germplasm, indicating a greater capture of variability within the core set. The statistical tests, specifically the Newman-Keuls test and t-test, showed that there were non-significant differences in means between the core set and the entire collection for all the traits. Levene's test revealed significant differences in sepal length trait (SL), but non-significant differences were observed the rest of traits.

The frequency distribution plots (**Figure 34**) demonstrated the representation of all classes from the entire collection within the core set, indicating the capture of quantitative trait variability. The interquartile range was mostly similar across traits, including SW, FL, PL, LL, LW, and SL, except for FW and TC (**Table 14**), which displayed symmetrical distributions of accessions between the core collection and the entire germplasm. To assess the distribution patterns of the eight qualitative traits, QQ plots and Kullback-Leibler distance calculations were performed for both the core set and the entire collection (**Figure A4**). The Kullback distances (**Figure A4**), ranging from 0.04 to 0.08 for all traits, indicated that the distribution of traits in the core collections was identical to that of the entire collection.

Table 14. Comparison of range, mean, coefficient of variation, interquartile range, and frequency distribution in the entire germplasm and core collection for various quantitative descriptors used in the formation of the core.

Trait	Entire germplasm					Core collection					Comparative statistics			
	Min	Max	**Mean ± SE**	CV	IQR	Min	Max	**Mean ± SE**	CV	IQR	$\bar{x}^a$	$\bar{x}^b$	V^c	F^d
FW	44.39	584.16	313.14±7.23	32. 45	125.13	108.63	517.43	326.39±14.08	34.61	141.13	ns	ns	ns	ns
SW	38.23	136.37	86.55±1.31	21.89	25.88	48.55	136.37	90.09±2.80	22.89	27.26	ns	ns	ns	ns
FL	3.46	18	11.54±0.2	22.12	4.3	3.46	16.63	11.12±0.44	26.13	3.73	ns	ns	ns	ns
PL	2.51	4.3	3.47±0.02	10.53	0.44	2.76	4.27	3.45±0.05	12.54	0.44	ns	ns	ns	ns
LL	5.22	36.91	22.66±0.44	27.56	8.52	11.71	32.38	22.91±0.86	26.97	7.99	ns	ns	ns	ns
LW	3.61	20.93	12.80±0.26	28.78	4.96	5.68	20.24	12.81±0.48	29.14	3.62	ns	ns	ns	ns
SL	1.16	5.21	3.51±0.06	21.34	1.06	2.02	4.69	3.70±0.09	25.56	0.6	ns	ns	**	ns
TC	22.93	147.74	104.37±1.68	22.89	27.93	49.22	142.42	103.47±3.13	23.29	31.15	ns	ns	ns	ns

FW: fruit weight; SW: seed weight; FL: fruit length; PL: pedicel length; LL: leaf length; LW: leaf width; SL: sepal length; TC: trunk circumference; CV: coefficient of variation; IQR: interquartile range. ns indicates not significant an*, ** and *** indicate $p \leq 0.05$, 0.01 and 0.001, respectively.

$\bar{x}^a$ Differences between means of entire collection and core set were tested by Newman–Keuls test.

$\bar{x}^b$ Differences between means of entire collection and core set were tested by t-test.

V^c Variance homogeneity as tested by Levene's test.

F^d Difference of frequency distribution by Wicoxon rank test.

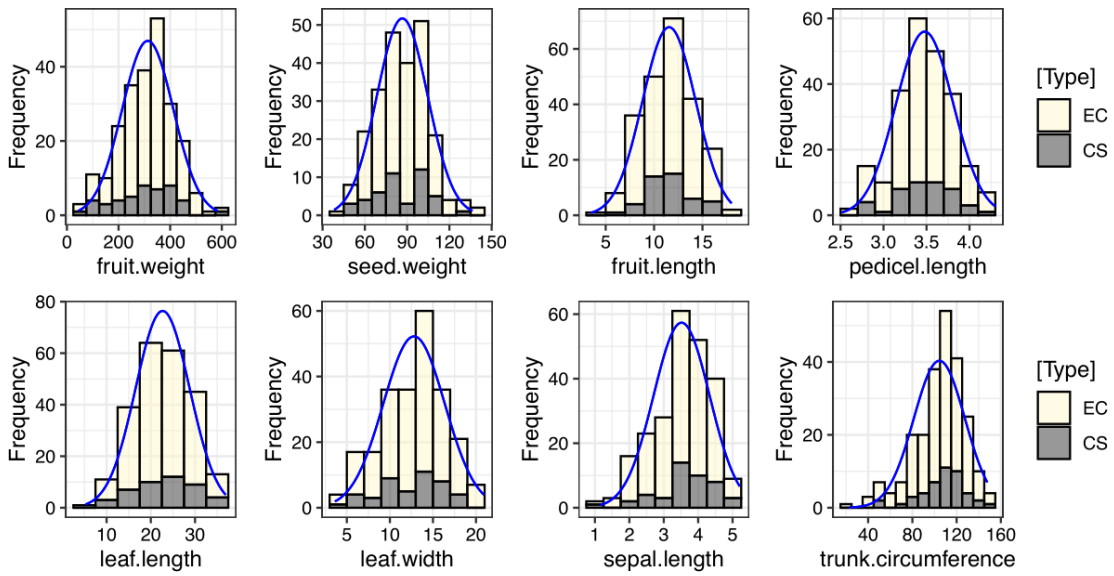


Figure 34. Frequency distribution plots showing the comparison of variability of quantitative traits in the entire germplasm (EC) and core collection (CS) of avocado.

The results of the calculation of the H' and Evenness for qualitative or categorical data in the entire germplasm and the core collection indicate that the extracted core sets successfully maximized the existing diversity, as evidenced by the increased values of H'

for all traits, except for a minimal difference observed in FS and SS (**Table 15**). It is worth noting that both FS and SS already exhibit maximum diversity in both the entire collection (2.18 and 2.06, respectively) and core collection (2.17 and 2.03, respectively), which are very close to the maximum possible values (H' max) of FS (2.20) and SS (2.08). The evenness value ranged from 0.88 to 1.

Table 15. Shannon diversity index of qualitative traits in the entire germplasm and core collections of native Guatemalan avocado.

Descriptor	Shannon-Weaver diversity index (H')		H' max		Evenness	
	Entire germplasm	Core collection	Entire germplasm	Core collection	Entire germplasm	Core collection
TS	1.00	1.16	1.10	1.10	0.81	0.88
CYT	1.48	1.55	1.61	1.61	0.88	0.96
CML	0.61	0.69	0.69	0.69	0.95	0.99
LS	2.02	2.17	2.20	2.20	0.91	0.96
AS	0.63	0.69	0.69	0.69	0.90	1.00
PP	0.93	1.10	1.10	1.10	0.93	1.00
PS	0.99	1.08	1.10	1.10	0.90	0.98
FSS	0.54	0.65	0.69	0.69	0.92	0.94
MFSC	1.75	1.94	1.95	1.95	0.96	1.00
FS	2.18	2.17	2.2	2.20	0.99	0.97
FT	1.38	1.48	1.39	1.39	0.95	0.98
SS	2.06	2.03	2.08	2.08	0.99	0.97
CS	1.09	1.10	1.10	1.10	0.99	0.99

TS: trunk surface; CYT: color young twig; CML: color mature leaf; LS: leaf shape; AS: leaf anise smell; PP: petal pubescent; PS: pedicel shape; FSS: fruit skin surface; MFSC: mature fruit skin color; FS: fruit shape; FT: flesh texture; SS: seed shape; CS: cotyledon surface.

The analysis of trait associations revealed significant and positive correlations among various traits. In the entire collection, a strong correlation was observed between SW and FW (r = 0.49), and between FW and TC (r = 0.24). Similarly, in the CC 03, with r-values of 0.42 for SW and FW, and 0.37 for FW and TC (**Figure 35**). Among all possible pairwise comparisons (r-values) between the eight quantitative traits, five correlations were significant in the entire collection, while six correlations remained significant in the core collection.

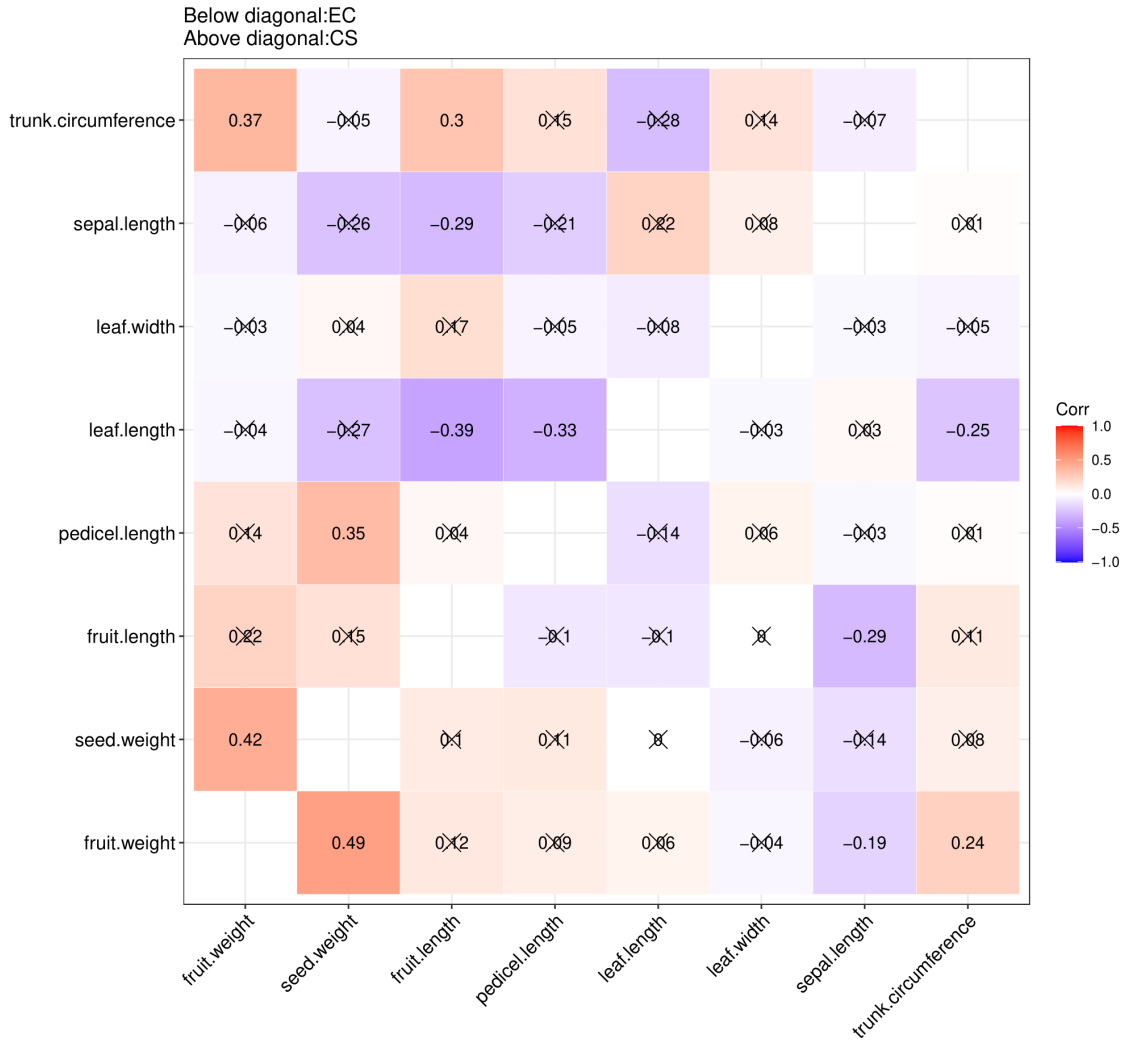


Figure 35. Correlogram of quantitative variables.

PCA was conducted based on the correlation between the eight quantitative traits to explore the spatial distribution of entries/samples in both the core collection and the entire germplasm of avocado. The first five principal components (PCs) accounted for a significant portion of the variance, explaining 77.9% of the variance in the core collection and 77.5% in the entire collection (**Table 16**).

Table 16. Comparison of first five principal components in entire germplasm and core collections of native Guatemalan avocado.

Collection	Statistics	Principal components				
		PC1	PC2	PC3	PC4	PC5
Entire germplasm	Standard deviation	1.334	1.111	1.092	1.043	0.949
	Proportion of variance	0.222	0.154	0.149	0.136	0.113
	Cumulative proportion	0.222	0.377	0.526	0.662	0.775
Core Collection	Standard deviation	1.364	1.108	1.103	1.023	0.938
	Proportion of variance	0.232	0.154	0.152	0.131	0.110
	Cumulative proportion	0.232	0.386	0.538	0.669	0.779

5.4 Comparison of the genetic diversity of Guatemalan and Ethiopian avocado germplasm

5.4.1 SSR polymorphism and genetic diversity

The variability of the SSR marker is presented in **Table 17**. A total of 352 alleles were detected with the 12 SSRs. The number of alleles per locus ranged from 5 (AVT 436) to 13 (AVAG 13) with an average of 10.2 alleles per locus. The PIC ranged from 0.71 to 0.91. A high level of polymorphism was obtained in most of the loci studied, since eight of the 12 loci revealed 10 or more alleles in the landraces studied. Consequently, all the SSR markers selected are highly informative with He= 0.72 and PIC = 0.83 on average except the locus AVT436 which shows the minimum number of alleles (5) and average PIC (0.79).

The observed heterozygosity ranged from 0.20 (AVAG25) to 0.70 (AVAG13). The overall mean He value was 0.52. The expected heterozygosity ranged from 0.59 (AVAG07) to 0.84 (AVD001) with a mean value of 0.72. The comparison between the two parameters was carried out based he Wright's fixation index (F). For all loci this parameter was positive, meaning a deficit of heterozygotes, with an overall average value of 0.2. SSR marker AVAG25 has the highest F value (0.68), and AVAG22 (0.05) with the lowest F value showing an overall heterozygote deficiency.

Significant departure over Hardy-Weinberg expectations ($p < 0.001$) was observed for 12 loci. The value of Ne ranges from 2.6 (AVAG07) to 8.2 (AVD001) with an average of 5.0. Shannon's Information Index (I) was high (2.12) and low (1.2) for AVD001 and AVAG07, respectively. The measurement of genetic diversity of the SSR

marker is very high (0.93) and low (0.74) for AVD001, and AVAG07, respectively. There is a significant difference in Ar among the 12 SSR markers (**Table 17**). The value ranges from 4.8 (AVT436) to 10.1 (AVD001) with an average of 7.9.

Table 17. Amplification performance of 298 avocado landraces collected from seven regions of Ethiopia (n = 109) and Guatemala (n = 189) as revealed by SSR markers.

SSR	A	Ne	I	Ho	He	F	PIC	Ar	G	HWE
AVAG05	10.00	3.61	1.47	0.42	0.62	0.30	0.83	6.86	0.85	***
AVAG11	10.14	3.50	1.53	0.35	0.68	0.49	0.74	6.97	0.77	***
AVAG13	13.43	7.11	2.08	0.70	0.83	0.14	0.92	9.87	0.93	***
AVT436	5.14	3.20	1.22	0.52	0.65	0.16	0.79	4.79	0.81	***
AUCR418	13.71	6.10	2.02	0.57	0.82	0.30	0.90	9.13	0.9	***
AVAG07	7.57	2.63	1.20	0.48	0.59	0.18	0.71	4.97	0.74	***
AVAG21	8.43	4.64	1.68	0.61	0.78	0.21	0.88	6.61	0.89	***
AVAG22	12.14	6.37	1.70	0.59	0.65	0.05	0.84	9.27	0.85	***
AVAG25	7.57	4.22	1.40	0.23	0.63	0.68	0.79	6.55	0.8	***
AVD001	13.57	8.17	2.12	0.58	0.84	0.32	0.92	10.11	0.93	***
AVD022	9.57	4.37	1.65	0.56	0.75	0.23	0.86	6.97	0.87	***
AVMIX04	11.29	6.33	1.80	0.61	0.75	0.20	0.88	9.04	0.89	***
Average	10.214	5.023	1.655	0.519	0.716	0.271	0.84±9	7.59	0.85	***

List Locus name, number of alleles (A), the effective number of alleles (Ne), Shannon information index (I), observed (Ho) and expected (He) heterozygosity, Wright's fixation index (F), *P* test for H-W equilibrium (HWE), Ar: average allelic richness and the Polymorphism information content (PIC) calculated for 12 SSRs markers in 298 avocado Landraces. ***Significant at $p < 0.01$.

5.4.2 AMOVA analysis

AMOVA indicated that the variation within individual landraces accounted for the highest variation (56%) followed by the variation between individual samples within regions (23%). The continent variation accounted for 18% of the total variance. The variation between region is low (2%). The permutation analyses result confirmed that all three levels significantly contributed to the overall genetic variation. The degree of population differentiation among landraces within the continent ($F_{SR} = 0.184$). The population differentiation among regions was small ($F_{SR} = 0.03$) but it was statistically different from zero ($p < 0.0001$) indicating the presence of intermediate genetic differentiation. This indicated that based seed movement contributed intermediate

(N_m = 0.95) number of migrants per generation between landraces leading to high genetic differentiation (**Table 18**).

Table 18. AMOVA analysis of genetic variances within and between avocado populations created by STRUCTURE software, and DAPC (the level of significance is based on 10,000 permutations).

Source of Variation	Sum of squares	Variance component	Percentage of variation	F-stat	p-value
Among Continent	310.05	1.05	18	F_{RC} = 0.18	< 0.001
Among Regions	82.65	0.14	2	F_{SR} = 0.03	< 0.001
Among Individuals in the region	1701.49	1.33	23	F_{IS} = 0.29	< 0.001
Within Individual	949.00	3.19	56	F_{IT} = 0.44	< 0.001
Total	3043.18	5.71	100	F_{ST} = 0.21 Nm =0.95	

F_{RC} = among continent; F_{SR} = among regions; F_{IS} = among populations within region; F_{IT}= within individual; Nm = Gene flow

5.4.3 Population stratification

The genetic variation found within the populations is high (**Table 19**). Shannon index (I) value of 2.30 for the western region, and 2.07 for the Central region. The Gamo (1.04), Sidama (1.14), and Wolaita (1.16), a region with low I value show less diversity compared to that of the more than one ancestral population of Wondo Genet with the I value of 1.87. The number of Na was 18.08, 16.25, 12.33, 8.08, 6.50, 5.50 and 4.75, for the western, central, northern, Wondo Genet, Wolaita, Sidama, and Gamo, respectively (**Table 19**).

Likewise, significant Ne was exhibited among the regions, where the western region has the highest effective number of alleles (8.16) while Gamo was with the least number of effective allele values (2.69). The largest number of privet alleles was found in the western region (3.42) while the lowest value of the number of private alleles is found in Sidama (0.42). Similarly, the expected heterozygosity is high for the western region (0.85) followed by the northern region (0.82) in Guatemala. The minimum value of He (0.55) value is found among the avocado landrace collected in the Gamo region (**Table 19**).

Table 19. Parameters of the genetic diversity of avocados based on the population derived from STRUCTURE.

Population	Na	Ne	I	#Private alleles	He
Central	16.25	5.91	2.01	2.92	0.80
Northern	12.33	7.08	2.07	1.67	0.82
Western	18.08	8.16	2.30	3.42	0.85
Gamo	4.75	2.69	1.04	0.50	0.55
SD	5.50	2.74	1.14	0.42	0.59
WG	8.08	5.80	1.87	1.08	0.81
WLT	6.50	2.77	1.16	0.67	0.59

Na: observed number of alleles; Ne: effective number of alleles; I: Shannon information index; He: expected heterozygosity. SD: Sidama, WLT: Wolaita, WG: Wondo Genet.

5.4.4 Genetic differentiation and population structure

The pairwise F_{ST} (genetic differentiation coefficient values) are shown in **Table 20**. The F_{ST} values between every two populations ranged from 0.01 to 0.27, and the genetic differentiation between every two populations reached a significant level ($p < 0.01$). The highest F_{ST} value was observed between the Northern and Gamo populations (0.27), while the lowest value was found for the Central and West populations (0.01) and the Wolaita and Sidama populations (0.01). Central, western, and northern populations showed fewer differentiation among the Wondo Genet population (0.06, 0.07, and 0.04) showed low-level differentiation compared to the other three populations of Ethiopia.

Table 20. Genetic differentiation coefficient (F_{ST}) and Gene flow (Nm) for seven avocado collecting regions. F_{ST} below and Nm above the diagonal.

	Central	Northern	Western	Gamo	SD	WG	WLT
Central	----	7.09	17.69	0.78	0.85	4.01	0.84
Northern	0.03	----	9.45	0.69	0.76	3.30	0.75
Western	0.01	0.03	----	0.95	1.05	6.22	1.04
Gamo	0.24	0.27	0.21	----	69.98	1.33	10421.78
SD	0.23	0.25	0.19	0.00	----	1.56	37.80
WG	0.06	0.07	0.04	0.16	0.14	----	1.43
WLT	0.23	0.25	0.19	0.00	0.01	0.15	----

SD: Sidama, WLT: Wolaita, WG: Wondo Genet, and Gamo from Ethiopia, whereas central, northern, and western indicate the regions from Guatemala.

These results suggested that all populations could be divided into two groups, central, western, and northern could be classified into one group, while the remaining three populations were assigned to the other group. Consistent with the results of genetic differentiation, seven populations could be clustered into three groups according to pairwise Nm values. The gene flow between Gamo, Wolaita, and Sidama is higher (**Table 20**, Nm < 1.05) and was much lower than that between gene flow between central, northern, and western avocado native trees in Guatemala. However, (Nm > 3.3) is present between the central, northern, western, and Wondo Genet avocado trees.

Consistent with the results of genetic differentiation, the central, western, and northern populations showed large genetic distances and low genetic identities when compared to the other four populations. As a result, all seven populations are grouped into two main genealogical branches in **Figure 36**. Western and central populations converged first, then gathered with Northern populations and separated from other populations (SD, Gamo, WLT and WG). The remaining four populations formed the other branch. Among them, WLT, Gamo, and SD clustered as a group. In the other group, the WG population was separated, which represented a separate relationship between the three.

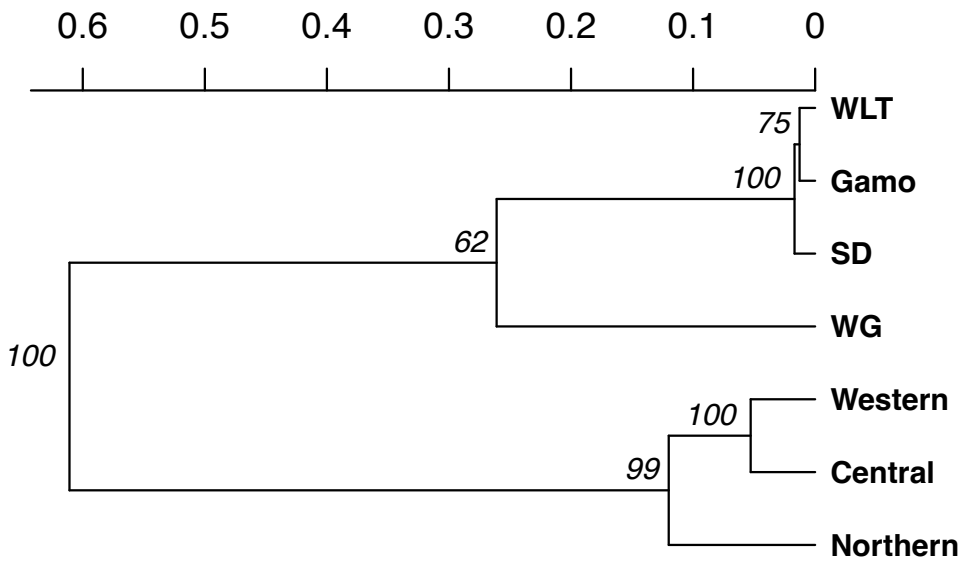


Figure 36. UPGMA clustering landraces of the avocado population from seven regions in two continents based on Nei genetic distance. SD: Sidama, WLT: Wolaita, WG: Wondo Genet, and Gamo from Ethiopia, whereas central, northern, and western indicate the regions from Guatemala. Values at the nodes represent the statistical bootstrap support of 1,000 iterations.

5.4.5 Similarity among the avocado genotypes

According to the SSR analysis, all individuals had genetic similarity coefficients ranging from 1 to 40. The multilocus SSR was found to be able to distinguish populations of avocado genotypes from two continents and seven different geographical areas according to the assignment tests and the UPGMA tree constructed from the matrix of pairwise allele-sharing distance among 298 individuals (**Figure 37**). The phylogeny based on twelve SSR markers demonstrated a distinct difference between avocado from Guatemala and Ethiopia. The Sidama, Gamo, and Wolaita landraces of avocado are distantly related to the Central landraces. However, the western and northern landraces overlap significantly with the Wondo Genet landraces.

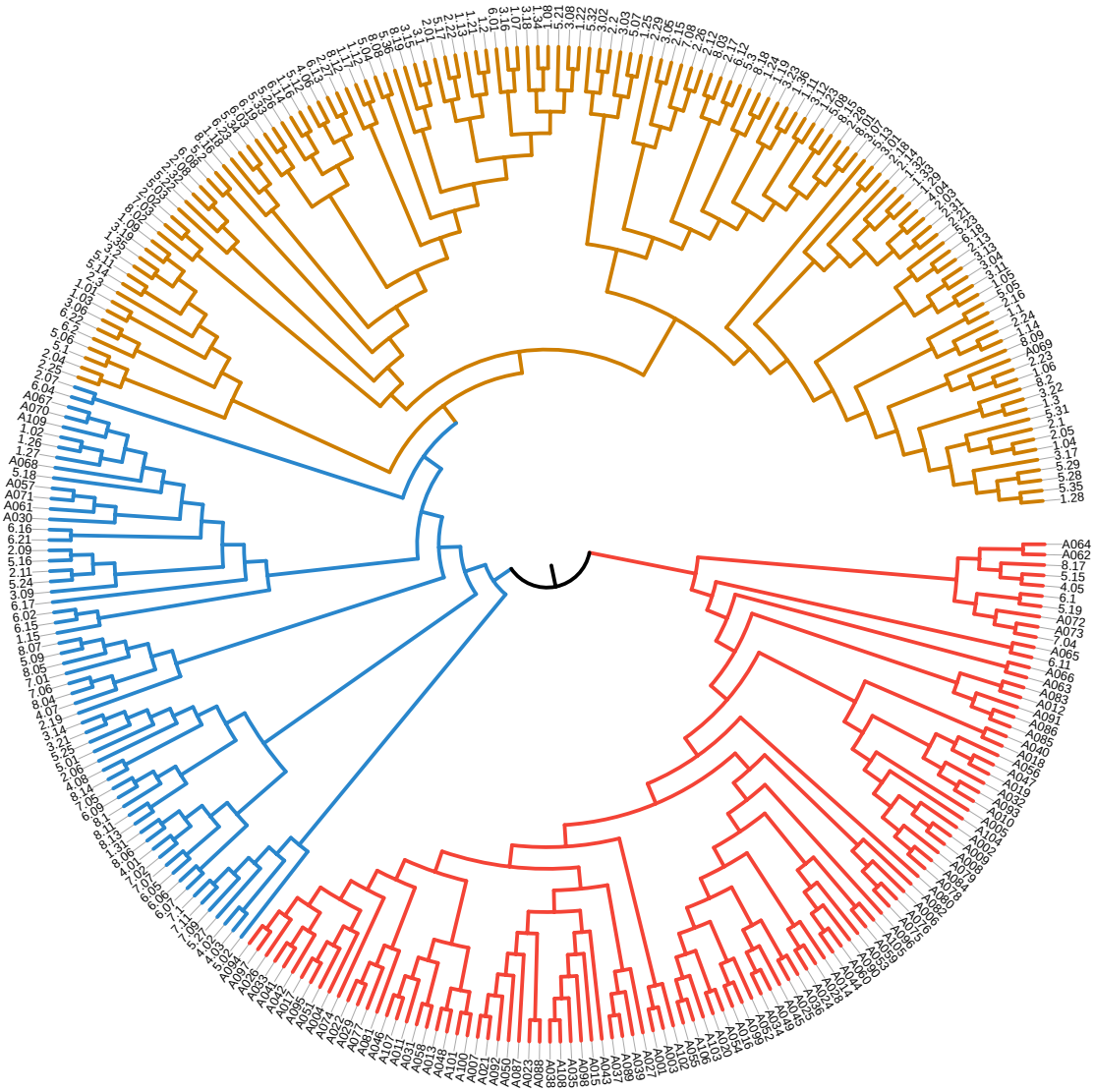


Figure 37. Dendrogram of 298 avocado genotypes based on UPGMA analysis using the similarity matrix generated by the Nei coefficient.

5.4.6 Population structure analysis

The model-based analysis of the population structure along with the Delta K method (Evanno et al. 2005) revealed the presence of three genetically distinct clusters (K = 3) corresponding to landraces of the north, central-west, Gamo, Wolaita, Sidama, and WG (**Figure A6**). With the arbitrary cut-off value of 80% ancestry for assignment, 49 landraces (16.4%) were attributed to cluster one (central, western, and South region) and 119 landraces (39.9%) to the second cluster (central, northern, western, WG, and Wolaita region). The third cluster is exclusively composed of 92 (30.9%) landraces from Sidama, Gamo, and Wolaita.

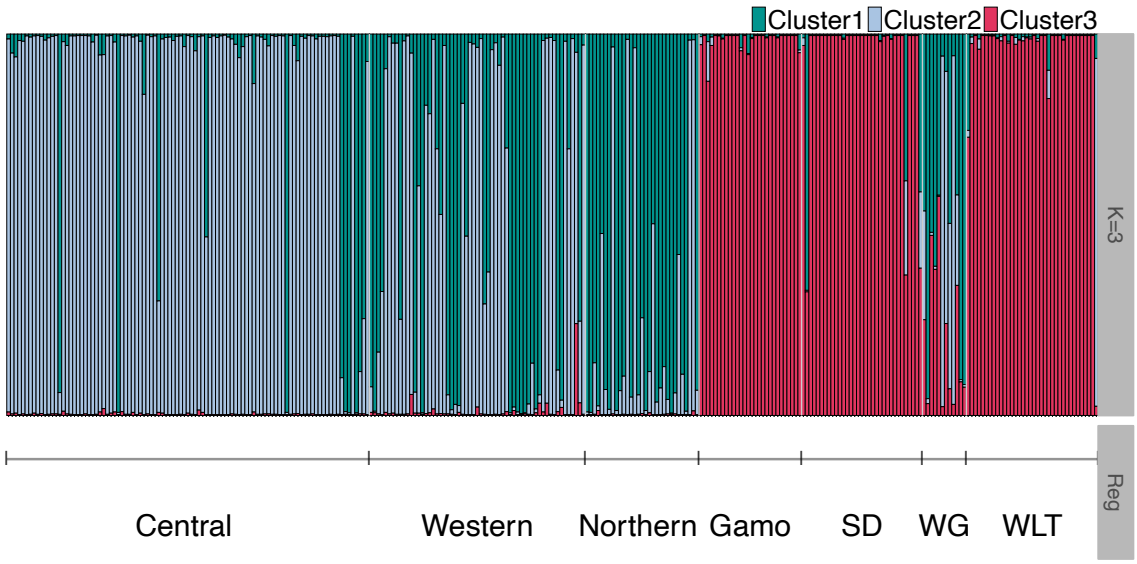


Figure 38. Population structure based on Bayesian analysis. Three subpopulations, or cluster, are indicated by color; sub-population one (red) accessions from central, northern, and western Guatemala; subpopulation two (green) predominately central, western, northern Guatemala, Wondo Genet (WG). Sub-population three (sky-blue) accessions exclusively from Sidama (SD), Wolaita (WLT), and Gamo. More than one ancestry is indicated by the presence of mixed color. The population names were given below the box plot with the individuals of different populations separated by vertical black lines. Each color represents one cluster.

However, 38 landraces (12.7%) were attributed to appeared to have ancestry from more than one cluster, having Q ancestry values of less than 80 %. Mixed-ancestry plants included 5, 7, 3, 15, 7, and 5 genotypes from central, northern, Sidama, western, Wondo Genet, and Wolaita, regions, respectively.

DAPC also obtained similar genetic structure among individuals from seven populations. The find cluster retained 21 PCs to apply the K-means algorithm and search the data pattern more thoroughly. Applying the elbow method, the optimal cluster number is 3 (**Figure 39**A). These findings are in line with STRUCTURE results. The three Guatemalan populations are grouped, and the same situation for the Ethiopian populations, while WG is like a transition group with mixed genetic information (**Figure 39**B). In the scatter plot (**Figure 39**C), groups 1, 2, and 3 include 83, 121, and 94 individuals, respectively.

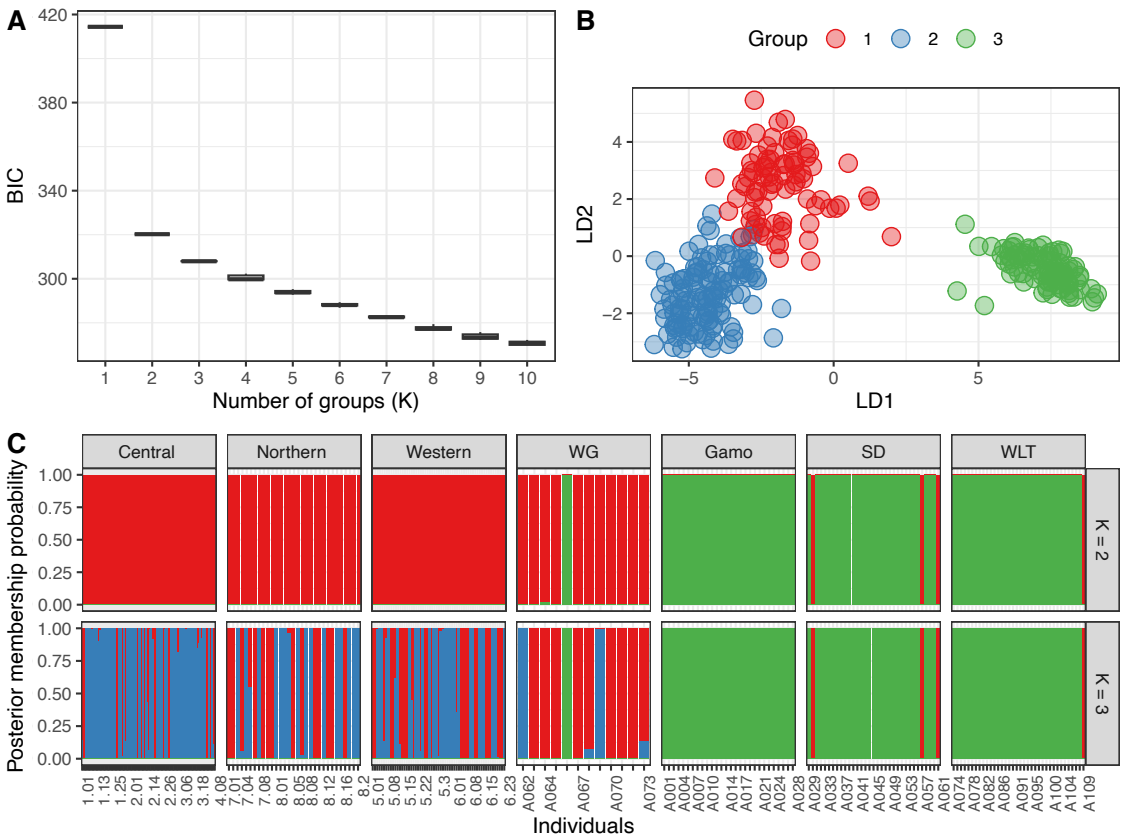


Figure 39. A) Using BIC, the most likely number of genetic groups is assumed to be three. B) A DAPC scatterplot of the 298 Avocado individuals divided into three genetic subgroups. C) A bar plot showing the DAPC outcomes. Different colors represent the various genetic groups are used to represent the probabilities of assignment to each genetic group. The posterior probability of assignments for K = 2 and K = 3 are displayed. WG: Wondo Genet; SD: Sidama; WLT: Wolaita.

5.4.7 Principal coordinates analysis

The first two principal components contributed 21% of the total variation, according to principal coordinate analysis. Variation was accounted for by 17.4% the first principal component (PC1), and 3.6% by PC 2 (**Figure 40**). This analysis reinforced the separation of Guatemala and Ethiopia on the right and left sides, respectively. Whereas the admixed genotypes appeared to be clustered along the centers of the two principal components.

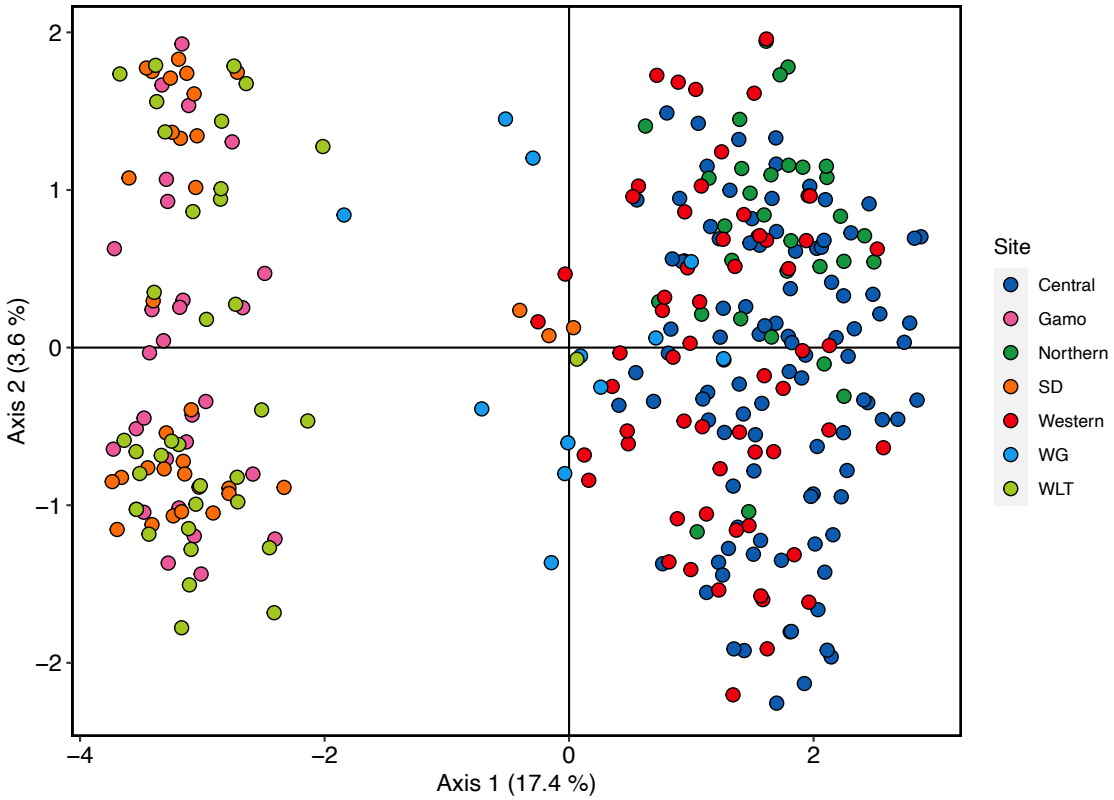


Figure 40. Principal coordinate analysis (PCoA) based on F_{ST} pairwise comparison showing the relationships among the Guatemalan and Ethiopian regions. The first principal coordinate explained 17.4% of the variation. The second principal coordinate explained 3.6% of the variation. WG: Wondo Genet; SD: Sidama; WLT: Wolaita.

6 Discussion

6.1 Variability analysis of native Guatemalan avocado germplasm based on agro-morphological traits

6.1.1 Quantitative traits

The assessed quantitative morphological traits revealed significant diversity in all three clusters, with more than 20% CV observed for 87.5% of the descriptors taken into consideration. A larger proportion for a property may suggest greater variability (Hidalgo 2003). This high level of variation suggests that each cluster possesses unique morphological characteristics, contributing to the overall diversity of native avocados. These findings are consistent with previous research conducted on Mexican avocado germplasm (Rincón-Hernández et al. 2011; López-Guzmán et al. 2015), further validating the importance of understanding and preserving the genetic diversity of native avocado populations. Based on Tukey's test, To-Qui and Sac populations exhibited the highest average values for fruit weight and length, while the lowest values were found in AV (**Table 5**). This finding has important implications for avocado breeding, variety selection, and horticulture. It allows targeted breeding efforts to enhance desirable traits, informs market segmentation based on fruit size, and guides orchard management for optimized yields. Additionally, preserving diverse genetic clusters supports avocado germplasm conservation and resilience. Overall, this finding contributes to sustainable and profitable avocado production, meeting industry demands.

The significant variation in FW among the avocado populations can be attributed to multiple factors. Firstly, the different genetic basis of each population contributes to the diversity in fruit size and weight. Different genetic backgrounds may result in variations in fruit development and maturation processes (Chen et al. 2007; Henao-Rojas et al. 2019; Cañas-Gutiérrez et al. 2022). Moreover, microenvironments and agro-ecological circumstances play a crucial role in shaping fruit characteristics. Variation in soil types, climate conditions, and management practices in different regions can influence the availability of nutrients, water, and other resources, affecting fruit development and size. The observed strong correlation between fruit weight and length further supports the notion that fruit weight can serve as a reliable indicator for yield estimation and monitoring changes in avocado production (Mokria et al. 2022). As FW and length are closely related, changes in one trait are likely to be reflected in the other, making it easier to estimate fruit yield before harvest.

These findings have practical implications for avocado growers and breeders. Understanding the sources of variability in FW can aid breeders in developing varieties with desirable fruit size and weight. Additionally, for growers, monitoring fruit weight can help optimize harvesting practices and manage orchards more effectively to achieve higher yields.

6.1.2 Qualitative traits

The prevalence of rough and very rough trunk surfaces among the avocado trees in this study is closely aligned with the characteristics of the study area. It is noteworthy that most of the sampled trees were situated in medium or highlands, while the number of trees in lowlands was relatively low. This observation correlates with previous studies indicating that the bark texture of Guatemalan and Mexican avocado races, primarily found at elevations above 1,500 masl, tends to be less rough. In contrast, the West Indian race, well-suited to lowland regions, exhibits rougher bark (Scora & Bergh 1992).

Delving into the aromatic realm, the alluring anise-scented smell that permeates the avocado groves can be attributed to the abundance of estragole—a distinctive organic compound detected exclusively in Mexican avocado cultivars (Pino et al. 2006; Pereira et al. 2013). This intriguing olfactory signature sets Mexican avocados apart from their counterparts, adding to the sensory allure of these fruits.

Turning the attention to the intricate world of leaf pubescence, the presence of fine hairs on avocado leaves has implications for photosynthesis. These delicate structures contribute to a reduction in the amount of light absorbed by the leaves, thus influencing the energy balance and significantly slowing down photosynthetic activity throughout the growth season (Ehleringer et al. 1976). Interestingly, leaf pubescence also demonstrates its adaptive prowess by enhancing water use efficiency through the promotion of condensation (Konrad et al. 2015). This adaptive trait holds particular significance in drier environments and for areas vulnerable to the pronounced environmental fluctuations associated with climate change.

Leaf morphology plays a critical role in understanding plant productivity, particularly in the context of avocado trees. The shape of avocado leaves is a crucial characteristic that serves as a key indicator of leaf area and its influence on light reflection, directly impacting plant productivity (Nkansah et al. 2013). This study examined the diverse range of leaf shapes exhibited by avocado germplasm, building upon previous

studies conducted in various regions including Colombia, Ghana, Mexico and Tanzania, (Abraham et al. 2018; Acosta-Díaz et al. 2020; Juma et al. 2020b; López-Galé et al. 2022). These studies have contributed valuable insights into the assortment of leaf shapes observed. The present study uncovered a noteworthy finding, revealing the presence of nine distinct leaf shapes. The observed richness of 9 leaf shapes aligns with previous findings of leaf shapes richness reported in Tanzania (9 leaf shapes), indicating a consistent pattern across different geographic landscapes (Juma et al. 2020). This could potentially be attributed to the larger sample size employed in both investigations, as they focused on describing native populations rather than solely relying on core collections.

Fruit shape plays a crucial role in consumer preferences and market appeal. The availability of a wide variety of fruit shapes and mature skin colors enables targeting a broader customer base. Our study observed fruit shapes that are consistent with previous research conducted by Juma et al. (2020), who explored the association between fruit shapes and avocado cultivars originating from different races in Tanzania. These findings indicate that the avocado trees in Guatemala possess genetic diversity encompassing all three avocado races.

The highly desirable buttery texture, favored by Guatemalan avocado buyers for its exceptional flavor, holds significant economic value. Previous studies have established a connection between certain Mexican and Guatemalan avocado varieties and their characteristic buttery flesh texture (Bost et al. 2013). Moreover, research has indicated that buttery and pastose flesh textures in Guatemalan and Mexican avocados are associated with moderate to high oil contents (Pereira et al. 2013; Espinosa-Alonso et al. 2017), suggesting genotypes from both regions also in our samples. In this study, the presence of buttery and pastose flesh textures suggests the presence of avocados from the Mexican and Guatemalan races. Conversely, the occurrence of watery flesh textures indicates avocados from the West Indian race. Notably, variations in fruit texture reported by producers between dry and rainy seasons suggest the potential influence of environmental factors on fruit quality. This phenomenon aligns with the findings of Juma et al. (2020) and emphasizes the importance of considering the potential effects of climate change on fruit quality, particularly in vulnerable regions like western departments in Guatemala.

The identification of a wide range of seed forms in our study aligns with previous reports on the morphological characterization of avocados from Tanzania, which described approximately 17 different seed forms (Juma et al. 2020). In contrast,

characterization reports from India and Colombia identified six and three different seed forms, respectively (Ranjitha et al. 2021; López-Galé et al. 2022). The substantial difference between findings in Guatemala and those from India and Colombia could be attributed to our larger sample size and the greater genetic diversity expressed through the diverse seed shapes observed. Popenoe (1974) established a correlation between spheroid, obovate, and oblong-conic seed shapes and the Guatemalan, West Indian, and Mexican avocado races, respectively. The presence of these seed morphologies in the present study suggests that the native avocado germplasm in Guatemala derives from all three avocado races.

The avocado plants examined in this study exhibited a diverse range of traits that hold significance in both economic and breeding contexts. Characteristics such as fruit shape, ripe fruit skin color, and flesh texture are valuable and can serve as selection criteria for future production by farmers and breeders aiming to develop improved cultivars. Previous research by Barrett et al. (2010) has highlighted the impact of external factors such as vibrant color, shine, and fruit form on enticing buyers and stimulating impulsive purchases. Once consumers taste the fruit, factors such as texture, freshness, and other flavor attributes become crucial in determining their satisfaction. Visual cues play a role in consumers' perception of freshness and flavor quality during the moment of purchase, although it is important to note that these cues can sometimes be misleading (Shewfelt 2000; Barrett et al. 2010).

6.1.3 Factorial analysis of mixed data (FAMD)

The observed variations in fruit shape, skin color, seed shape, flesh texture, and anise odor in the leaves align with previous studies that have associated specific traits with different avocado races. For instance, the presence of anise odor in the leaves has been linked to the Mexican race, while rough fruit surfaces and rough cotyledon surfaces have been associated with the Guatemalan and West Indian races, respectively (Popenoe 1935; Bergh 1992; Janick 2005). These findings highlight the value of these traits as reliable indicators for distinguishing between avocado horticultural races.

In this study, the FAMD proved to be highly valuable in demonstrating that the analyzed trees, representing native Guatemalan avocado germplasm, displayed a combination of varying levels of each analyzed qualitative characteristic. This observation could be attributed to the occurrence of hybridization between the three

horticultural races facilitated by the absence of sterility barriers and favorable floral biology that promotes cross-pollination (Alcaraz & Hormaza 2011; Gross-German & Viruel 2013). However, the FAMD analysis highlighted that the distinguishing features among individuals were primarily related to dimensions 1 and 2, notably fruit shape and skin color, seed shape, flesh texture, and anise odor in the leaves, which made the most significant contributions.

6.1.4 Population structure

The FAMD analysis provided valuable insights into the diversity and variability of avocado germplasm in this study. The visualization of individual data points in the new feature space allowed for a comprehensive assessment of the avocado's qualitative and quantitative morphological attributes. Interestingly, no prominent grouping of samples was observed, indicating significant variance among the examined trees at the population level. This suggests a high degree of phenotypic divergence within each area and a lack of differentiation among populations.

The absence of distinct sample groupings may be attributed to the unique characteristics of avocado as a highly heterozygous cross-fertilizing species. Avocado plants produce fruits with monoembryonic seeds, resulting in a high level of genetic variability within the progeny (Alberti et al. 2018). This inherent variability contributes to the observed phenotypic diversity among the evaluated populations. Furthermore, the natural repopulation processes that occur with minimal selection interference from producers likely enhance the morphological variability, particularly in fruit and seed characteristics, which exhibit significant natural variability (López-Galé et al. 2022).

These findings have important implications for avocado breeding programs and germplasm conservation efforts. The high degree of variability observed highlights the potential for selecting and breeding avocado cultivars with desirable traits. The wide range of phenotypic variation within the avocado germplasm can serve as a valuable resource for developing new cultivars that meet the diverse preferences of consumers and the needs of the avocado industry.

The absence of distinct sample groupings also indicates that the populations studied share a common genetic background and have not undergone significant differentiation. This suggests a high degree of gene flow and genetic exchange among avocado trees, potentially facilitated by the absence of sterility barriers and the favorable

floral biology of the species (Alcaraz & Hormaza 2011; Schaffer et al. 2012; Gross-German & Viruel 2013; Stern et al. 2021). Understanding the patterns of gene flow and genetic exchange in avocado populations can contribute to the development of effective conservation strategies and the sustainable management of genetic resources.

6.1.5 Hierarchical cluster of principal components (HCPC)

The results of the study revealed interesting findings regarding the genetic diversity and population structure of avocado trees in Guatemala. The clustering analysis based on principal components highlighted the presence of three distinct clusters, suggesting a genetic mixing among avocado populations from different geographical locations. This mixing could be attributed to seed exchanges between farmers, where avocado seeds are frequently traded among friends and family members, as well as the practice of swapping fruits between populations for seed sowing. Additionally, the cultural significance of avocados in Guatemala, coupled with their religious, mythical, economic, and therapeutic value (Galindo-Tovar et al. 2008; Landon 2009), may have contributed to the observed genetic mingling among populations. These findings highlight the ongoing influence of traditional practices and cultural factors in shaping the genetic diversity and population structure of avocados in the region.

The results of the HCPC analysis revealed statistically significant associations between the clusters' partition and both quantitative and qualitative characters. The highest explained variances were observed for FW, SW, and FL, indicating that these traits strongly contributed to the clustering patterns. These findings are consistent with previous studies that have highlighted the importance of these morphological characteristics in distinguishing avocado races (Bergh, 1992; Janick, 2005; Popenoe, 1935).

In the first cluster, most individuals exhibited a rugged fruit skin surface and a smooth cotyledon surface, which aligns with the typical traits associated with the Guatemalan race. However, the presence of anise odor in the leaves was not highly represented in this cluster, indicating a deviation from the expected pattern. Additionally, the lowest seed weight was observed in this cluster. These characteristics further support the assignment of this cluster to the Guatemalan race, as they align with the race's description of fruits with rough surfaces and seed sizes that are proportionate to the fruit size.

The similarity in morphological characteristics between the individuals in the first cluster and the traits associated with the Guatemalan race suggests that the genetic pool of this cluster predominantly originates from the Guatemalan race. These results provide insights into the genetic composition of the avocado populations studied and contribute to our understanding of the diversity within the Guatemalan race.

Consistent with the observations, the individuals in the second cluster demonstrated the lowest fruit weight compared to the other clusters, as evidenced by the statistical analysis (v.test -2.98). On the other hand, they exhibited significantly higher seed weight (v.test = 2.58), further emphasizing the contrast in quantitative characteristics within this cluster. The qualitative and quantitative traits identified within the second cluster align with the botanical description of the Mexican race. The Mexican race is known for its fruits with high oil content, resulting in a buttery texture, as well as smooth fruit surfaces. The characteristic anise-like odor in the leaves is also a distinguishing feature of the Mexican race (Ayala-Silva and Ledesma, 2014; Bergh, 1992). These findings provide further support for the hypothesis that the genetic composition of the individuals within the second cluster is primarily derived from the Mexican race.

The morphological traits observed in the third cluster closely resemble the description of the West Indian race. The absence of anise scent in the leaves, larger fruit and seed sizes that range from large to very large, and the typical pear-shaped or round fruit shape are characteristics associated with the West Indian race (Ayala-Silva and Ledesma, 2014; Popenoe, 1935). The rough cotyledon surface observed in this cluster further supports its resemblance to the West Indian race. Understanding the genetic composition of different avocado populations, including the presence of distinct races like the West Indian race, is significant for avocado breeders and conservationists. Such knowledge can facilitate targeted breeding programs and conservation efforts aimed at preserving and utilizing the unique genetic diversity within the West Indian race.

The provided data offers valuable insights into identifying individuals with genotypes that closely resemble the mentioned avocado races. This information can be utilized to establish a germplasm collection aimed at preserving the natural genetic variability associated with each Guatemalan avocado race, which can then be effectively utilized in future breeding programs.

The findings of this study demonstrate a clear genetic basis for each cluster, aligning them with specific avocado races. However, it is important to note that the clusters also exhibit traits that deviate from the typical characteristics of their related races.

This suggests a significant degree of hybridization, likely attributed to the absence of sterility barriers between avocado horticultural races (Alcaraz and Hormaza 2011; Gross-German and Viruel 2013). As a result, inter-breeding has occurred, leading to the emergence of highly diverse genotypes that are well-adapted to the climatic conditions of the region.

The identification of these diverse genotypes and the absence of clear population structure have significant implications for avocado breeding and conservation efforts. It highlights the potential for exploiting the wide range of genetic resources available in Guatemala to develop improved avocado varieties that exhibit desirable traits such as disease resistance, yield, and fruit quality. Moreover, the preservation of this genetic diversity through germplasm collection is crucial for maintaining the long-term sustainability and resilience of avocado crops in the face of changing environmental conditions and emerging threats.

6.2 Genetic diversity and population structure of native Guatemalan avocado using AFLP and SSRs markers

6.2.1 Genetic diversity

Efforts to assess the extent of genetic diversity in avocado materials have led to the use of the AFLP molecular marker, which has proven to be useful (Cerda-Hurtado et al. 2015). Additionally, other dominant molecular markers such as RAPD and ISSR (Fiedler et al. 1998; Reyes-Alemán et al. 2013) have facilitated the molecular characterization and identification of genetic diversity within the *Persea* genus and also within the species *P. americana*. The results obtained by Cuiris-Pérez et al. (2009), who evaluated genetic diversity within the Mexican race using the AFLP marker, showed a very similar pattern to the present study when comparing the level of organization through cluster analysis using generated dendrograms.

The characterization experience of native materials in the state of Nuevo León, Mexico, using the same AFLP molecular marker (Gutiérrez-Díez et al., 2009), also revealed a high degree of genetic diversity. However, this study achieved a certain level of arrangement through the dendrogram, thanks to the previous characterization at the race level of the 42 analyzed accessions. Nevertheless, due to the scope of the present study, there was no prior characterization at the race level, making it impossible to

establish a relationship as in the aforementioned study. However, it was possible to delve into population indices (Gutiérrez-Díez et al., 2009).

In this study, the 12 SSR loci yielded 286 alleles with alleles per locus ranging from 9 to 32, while Juma et al. (2020) detected 167 alleles using 10 SSR loci across 226 Ghanaian avocado landraces, with the number of alleles ranging from 10 to 23 per locus. Similarly, Schnell et al. (2003), analyzing genotypes from the National Germplasm Repository (FL, USA) identified 8 to 30 alleles per locus in their study. Other studies have reported relatively lower values. Guzmán et al. (2017), using wild genotypes from the Experimental Station of the National Forestry, Crops, and Livestock Research Institute in Mexico, documented 20 alleles per locus. Gross-German & Viruel (2013), encompassing accessions that included rootstocks, commercial varieties, and Spanish local selections from avocado collections in Spain, reported 19 alleles per locus. Similarly, Abraham & Takrama (2014), investigating landraces Ghanaian genotypes, reported 12 alleles per locus.

The observed differences in allele numbers among these studies could be attributed to various factors, including variations in sample size, the use of different molecular markers, the diversity of the investigated germplasm, and the genotyping platform's accuracy (Vieira et al. 2016). Additionally, factors such as the quality of genomic DNA used for SSR PCR amplification, and the optimization of PCR procedures may contribute to the discrepancies (Juma et al. 2020).

The mean observed heterozygosity (Ho) across the analyzed SSR loci in the present study was determined to be 0.53. This value is comparable to the 0.56 reported by Boza et al. (2018), who studied non-commercial avocado trees from Mexico, suggesting that both gene pools exhibit similar levels of genetic diversity. Other studies have reported higher Ho values, such as 0.65 by Juma et al. (2020), 0.64 by Schnell et al. (2003), and 0.61 by Guzmán et al. (2017). In contrast, lower Ho values have been reported in avocado genetic reservoirs by Abraham & Takrama (2014) with 0.48. Additionally, Liu et al. (2020), analyzing Chinese avocado landraces, reported 0.39, indicating relatively lower genetic diversity in the studied germplasm compared to our findings. The variations in Ho values could be attributed to several factors, including the use of different SSR loci among the studies.

The analysis of genetic diversity among populations revealed notable differences in various diversity measures. The populations of Sac, Sol, To-Qui, and Hue exhibited the highest levels of observed and expected heterozygosity, Shannon's information index,

and allelic richness, respectively **(Table 9)**. These results indicate that the populations from the central and western regions possess greater genetic diversity compared to those from the northern region. Consequently, these central and western populations may serve as valuable sources of elite genotypes for breeding programs aiming to incorporate genetic diversity into new commercially successful cultivars. Moreover, the populations from the central and western regions may exhibit a higher potential for adapting to climate fluctuations when compared to the northern populations. On the other hand, the BV population displayed the lowest values across most diversity measures, suggesting a lower level of genetic diversity. This observation can be attributed to the extensive replacement of native avocado genotypes with commercially favored cultivars, particularly the Hass variety, driven by the increasing market demand. This replacement process has led to a reduction in the genetic diversity within the avocado gene pool.

Avocado's unique floral morphology, heterodichogamy, not only increase genetic diversity by promoting cross-pollination (Alcaraz & Hormaza 2011) but also, in conjunction with the absence of sterility barriers between horticultural races, fosters natural or human-induced interracial hybridization (Gross-German & Viruel 2013), further contributing to overall diversity. Scientific evidence confirms the separation of the three horticultural avocado races during domestication (Furnier et al. 1990; Ashworth & Clegg 2003). However, with the arrival of the Spanish in the 16th century, these races came into contact, resulting in interracial hybridization due to the genetic flow facilitated by human activity. This situation helps us understand the influence of human-plant interaction on the current population structure of avocados.

6.2.2 Population genetic divergence

The analysis of the studied populations revealed a potential inbreeding effect, resulting in a departure from Hardy-Weinberg equilibrium across the 12 SSR loci (Table 8). Wright's fixation indices (F_{IT}, F_{ST}, and F_{IS}) were employed to assess the genetic variation within and among populations. F_{IT} estimates the deficit ($F_{IT} > 0$) or excess ($F_{IT} < 0$) of heterozygosity across all populations (Lachance 2009). The computation of AMOVA without considering regions yielded a global heterozygosity deficit (F_{IT}) of 0.326 ($p < 0.001$; Table 12), indicating an approximately 33% increase in observed homozygotes. This finding is consistent with the G_{IS} values (Table 8), which indicate a significant reduction in heterozygotes.

Similarly, F_{IS} measures the deviation from panmixia at a local scale, reflecting the deficit ($F_{IS} > 0$) or excess ($F_{IS} < 0$) of heterozygotes within subpopulations (de Meeûs et al. 2007). The average F_{IS} was determined to be 0.313 ($p < 0.001$), indicating a deficit of approximately 31% of heterozygotes within subpopulations. The observed high frequency of homozygotes can be attributed to various factors. One contributing factor is the presence of null alleles, which may result from PCR failures during microsatellite amplification (Wattier et al. 1998; Stadhouders et al. 2010), causing individuals with heterozygous genotypes for these alleles to be misclassified as homozygotes for dominant alleles (Lemer et al. 2011). Another factor that can lead to a decrease in the observed frequency of heterozygotes is the Wahlund effect, which refers to the apparent excess of homozygotes and deficit of heterozygotes observed within a sample of individuals due to population subdivision (de Meeûs 2018).

F_{ST} is a metric that quantifies the average deficit of expected heterozygotes among subpopulations compared to that expected for the total population. It is commonly used to assess differentiation between subpopulations. In the case of the eight geographical populations of avocados, the genetic differentiation was found to be low ($F_{ST} = 0.018$, $p < 0.001$). However, a significant subdivision was observed within the Guatemalan avocado population. The results of the AMOVA indicated that the majority (approximately 98%) of the genetic variation was shared within and among populations, indicating a high level of gene flow and minimal genetic differentiation (Wright 1931). This is supported by the high gene flow value (Nm = 12.25).

The F_{ST} value obtained in this study (0.018) is similar to 0.06 and 0.05 reported by Juma et al. (2020) (0.061) and Cañas-Gutiérrez et al. (2019), respectively, but lower than those 0.19, 0.25 and 0.63 reported by Boza et al. (2018) and Gross-German & Viruel (2013), and Talavera et al. (2019), respectively. The discrepancies can be attributed to the composition of the genotyped avocado trees. In our study, only native avocado varieties were genotyped, excluding commercial varieties, as noted by Juma et al. (2020) and Cañas-Gutiérrez (2015). On the other hand, Boza et al. (2018) included *P. americana*, *P. nubigena*, and *P. krugii* trees in their analysis, while Gross-German & Viruel (2013) examined a set of 42 avocado accessions, including rootstocks, commercial varieties, and Spanish local accessions.

The AMOVA analysis revealed that the genetic variation among the eight populations accounted for 2.5% and 1.8% of the total variation for AFLP and SSR data, respectively (**Table 11**, **Table 12**). Furthermore, when grouping the samples based on the

three main geographical regions, the genetic variation among groups was also 1.8% with a significant F_{CT} value ($F_{CT} = 0.018$, $p < 0.001$), indicating a subtle substructure. These findings align with the results reported by Juma et al. (2020), who observed a genetic variation among groups of 1.98% ($F_{CT} = 0.019$, $p < 0.05$) when grouping the geographical populations based on regions.

The substructure identified through AMOVA is consistent with the clustering pattern observed in the UPGMA dendrogram at the population level (**Figure 22**A-B), where populations are grouped according to their geographical region. However, the Sol population exhibited a deviation from the expected grouping, as it showed closer genetic affinity with the western population despite its initial classification within the central region. This observation can be explained by the relatively short geographical distance between the Sol and western populations. The substructure and organization revealed in the dendrogram are further supported by the Mantel test (**Figure 22**C-D), which indicated a small but significant correlation between geographical and genetic distances among the eight populations.

6.2.3 Population genetic structure

Upon delving into the genetic relationships among all sampled trees at the individual level using UPGMA cluster analysis (**Figure 23**A-B), it was not possible to discern a clear separation of trees based on their populations or regions. This observation was further supported by the pairwise population Phi_{PT} and $G'_{ST(Nei)}$ values (**Figure 22**A-B), which indicated a lack of differentiation between pairs of populations, such as AV versus BV, Sac versus Chi, and Sol versus To-Qui. The absence of distinct genetic boundaries among avocado populations in Guatemala can be attributed to the country's status as one of the three domestication centres for avocado, coupled with its religious, mythological, economic, and medicinal significance (Galindo-Tovar et al. 2008). This historical and ongoing cultural significance has contributed to genetic admixture among avocado populations, thereby influencing the genetic diversity and structure of these populations.

AMOVA, UPGMA, STRUCTURE, and DAPC consistently revealed significant genetic similarities and low differentiation among native Guatemalan avocado populations. These results suggest that diverse ecological conditions and geographical disparities do not anymore contribute significantly to the formation of distinct genetic structures within the studied populations, likely due to the extensive gene flow facilitated

by human activities (Nm = 12.25), as previously suggested by Galindo-Tovar et al. (2008). These human-mediated factors have acted as connectors, enabling genetic exchange between different regions and populations, thus reducing the impact of ecological and geographical barriers on genetic differentiation. Human-mediated gene flow can be considered the primary factor influencing the diversity and genetic structure of Guatemalan avocado populations.

The genetic status of native avocado trees in Guatemala deserves special attention, as the current genetic structure is the result of human interaction through domestication processes, which are believed to have begun with the arrival of humans in Mesoamerica approximately 15,000 years ago (Goebel et al. 2008). Since then, historical, paleohistorical and paleoecological evidences suggests that after the domestication processes initiated in Mexico (Galindo-Tovar, Lee-Espinoza, Murgía-González, Leyva-Ovalle, & Landero-Torres, 2013), avocado materials were transported to the Yucatan Peninsula, Guatemala, and Belize, where the Maya culture established one of the three primary centers of avocado domestication (Colunga-GarcíaMarín & Zizumboo, 2004; Gama & Gomez, 1992). From there, the religious, mythological, economic and medicinal implications of avocados in the Mesoamerican region (Galindo-Tovar, Ogata-Aguilar, & Arzate-Fernández, 2008) fostered the migration of avocado materials beyond their original populations. This situation persists today and has become the primary influencing factor in the diversity and genetic structure of this valuable resource.

The present richness of native Guatemalan avocado germplasm is a consequence of ongoing gene flow between populations. However, this valuable gene flow is currently under threat due to deforestation and the widespread introduction of commercial varieties, which displace native genotypes and lead to a reduction in their population sizes (Rincón-Hernández et al. 2011; Bullock et al. 2020). Therefore, it is crucial to prioritize measures that prevent further population decline and maintain connectivity among populations to ensure the conservation of genetic diversity.

6.2.4 Joint analysis of phenotypic and molecular data

High cophenetic coefficients were observed for both phenotypic and molecular data, signifying a substantial alignment between each data type and its corresponding clustering dendrogram. The cophenetic coefficient's significance lies in its ability to gauge the concordance between dendrograms and their respective distance matrices

(Barrett et al. 2010). A correlation coefficient exceeding 80% indicates a robust alignment between these matrices (Shewfelt 2000; Allendorf et al. 2022). These results underscore the effectiveness of phenotypic evaluations and SSR markers in independently identifying genetic diversity and structuring wild avocado populations.

Nonetheless, it is noteworthy that despite the strong alignment observed between phenotypic and molecular data with their respective dendrograms, the tanglegram analysis revealed an entanglement value of 0.34. This value implies a certain degree of discrepancy or partial misalignment between the two dendrograms, representing the microsatellite and phenotypic data of wild avocados. Essentially, this indicates that while phenotypic and molecular data individually align well with their corresponding clustering, slight variations emerge when these two datasets are directly compared (Simionca Mărcășan et al. 2023). The entanglement value of 0.34 signifies that these distinctions exist but are not pronounced, falling between complete congruence (a value closer to 0) and substantial disparity (a value closer to 1).

This discrepancy between the dendrograms suggests that the genetic structure and morphological structure of the wild avocado populations may not be fully aligned. It is possible that some individuals or groups of individuals that are genetically close show significant morphological differences, and *vice versa*. The reasons behind this discrepancy could be diverse (Liu et al. 2020). Genetic variability within populations, the influence of the environment on the expression of morphological traits, and the evolution of specific traits in different geographical regions, along with the marker system itself, which primarily amplifies non-coding regions and may not necessarily be associated with features (Vieira et al. 2016; Allendorf et al. 2022), are factors that could contribute to this discordance between genetic and morphological data. These results suggest that a single data source may not fully capture the diversity and structure of wild avocado populations. It is important to consider multiple approaches and data sources to obtain a more comprehensive understanding of the genetic and morphological variation in these populations.

The observed low correlation between phenotypic and genotypic distance matrices confirms their independence and complementary nature rather than a limitation (Singh et al. 1991). This discordance and observed low correlation can be explained by the molecular marker's capacity to identify genetic-level variations, unaffected by natural or artificial selection, unlike phenotypic markers (Alves et al. 2013). Furthermore, molecular markers are selectively neutral, in contrast to the genomic region linked to the

phenotypic trait, which is often subject to selection influenced by the environment (Collard et al. 2005; Sunil et al. 2011). Consequently, the genetic diversity captured by molecular markers may not always correspond directly to the phenotypic diversity due to the complex interplay of genetic and environmental factors affecting trait expression. The development of trait-related markers, such as EST markers, could serve as a valuable and potentially superior tool for bridging the gap between morphological and genetic data.

Previous studies of other crops such as cowpea (Nkhoma et al. 2020), yam (Agre et al. 2019), and common bean (Guidoti et al. 2018) also reported inconsistences between phenotypic and genotypic matrices. To address this, using a joint matrix derived from both phenotypic and genotypic data is highly recommended for increased precision (Sartie et al. 2012; Vinu et al. 2013). The strong correlations exhibited by phenotypic and genotypic matrices with the joint matrix further support their use for enhanced precision without overlapping. Previous studies also support the combined use of molecular and phenotypic data for assessing genetic diversity (Sunil et al. 2011; Sartie et al. 2012; Vinu et al. 2013; de Andrade et al. 2017).

6.3 Selection of a core collection of avocado genotypes for long term conservation based on agro-morphological traits and molecular markers

The creation of a core collection for wild Guatemalan avocado is crucial to safeguard genetic diversity and ensure adaptability, resilience, and sustainability in the face of modern agricultural challenges and environmental changes. This core collection serves as an essential genetic resource, preserving vital genes for future breeding and cultivation (Brown 1989; Franco et al. 2006).

The concept of core collections was introduced to enhance the efficiency of evaluating and utilizing genetic resources while preserving maximum diversity. Core Hunter was utilized to develop the core collection, prioritizing both diversity and usefulness. This approach aims to strike a balance between representing total diversity and meeting the needs of breeding programs, ensuring a multipurpose core set with maximum genetic potential (Thachuk et al. 2009; De Beukelaer et al. 2018; De Beukelaer & G 2023).

6.3.1 Assembly and quality evaluation of the core collections

The CC 03, which was developed by giving equal weightage of 1:1 to both E-NE and A-NE, exhibited maximum diversity with high E-NE and E-E genetic distances and maximum representativeness with low A-NE genetic distances, as revealed by the detailed comparative statistical analyses (**Table 13**). Previous studies have suggested that maximizing the average genetic distance within a core collection is a desirable quality criterion for core collections intended for plant breeders (Franco et al. 2006; Thachuk et al. 2009).

Furthermore, the assessment of mean difference (MD%), variance difference (VD%), coefficient of range (CR%), and variation range (VR%) between the whole studied germplasm and various core sets indicated that the CC 03 had a VD of 95.5%, CR of 92.06%, and VR of 108.71%. To ensure a core collection is more diverse and representative, it is desirable to have a lower MD value (< 20%), larger VD and CR values (> 80%), and a VR value (> 100%) (Hu et al. 2000; Agrama et al. 2009). Similar parameters were also employed in the evaluation of core sets in avocado (Guzmán et al. 2017), and other crops such as Indian mustard (Nanjundan et al. 2022), rice (Agrama et al. 2009; Ndjiondjop et al. 2023), and wheat (Phogat et al. 2021). The geographical representativeness of the extracted core set is evident from the relative distribution of areas of collection of indigenous and exotic germplasm in the entire collection and core set.

6.3.2 Comparative evaluation of the core collections with the whole native Guatemalan avocado

To comprehensively assess quality, we compared the entire avocado germplasm with the core set CC 03 using various statistical measures, including summary statistics, diversity indices, correlation analysis, and PCA. The core set CC 03 exhibited a higher coefficient of variation (CV) for all traits compared to the whole collection, indicating its ability to capture greater variability.

The relative frequency bar plots for qualitative traits in both the whole collection and the core set demonstrated the capture of the entire range of variation (**Figure 34**). Similarly, the distribution of the eight quantitative traits in avocado, as depicted in the boxplots, showed consistent patterns between the entire germplasm and the core collection (**Figure A5**). These findings suggest that the core collection is representative

of the entire germplasm collection, and that the core collection can be used to study the genetic diversity of the entire collection. The core collection can also be used to identify accessions with desirable traits for breeding programs (Rubinstein et al. 2019).

Quantile-quantile (QQ) plots and Kullback–Leibler distance analysis (Kullback & Leibler 1951; Wilk & Gnanadesikan 1968) confirmed the core set's representation of trait distribution, with values ranging between 0.037 and 0.08 (**Figure A4**). These results imply a high degree of similarity between the core set and the whole germplasm. Furthermore, the core set effectively maximized existing diversity, as evident from increased Shannon–Weaver diversity index (H′) values for most traits, except for slight differences in a few cases. This reaffirms the core set's role in preserving genetic diversity.

Correlation coefficient analysis has been widely employed in various crop species, including avocado, to examine the inter-relationships among different traits (Juma et al. 2021; Cañas-Gutiérrez et al. 2022; Awachare et al. 2023). In both, the whole collection and the core set, strong positive correlations were observed between traits such as SW and FW, and FW and TC (**Figure 35**). These findings support the preservation of trait associations within the core collection. Evaluating the quality of core collections has often involved comparing the correlation coefficients of the whole collection with those of the core collection (Reddy et al. 2005; Mahajan et al. 2007).

The results presented here demonstrate the presence of a broad range of variability in phenotypic traits within the native avocado germplasm, and this variability is preserved in the proposed core set. These findings emphasize the importance of phenotypic characterization-based evaluation in assessing genetic variability, serving as a crucial foundation for the effective utilization and conservation of germplasm resources, even in the face of reduced overall genetic diversity (Hu et al. 2022).

6.4 Comparison of the genetic diversity of Guatemalan and Ethiopian avocado germplasm.

6.4.1 Polymorphism diversity of SSR markers

Many molecular markers have been used to quantify the diversity of avocados throughout the world (Sandoval-Castro et al. 2021; López-Guzmán et al. 2021; Ramos-Aguilar et al. 2021; Ninh et al. 2022; Ruiz-Chután et al. 2022; Wienk et al. 2022). Most of the research employed the SSR marker to clearly distinguish the avocado races and the presence of genetic diversity within and between populations. The random distribution, abundance,

and wide genome-wide coverage of SSR markers make them feasible for avocado diversity studies. In this work, 12 microsatellites that had previously been evaluated by Alcaraz & Hormaza (2007) (Sharon et al. 1997; Ashworth & Clegg 2003) were used to carry out molecular characterization on 298 avocado landraces collected from Guatemala and Ethiopia (seven populations).

With a mean number of distinct alleles per locus ranging from 3.4 to 13.7 and an overall average of 10.2, the genetic diversity present in the current continental populations of avocados is astounding (**Table 19**). This discovery highlights the significant genetic diversity seen in avocado germplasm from Guatemala and Ethiopia. Interesting trends in allele diversity are shown when results from related studies are taken into consideration. For instance, Juma et al.'s (2020) study analyzed 226 avocado trees and found that the average allele count was 9.4. Comparable results were obtained by Boza et al. (2018), who focused on three different horticultural groups and discovered an overall mean of 9.1 alleles, which is comparable but somewhat higher. In contrast, Gross-German and Viruel (2013) found an average of 5.6 alleles across 41 avocado trees. In a different light, Schnell et al.'s (2003) analysis of 221 samples revealed an average of 10.3 alleles, which was comparatively similar to the current analysis. Notably, Cañas-Gutiérrez et al. (2019) observed an overall mean of 4.3 alleles for 90 Colombian avocado cultivars, which is noticeably lower. This shows that the SSR loci can create enough variation to be used as molecular markers in the current analysis.

6.4.2 Population genetic diversity

The avocado collection's Shannon-Weaver diversity score ranged from 1.22 to 2.12, indicating that the genetic diversity of avocados from Guatemala and Ethiopia is variable, with Gamo showing the lowest values. The fact that the bulk of the avocado samples were from western Guatemala may help to explain these findings. Additionally, it was discovered that different populations had heterozygosities that ranged from 0.55 (Gamo, Ethiopia) to 0.85 (west Guatemala). Ho and He ranged in 0.23 to 0.70. This disparity can be explained by the presence of different sample sizes that may potentially be the cause of variations in heterozygosity estimates (Ashworth et al. 2004; Borrone et al. 2008; Gross-German & Viruel 2013).

Samples from the three avocado botanical races and the species *P. steyermarkii* were determined to have Ho = 0.61 with 25 microsatellites standardized by Ashworth et

al. (2004). In the study carried out by Borrone et al. (2008), they included the three botanical races and *P. schiedeana* and calculated Ho between 0.17 to 0.73. In the investigation done by Schnell et al. (2003), Ho values varied from 0.73 to 0.97. Gross-German and Viruel (2013) yielded Ho estimations that were 0.67 on average from 42 accessions from Spain that were investigated using 47 microsatellites. Additionally, Juma et al. (2019) examined 226 accessions from Tanzania, using a comparable sample size to the current study; their Ho values varied from 0.46 to 0.84.

HWE study revealed that all the samples from Ethiopia (Gamo, Wolaita, Sidama, WG) and Guatemala (central, norther, and western) are not in HWE. A similar study by Juma et al. (2019) discovered eight of the SSR markers were deviated from HWE. The major explanations of deviating from HWE is the domestication process that creates a bottleneck effect in an effort to promote particular linkages through human-induced selection.

The findings highlight the noteworthy differences in genetic diversity between Guatemalan and Ethiopian avocado populations. Specifically, Guatemalan avocados displayed a considerably higher genetic diversity compared to their Ethiopian counterparts. This variation can be attributed to Guatemala's status as the center of origin for avocado (Storey et al. 1986; Bergh 1992). The intrinsic diversity of avocado in this region has long been acknowledged (Gross-German & Viruel 2013; Rubinstein et al. 2019), and our study reaffirms and quantifies this inherent richness.

Preserving genetic diversity within avocado populations is critical to bolstering the crop's resilience against environmental challenges and adapting to evolving agricultural demands (Juma et al. 2020a; Sandoval-Castro et al. 2021). The notable disparity in genetic diversity between Guatemala and Ethiopia emphasizes the significance of focused conservation initiatives. This is particularly crucial in regions such as Ethiopia, where genetic diversity seems comparatively limited.

6.4.3 Genetic differentiation and genetic structure

The AMOVA results showed significant differences between the countries and among the populations, indicating the presence of population structure. The trend explains the difference in climatic factors. Furthermore, the ancient human cultivation of avocados in Guatemala may have influenced the genetic structure in the country. Guzmán et al. (2017) found that Mexican avocados are genetically divided into two groups, implying that

differences in agroecology might contribute to such differentiation. An AMOVA analysis of 226 avocados from Tanzania revealed a substantially higher level of genetic diversity among populations (77%) (Juma et al. 2019), which is higher than our results of 56% (**Table 18**).

The results of the phylogenetic tree, an ancestral group based on STRUCTURE software, and cluster patterns in DAPC and PCA reveal the connection between Wondo Genet avocados and the western and northern Guatemalan populations. The high diversity and relatedness to Mesoamerican germplasm here is due to Wondo Genet being a research center and harbouring released varieties, that can be traced back to Guatemala. In contrast, the limited diversity observed in the Sidama, Wolaita, and Gamo populations in Ethiopia indicates a narrower genetic basis, likely due to their descent from a small group of trees, leading to a founder effect. The founder effect may cause the gene pool not to be an accurate reflection of the population that gave rise to them. Therefore, it is possible that the founder effect may result in lower genetic diversity in a population (Parisod et al. 2005). Moreover, avocado cultivation hastened regional adaptation and may have led to genetic drift in mentioned Southern populations (Smýkal et al. 2018).

This statement can be confirmed by the fact that there are fewer alleles in the Gamo population than in Wondo Genet, as shown by the lower Ne value. Furthermore, the avocado population in Wondo Genet showed better-effective alleles. Compared to the other populations, western Guatemala has moderate genetic diversity Chanderbali et al. (2008). The western Guatemalan highlands may have provided ideal environmental conditions for the diversification of avocados. High Nm and low F_{ST} in western Guatemala and Wondo Genet suggest frequent gene exchange with other regions, which was also a factor in avocado diversification (Galindo-Tovar et al. 2007).

The observed patterns of genetic diversity and differentiation in avocado populations provide valuable insights into the crop's historical dispersal and contemporary genetic relationships. These findings align with existing evidence suggesting a notable genetic divergence between avocado populations in Ethiopia, specifically Sidama, Wolaita, and Gamo, and the Wondo Genet population, which exhibits a closer affiliation with western Guatemala. The historical transportation of avocado seeds by missionaries from the avocado-growing region of Spain to southern Ethiopia (Kamaraj et al. 2020) may explain the migration path of the Guatemalan avocado germplasm likely played a pivotal role in establishing the genetic connection with Wondo Genet. The high values of Nm detected between central, western, and northern Guatemala,

and Wondo Genet in Ethiopia provide strong support for ongoing gene flow between these geographically distant populations.

The limited genetic affinity and reduced gene exchange between Wondo Genet and the southern Ethiopian populations suggest that the avocados in southern Ethiopia have a distinct ancestry not connected to the Guatemalan germplasm found in Wondo Genet. To clarify the ancestry of avocado germplasm from southern Ethiopia, it should be compared with other germplasm from African countries like Tanzania (Juma et al. 2020a), South Africa (Wienk et al. 2022), Ghana (Abraham & Takrama 2014) or international collections (Alcaraz & Hormaza 2007; Guzmán et al. 2017).

In Ethiopia, growers select avocado trees based on desirable fruit traits from local markets or neighbors, and these seedlings evolve from indigenous germplasm (Biazin et al. 2018). This situation underscores the limited utilization of the Guatemalan genetic resource within Wondo Genet for developing avocado varieties aligned with Ethiopian market demands. Nevertheless, it also presents a promising opportunity to take advantage of the genetic basis in Wondo Genet, facilitating the integration of agriculturally advantageous genes into local cultivars via targeted breeding programs.

The genetic distinction between southern Ethiopian populations and Guatemalan germplasm highlights the critical need to conserve the genetic diversity inherent in the southern Ethiopian avocado populations. These populations possess alleles absent in Guatemalan germplasm, indicating an evolutionary process that has driven the adaptation of Ethiopian germplasm. These unique genetic traits could be valuable for developing cultivars tailored to local environmental conditions.

Expansion of avocado cultivation in a changing climate must involve the selection and hybridization of genotypes from different lineages for higher success rates and better adoption (Marsh et al. 2021). The avocado germplasm from Guatemala and Ethiopia exhibits substantial genetic diversity, suggesting that the SSR loci may be able to interpret the genetic diversity already present. Molecular data clearly show the presence of three ecotypes: northern, central-western, and Ethiopian. This points to three potential areas for improvement in avocado cultivation. The identification of these ecotypes could be useful for the development of breeding programs aimed at improving the genetic diversity of avocado cultivars. The genetic diversity of the germplasm could be harnessed to develop new cultivars that are better adapted to the changing climate (Singh & Behera 2022). The identification of the three ecotypes could also be useful for the conservation and management of avocado genetic resources in these regions.

The present study helps to find out the available genetic repository of avocados for breeding programs. It also lays the foundation for further research, such as the construction of genetic linkage maps, association studies, and population studies.

7 Conclusions

The objective of this study was to comprehensively investigate the genetic resources of native Guatemalan avocado germplasm, aiming to analyze their genetic diversity, relatedness, and population structure. This analysis was conducted by utilizing a combination of agro-morphological traits, AFLPs, and SSRs markers. Additionally, the study sought to establish a core collection of avocado genotypes for long-term conservation, considering both agro-morphological traits and molecular markers. Furthermore, the objective involved comparing the genetic diversity between Guatemalan and Ethiopian avocado germplasm, providing valuable insights into their genetic relationships and differentiation.

The study commenced by characterizing the native Guatemalan avocado trees using morphological markers, leading to the identification of effective traits such as fruit and seed weight, seed and fruit shapes, and fruit length. Through the application of clustering methods, the study successfully grouped the avocado trees into three main clusters, showcasing the utility of these morphological descriptors in characterizing both cultivated and native avocado trees. These findings contribute significantly to the understanding of avocado tree diversity and establish a foundation for future research in the characterization and conservation of avocado germplasm.

The study examined genetic diversity in native Guatemalan avocado trees using AFLP and SSR markers, highlighting the richness of this germplasm. Population differentiation was low, attributed to ongoing gene flow and historical human-mediated interactions. However, the ongoing deforestation and introduction of commercial varieties pose a threat to this gene flow, endangering native genotypes and reducing their population size. The genetic diversity was compared to phenotypic traits, revealing some discordance, emphasizing the need for combined data analysis. The strong correlation between the joint matrix and phenotypic/genotypic matrices confirmed their complementary nature. This approach offers a robust method for understanding genetic diversity by capturing both genotypic and phenotypic aspects. It has practical implications for breeding and conservation programs, ensuring the preservation and utilization of this valuable resource.

Based on the observed genetic variability, the study proceeded to develop an avocado core set through a comprehensive quality evaluation, comparing different core sets derived using specialized software and strategies. The final proposed core set,

consisting of 38 accessions, successfully captured maximum diversity and representativeness of the entire avocado germplasm under study. The effectiveness of the core set lies in its potential to provide expedited access to genetically diverse and agriculturally significant resources, thereby enhancing breeding programs and expanding the genetic base of avocado. Additionally, the diverse germplasm represented by the core set can undergo field evaluations for various stresses and subsequent genome-wide association studies, enabling the identification of genes, alleles, and markers associated with important traits.

Moreover, this study sheds light on the genetic differentiation and diversity of avocado populations in Ethiopia and Guatemala. It revealed notable genetic distinctions in Sidama, Wolaita, and Gamo populations compared to the genetically closer Wondo Genet group, which shares an affinity with Guatemalan germplasm. Historical seed transport by missionaries may explain this link.

Southern avocado populations shared genetic traits due to unrestricted gene exchange caused by the absence of geographical barriers. In contrast, Wondo Genet exhibited limited genetic affinity and minimal gene exchange with these populations. These findings indicate that the southern Ethiopian avocados have a distinct ancestry unrelated to the Guatemalan germplasm present in Wondo Genet. To fully comprehend the origins of southern Ethiopian avocado germplasm, comparisons with other sources are essential.

Significant genetic diversity in both Guatemalan and Ethiopian avocado germplasm, highlights their potential for breeding programs. Identifying three ecotypes and their origins provides a basis for future research, emphasizing the importance of diverse genotypes for resilient avocado cultivation in changing climates.

8 Application and practical implications of the research findings

The research on the variability of native Guatemalan avocado germplasm has several practical applications and uses that can benefit various stakeholders and industries.

Political frameworks, nature conservation elements, and certification schemes: Politicians and policymakers play a pivotal role in the conservation and sustainable use of plant genetic resources. The research findings should inform the development of policies that encourage and support germplasm conservation efforts. Nature conservation elements, integrated into agricultural policies, can enhance the protection of avocado diversity by safeguarding natural habitats and promoting sustainable cultivation practices. Certification schemes can be designed to recognize and incentivize farmers and stakeholders engaged in the preservation of native avocado germplasm, thus contributing to the broader goals of biodiversity conservation.

Germplasm conservation and genetic resource management: The study's findings have practical implications for germplasm conservation efforts. The high level of genetic diversity identified within the native Guatemalan avocado germplasm highlights the importance of preserving these diverse genetic resources. Conservation initiatives can now focus on collecting and conserving avocado germplasm from different populations and regions to maintain the genetic diversity and ensure the long-term survival of valuable genetic traits. Preserving the native germplasm ensures that breeders and researchers have access to a broad range of genetic material for future breeding efforts, making avocados more resilient to environmental challenges, pests, and diseases.

Avocado breeding and cultivar development: One of the primary practical uses of this research is in avocado breeding programs. The significant diversity observed in both quantitative and qualitative traits provides breeders with valuable information for selecting and developing improved avocado cultivars. Breeders can now focus on traits such as fruit weight, length, shape, flesh texture, and seed form to develop cultivars with desired characteristics, such as larger fruits, better texture, or specific shapes. This research allows breeders to make informed decisions about which traits to prioritize and select for in their breeding programs, leading to the development of new avocado cultivars that meet consumer preferences and market demands.

Cultural preservation and historical significance: The research provides invaluable insights into the historical and cultural importance of avocados in Guatemala. To translate this knowledge into tangible actions, initiatives such as the creation of

reserves, involvement of indigenous communities, and the establishment of a labeling system for wild-collected products/foods could be considered. Creating reserves dedicated to native avocado landraces ensures the protection of their unique genetic heritage. Involving indigenous communities in cultivation practices not only safeguards traditional knowledge but also fosters sustainable agricultural traditions. Implementing a labeling system for products derived from wild-collected avocados not only adds value to these items but also raises awareness about their cultural significance, contributing to a sense of identity and pride among local communities.

Research methodology advancement: The research contributes to the advancement of research methodologies in the field of plant genetics and breeding. The use of molecular markers, such as AFLP and SSR markers, combined with statistical analyses like FAMD and HCPC, provides a comprehensive approach to studying genetic diversity and population structure. The methodology developed in this study can be applied to other crop species to assess genetic resources, understand diversity patterns, and inform breeding and conservation programs. This research methodology can help researchers and breeders gain insights into the genetic diversity of various crops, facilitating the development of improved cultivars and the conservation of valuable genetic resources.

Continuation of core collection: The core collection's establishment marks a crucial stride in preserving and leveraging native Guatemalan avocado germplasm. While currently embedded within their native habitats, these identified trees offer a promising path for future research and conservation. A practical next step involves revisiting these earmarked trees, collecting vegetable material, and establishing them in dedicated spaces, like research institutes. This proactive measure ensures the long-term conservation of these genetic treasures, actively safeguarding them for sustained and future agricultural pursuits.

In conclusion, the practical use of the research results on the variability of native Guatemalan avocado germplasm is diverse and far-reaching. The findings have practical applications in germplasm conservation and genetic resource management, avocado breeding and cultivar development, cultural preservation and historical significance, and the advancement of research methodologies. By leveraging this knowledge and applying it in various sectors, stakeholders can enhance avocado production, preserve genetic diversity, promote cultural heritage, and contribute to sustainable agriculture and food security.

9 References

Abraham J, Abraham J, Takrama JF. 2018. Morphological characteristics of avocado (*Persea americana* Mill.) in Ghana. African Journal of Plant Science **12**:88–97.

Abraham JD, Takrama JF. 2014. Genetic characterisation of avocado (*Persea americana* Mill.) in two regions of Ghana. African Journal of Biotechnology **13**:4620–4627.

Acosta-Díaz E, Álvarez-Ojeda MG, Rodríguez-Guerra R, Almeyda-León IH. 2020. Variabilidad morfológica y genética del germoplasma de aguacate criollo en Nuevo León, México. Biotecnología & Sustentabilidad **5**:91–108.

Adamack AT, Gruber B. 2014. PopGenReport: simplifying basic population genetic analyses in R. Methods in Ecology and Evolution **5**:384–387.

Adhikari S, Saha S, Biswas A, Rana TS, Bandyopadhyay TK, Ghosh P. 2017. Application of molecular markers in plant genome analysis: a review. The Nucleus **60**:283–297.

Adjei PY, Banful BK, Idun IA. 2011. Seed size and seed cut-length effects on germination behavior and seedling growth of avocado (*Persea americana* Mill.). International Journal of Agricultural Research **6**:299–305.

Agarwal M, Shrivastava N, Padh H. 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Reports **27**:617–631.

Agrama HA, Yan W, Lee F, Fjellstrom R, Chen M-H, Jia M, McClung A. 2009. Genetic assessment of a mini-core subset developed from the USDA rice genebank. Crop Science **49**:1336–1346.

Agre P, Asibe F, Darkwa K, Edemodu A, Bauchet G, Asiedu R, Adebola P, Asfaw A. 2019. Phenotypic and molecular assessment of genetic structure and diversity in a panel of winged yam (*Dioscorea alata*) clones and cultivars. Scientific Reports **9**:18221.

Ahsan MU, Hayward A, Alam M, Bandaralage JH, Topp B, Beveridge CA, Mitter N. 2019. Scion control of miRNA abundance and tree maturity in grafted avocado. BMC Plant Biology **19**:382.

Alberti MF, Brogio B do A, Silva SR da, Cantuarias-Avilés T, Fassio C. 2018. Avances en la propagación del aguacate. Revista Brasileira de Fruticultura **40**:1–18.

Alcaraz ML, Hormaza J. 2019. Reproductive biology of avocado (*Persea americana*). Acta Horticulturae **1231**:23–28.

Alcaraz ML, Hormaza JI. 2007. Molecular characterization and genetic diversity in an

avocado collection of cultivars and local Spanish genotypes using SSRs. Hereditas **144**:244–253.

Alcaraz ML, Hormaza JI. 2009. Avocado pollination and fruit Set - A perspective from Spain. California Avocado Association Yearbook **92**:113–135.

Alcaraz ML, Hormaza JI. 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**:354–361.

Alcaraz ML, Hormaza JI. 2014. Optimization of controlled pollination in avocado (*Persea americana* Mill., Lauraceae). Scientia Horticulturae **180**:79–85.

Alcaraz ML, Hormaza JI. 2021. Fruit set in avocado: pollen limitation, pollen load size, and selective fruit abortion. Agronomy **11**:1603.

Alcaraz ML, Hormaza JI, Rodrigo J. 2013. Pistil starch reserves at anthesis correlate with final flower fate in avocado (*Persea americana*). PLOS ONE **8**:e78467.

Allendorf FW, Funk WC, Aitken SN, Byrne M, Luikart G. 2022. Phenotypic variation in natural populations. Pages 19-38 in Allendorf FW, Funk WC, Aitken SN, Byrne M, Luikart G, Antunes A, editors. Conservation and the Genomics of Populations. Oxford University Press.

Alseekh S, Kostova D, Bulut M, Fernie AR. 2021. Genome-wide association studies: assessing trait characteristics in model and crop plants. Cellular and molecular life sciences **78**:5743–5754..

Alves AA, Bhering LL, Rosado TB, Laviola BG, Formighieri EF, Cruz CD. 2013. Joint analysis of phenotypic and molecular diversity provides new insights on the genetic variability of the Brazilian physic nut germplasm bank. Genetics and Molecular Biology **36**:371–381.

Amiteye S. 2021. Basic concepts and methodologies of DNA marker systems in plant molecular breeding. Heliyon **7**:e08093.

Arpaia ML, Collin S, Sievert J, Obenland D. 2018. 'Hass' avocado quality as influenced by temperature and ethylene prior to and during final ripening. Postharvest Biology and Technology **140**:76–84.

Asfaw Z, Ågren GI. 2007. Farmers' local knowledge and topsoil properties of agroforestry practices in Sidama, Southern Ethiopia. Agroforestry Systems **71**:35–48.

Ashman TL et al. 2004. Pollen limitation of plant reproduction: Ecological and evolutionary causes and consequences. Ecology **85**:2408–2421.

Ashton OBO, Wong M, McGhie TK, Vather R, Wang Y, Requejo-Jackman C, Ramankutty P, Woolf AB. 2006. Pigments in avocado tissue and oil. Journal of Agricultural and Food Chemistry **54**:10151–10158.

Ashworth VETM, Clegg MT. 2003. Microsatellite markers in avocado (*Persea americana* Mill.): Genealogical relationships among cultivated avocado genotypes. Journal of Heredity **94**:407–415.

Ashworth VETM, Kobayashi MC, Cruz MD La, Clegg MT. 2004. Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. Scientia Horticulturae **101**:255–267.

Avendaño-Arrazate CH, Campos-Rojas E, Andres Agustin J, Espinosa-Zaragoza S, Sandoval Cancino G. 2019. Crop wild relatives of avocado, mamey sapote, and cherimoya. Pages 387–415 in Greene SL, Williams KA, Khoury CK, Kantar MB, Marek LF, editors. North American Crop Wild Relatives, Volume 2: Important Species. Springer International Publishing, Cham.

Awachare C, Karunakaran G, Madhavi M, Sakthivel T. 2023. Studies on morphological characterization of 72 avocado (*Persea Americana* Mill.) accessions. The Pharma Innovation Journal **12**:1970–1975.

Ayala-Silva T, Ledesma N. 2014. Avocado history, biodiversity and production. Sustainable Horticultural Systems **2**:157–205.

Bangar SP, Dunno K, Dhull SB, Kumar Siroha A, Changan S, Maqsood S, Rusu AV. 2022. Avocado seed discoveries: Chemical composition, biological properties, and industrial food applications. Food chemistry: X **16**:100507.

Bardakci Fİ. 2001. Random amplified polymorphic DNA (RAPD) markers. Turkish Journal of Biology **25**:185–196.

Barrera-Redondo J, Piñero D, Eguiarte LE. 2020. Genomic, transcriptomic and epigenomic tools to study the domestication of plants and animals: A field guide for beginners. Frontiers in Genetics **11**:742

Barrett DM, Beaulieu JC, Shewfelt R. 2010. Color, flavor, texture, and nutritional quality of fresh-cut fruits and vegetables: Desirable levels, instrumental and sensory measurement, and the effects of processing. Critical Reviews in Food Science and Nutrition **50**:369–389.

Ben-Ya'acov A, Barrientos A. 2003. The Persea germplasm resources potential as discovered during an international collection project. Proceedings V World Avocado Congress:21–26.

Ben-ya A, Bufler G, Agriculture M, Jose S, Rica C. 2003. the wild avocado of Monteverde , Costa Rica. Proceedings V World Avocado Congress:35–38.

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

Bergh B, Ellstrand N. 1986. Taxonomy of the avocado. California Avocado Society Yearbook **70**:135–146.

Bergh B, Scora RW, Storey WB. 1973. A comparison of leaf terpenes in *Persea* subgenus *Persea*. Botanical Gazette **134**:130–134.

Bill M, Sivakumar D, Thompson AK, Korsten L. 2014. Avocado fruit quality management during the postharvest supply chain. Food Reviews International **30**:169–202.

Blears MJ, De Grandis SA, Lee H, Trevors JT. 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. Journal of Industrial Microbiology and Biotechnology **21**:99–114.

Borrone J, Olano C, Kuhn D, Brown J, Schnell RJ, Violi H. 2008. Outcrossing in Florida avocados as measured using microsatellite markers. Journal of the American Society for Horticultural Science **133**:255–261.

Bost JB, Smith NJ., Crane J. 2013. History, distribution and uses. Pages 10–30 in Schaffer B, Wolstenholme B, Whiley AW, editors. The avocado: Botany, production and uses, 2nd edition. CAB Internatinoal.

Bougeard S, Dray S. 2018. Supervised multiblock analysis in R with the ade4 package. Journal of Statistical Software **86**:1–17.

Boza EJ, Tondo CL, Ledesma N, Campbell RJ, Bost J, Schnell RJ, Gutiérrez OA. 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.

Brouwer M, Blok de R. 2022. coreCollection: creating a core collection. Wageningen University and Research, department Plant Breeding, Wageningen, The Netherlands.

Brown AHD. 1989. Core collections: a practical approach to genetic resources management. Genome **31**:818–824.

Bullock EL, Nolte C, Segovia AR, Woodcock CE. 2020. Ongoing forest disturbance in Guatemala's protected areas. Remote Sensing in Ecology and Conservation **6**:141–152.

Burd M. 1994. Bateman's principle and plant reproduction: The role of pollen limitation

in fruit and seed set. The Botanical Review **60**:83–139.

Cañas-Gutiérrez GP, Alcaraz L, Hormaza JI, Arango-Isaza RE, Saldamando-Benjumea CI. 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**:437–449.

Cañas-Gutiérrez GP, Galindo-López LF, Arango-Isaza R, Saldamando-Benjumea CI. 2015. Genetic diversity of avocado (*Persea americana* Mill.) cultivars in Antioquia, Colombia. Agronomia Mesoamericana **26**:129–143.

Cañas-Gutiérrez GP, Sepulveda-Ortega S, López-Hernández F, Navas-Arboleda AA, Cortés AJ. 2022. Inheritance of yield components and morphological traits in avocado cv. Hass from "Criollo" "Elite Trees" via half-sib seedling rootstocks. Frontiers in Plant Science **13**:1–17.

Cao Z, Deng Z. 2020. Morphological, cytological and molecular marker analyses of 'Tapestry' caladium variants reveal diverse genetic changes and enable association of leaf coloration pattern loci with molecular markers. Plant Cell, Tissue and Organ Culture (PCTOC) **143**:363–375.

Cárceles Rodríguez B, Durán Zuazo VH, Franco Tarifa D, Cuadros Tavira S, Sacristan PC, García-Tejero IF. 2023. Irrigation alternatives for avocado (*Persea americana* Mill.) in the Mediterranean subtropical region in the context of climate change: A review. Agriculture **13**:1049.

Carr MK V. 2013. The water relations and irrigation requirements of avocado (*Persea americana* Mill.): A review. Experimental Agriculture **49**:256–278.

Cerda-Hurtado I, Ojeda-Zacarías M, Iracheta-Donjuan L, de la Cerda J, Torres-Castillo J, Gutiérrez-Díez A. 2015. Variabilidad genética de cultivo in vitro de aguacate raza mexicana. Revista Mexicana de Ciencias Agrícolas **6**:243–250.

Cervantes-Paz B, Yahia EM. 2021. Avocado oil: Production and market demand, bioactive components, implications in health, and tendencies and potential uses. Comprehensive Reviews in Food Science and Food Safety **20**:4120–4158.

Chanderbali A, Soltis D, Soltis P, Wolstenholme BN. 2013. Taxonomy and botany. Pages 31–50 The avocado: Botany, production and uses. CABI.

Chaudhary R, Maurya GK. 2019. Restriction Fragment length polymorphism. Pages 1–3 in Vonk J, Shackelford T, editors. Encyclopedia of Animal Cognition and Behavior. Springer International Publishing, Cham.

Chen H, Ashworth V, Xu S, Clegg M. 2007. Quantitative genetic analysis of growth rate

in avocado. Journal of the American Society for Horticultural Science **132**:691–696.

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**:56–65.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica **142**:169–196.

Colunga P, Zizumboo D. 2004. Domestication of plants in Maya lowlands. Economic Botany **58**:101–110.

Corrales-García JE, Del Rosario García-Mateos M, Martínez-López E, Barrientos-Priego AF, Ybarra-Moncada MC, Ibarra-Estrada E, Méndez-Zúñiga SM, Becerra-Morales D. 2019. Anthocyanin and oil contents, fatty acids profiles and antioxidant activity of Mexican landrace avocado fruits. Plant foods for human nutrition (Dordrecht, Netherlands) **74**:210–215.

Cruz-López DF, Caamal-Cauich I, Pat-Fernández VG, Reza Salgado J. 2022. Competitividad de las exportaciones de aguacate Hass de México en el mercado mundial. Revista Mexicana de Ciencias Agrícolas **13**:355–362.

Cruz-Maya ME, Barrientos-Priego AF, Zelaya-Molina LX, Rodríguez-De La O JL, Reyes-Alemán JC. 2018. Phylogenetic analysis of some members of the subgenus Persea (*Persea*, Lauraceae). Revista Chapingo, Serie Horticultura **24**:133–150.

Cuiris-Pérez H, Guillén-Andrade H, Pedraza-Santos M, López-Medina J, Vidales-Fernández I. 2009. Genetic variability within Mexican race avocado (*Persea americana* Mill.) germplasm collections determined by ISSRs. Revista Chapingo Serie Horticultura **15**:169–175.

Dabas D, Shegog RM, Ziegler GR, Lambert JD. 2013. Avocado (*Persea americana*) seed as a source of bioactive phytochemicals. Current pharmaceutical design **19**:6133–6140.

Davenport T. 1986. Avocado Flowering. Horticultural Reviews **8**:257–289.

Davis J, Henderson D, Kobayashi M, Clegg MT, Michael T, Allen PCK. 1998. Genealogical relationships among cultivated avocado as revealed through RFLP analyses. The Journal of Heredity **89**:319–323.

de Andrade EKV, de Andrade Júnior VC, de Laia ML, Fernandes JSC, Oliveira AJM, Azevedo AM. 2017. Genetic dissimilarity among sweet potato genotypes using morphological and molecular descriptors. Acta Scientiarum - Agronomy **39**:447–455.

De Beukelaer H, Davenport GF, Fack V. 2018. Core Hunter 3: flexible core subset selection. BMC Bioinformatics **19**:203.

De Beukelaer H, G D. 2023. Corehunter: Multi-purpose core subset selection. R package.

de Meeûs T. 2018. Revisiting FIS, FST, Wahlund effects, and null alleles. Journal of Heredity **109**:446–456.

de Meeûs T, McCoy KD, 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**:308–332.

Denvir A, Arima EY, González-Rodríguez A, Young KR. 2022. Ecological and human dimensions of avocado expansion in México: Towards supply-chain sustainability. Ambio **51**:152–166.

Devalaraja S, Jain S, Yadav H. 2011. Exotic fruits as therapeutic complements for diabetes, obesity and metabolic syndrome. Food Research International **44**:1856–1865.

Ding H, Chin Y-W, Kinghorn AD, D'Ambrosio SM. 2007. Chemopreventive characteristics of avocado fruit. Seminars in Cancer Biology **17**:386–394.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh tissue. Phytochemical Bulletin **19**:11–15.

Dreher ML, Cheng FW, Ford NA. 2021. A Comprehensive review of Hass avocado clinical trials, observational studies, and biological mechanisms. Nutrients **13**.

Dreher ML, Davenport AJ. 2013. Hass avocado composition and potential health effects. Critical Reviews in Food Science and Nutrition **53**:738–750.

Dutech C, Enjalbert J, Fournier E, Delmotte F, Barrès B, Carlier J, Tharreau D, Giraud T. 2007. Challenges of microsatellite isolation in fungi. Fungal Genetics and Biology **44**:933–949.

Eagles HA, Bariana HS, Ogbonnaya FC, Rebetzke GJ, Hollamby GJ, Henry RJ, Henschke PH, Carter M. 2001. Implementation of markers in Australian wheat breeding. Australian Journal of Agricultural Research **52**:1349–1356.

Ehleringer J, Björkman O, Mooney HA. 1976. Leaf pubescence: effects on absorptance and photosynthesis in a desert shrub. Science **192**:376–377.

Ernst WHO. 1999. Biomarkers in plants. Pages 135–151 in Peakall DB, Walker CH, Migula P, editors. Biomarkers: A pragmatic basis for remediation of severe pollution in eastern Europe. Springer Netherlands, Dordrecht.

Espinosa-Alonso LG, Paredes-López O, Valdez-Morales M, Oomah BD. 2017. Avocado

oil characteristics of Mexican creole genotypes. European Journal of Lipid Science and Technology **119**:1600406.

Estoup A, Jarne P, Cornuet J-M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology **11**:1591–1604.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology **14**:2611–2620.

Fedick SL, Morell-Hart S, Dussol L. 2023. Agriculture in the ancient Maya lowlands (Part 2): Landesque capital and long-term resource management strategies. Journal of Archaeological Research.

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

Figueroa DM, Bass HW. 2010. A historical and modern perspective on plant cytogenetics. Briefings in Functional Genomics **9**:95–102.

Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderegger R, Widmer A. 2017. Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri.* BMC Genomics **18**:69.

Floate KD, Whitham TG, Keim P. 1994. Morphological versus genetic markers in classifying hybrid plants. Evolution **48**:929–930.

Ford NA, Liu AG. 2020. The forgotten fruit: A case for consuming avocado within the traditional Mediterranean diet. Frontiers in Nutrition **7**.

Francis RM. 2017. pophelper: an R package and web app to analyse and visualize population structure. Molecular Ecology Resources **17**:27–32.

Franco J, Crossa J, Warburton ML, Taba S. 2006. Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. Crop Science **46**:854–864.

Fry NK, Savelkoul PHM, Visca P. 2009. Amplified fragment length polymorphism analysis. Pages 89–104 in Caugant DA, editor. Molecular Epidemiology of Microorganisms: Methods and Protocols. Humana Press, Totowa, NJ.

Furnier GR, Cummings MP, Clegg MT. 1990. Evolution of the avocados as revealed by DNA restriction fragment variation. Journal of Heredity **81**:183–188.

Galang FG. 1940. Avocado culture. The East African Agricultural Journal **5**:340–342.

Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics **31**:3718–3720.

Galindo-Tovar ME, Arzate-Fernández A. 2010a. Consideraciones sobre el origen y primera dispersión del aguacate Lauraceae. Cuadernos de Biodiversidad **33**:11–15.

Galindo-Tovar ME, Arzate-Fernández A. 2010b. West Indian avocado: where did it originate? Phyton **79**:203–207.

Galindo-Tovar ME, Arzate-Fernández AM, Ogata-Aguilar N, Landero-Torres I. 2007. The avocado (*Persea americana*, Lauraceae) crop in Mesoamerica: 10,000 years of history. Harvard Papers in Botany **12**:325–334.

Galindo-Tovar ME, Lee-Espinoza HE, Murgía-González J, Leyva-Ovalle OR, Landero-Torres I. 2013. Domesticación y distribución geográfica de *Persea americana* Mill. en la época precolombina. Revista de Geografía Agrícola **50**–**51**:65–70.

Galindo-Tovar ME, 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**:441–450.

Gama L, Gomez P. 1992. An ethnoecological approach for the study of Persea: a case study in the Maya area. Proceedings II World Avocado Congress:11–17.

Ge Y, Zang X, Tan L, Wang J, Liu Y, Li Y, Wang N, Chen D, Zhan R, Ma W. 2019. Single-molecule long-read sequencing of avocado generates microsatellite markers for analyzing the genetic diversity in avocado germplasm. Agronomy **9**:512.

Goebel T, Waters M, O´Rourke D. 2008. The late Pleistocene dispersal of modern humans in the Americas. Science **319**:1497–1502.

Golukcu M, Ozdemir F. 2010. Changes in phenolic composition of avocado cultivars during harvesting time. Chemistry of Natural Compounds **46**:112–115.

Goudet J. 1995. Computer program to calculate F-statistics. Journal of Heredity **86**:485–486.

Govindaraj M, Vetriventhan M, Srinivasan M. 2015. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genetics research international **2015**:431487.

Graebner R, Cuesta-Marcos A. 2016. GeneticSubsetter: Identify Favorable Subsets of Germplasm Collections. R package.

Gross-German E, Viruel MA. 2013. Molecular characterization of avocado germplasm with a new set of SSR and EST-SSR markers: Genetic diversity, population structure, and identification of race-specific markers in a group of cultivated genotypes. Tree

Genetics and Genomes **9**:539–555.

Gross J, Gabai M, Lifshitz A, Sklarz B. 1973. Carotenoids in pulp, peel and leaves of *Persea americana*. Phytochemistry **12**:2259–2263.

Guichoux E et al. 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources **11**:591–611.

Guidoti DT, Gonela A, Vidigal MCG, Conrado TV, Romani I. 2018. Interrelationship between morphological, agronomic and molecular characteristics in the analysis of common bean genetic diversity. Acta Scientiarum - Agronomy **40**:1–9.

Gustafson CD, Marsh AW, Branson RL, Davis S. 1979. Drip irrigation on avocados. California Avocado Association Yearbook **63**:95–134.

Gutiérrez-Díez A, Martínez-de la Cerda J, García-Zambrano EA, Iracheta-Donjuan L, Ocampo-Morales JD, Cerda-Hurtado IM. 2009. Estudio de diversidad genética del aguacate nativo en Nuevo León, México. Revista de Fitotecnia Mexicana. **32**:9–18.

Guzmán LF, Machida-Hirano R, Borrayo E, Cortés-Cruz M, Espíndola-Barquera M del C, Heredia García E. 2017a. Genetic structure and selection of a core collection for long term conservation of avocado in Mexico. Frontiers in Plant Science **8**:1–10.

Guzmán LF, Machida-hirano R, Borrayo E, Cortés-cruz M, Jarret RL. 2017b. Genetic structure and selection of a core collection for long term conservation of avocado in Mexico. Frontiers in Plant Science **8**:1–10.

Haberman A, Tsror Lahkim L, Lazare S, Hazanovsky M, Lebiush S, Zipori I, Busatn A, Simenski E, Dag A. 2020. Management of Verticillium Wilt of avocado using tolerant rootstocks. Plants **9**:531.

Henao-Rojas JC, Lopez JH, Osorio NW, Ramírez-Gil JG. 2019. Fruit quality in Hass avocado and its relationships with different growing areas under tropical zones. Revista Ceres **66**:341–350.

Hidalgo R. 2003. Variabilidad genética y caracterización de especies vegetales. Page 2.26 in Franco TL, Hidalgo R, editors. Análisis estadístico de datos de caracterización morfológica de recursos fitogenéticos. Instituto Internacional de Recursos Fitogenéticos (IPGRI), Colombia.

Hu J, Mei M, Jin F, Xu J, Duan S, Bian C, Li G, Wang X, Jin L. 2022. Phenotypic variability and genetic diversity analysis of cultivated potatoes in China. Frontiers in Plant Science **13**.

Hu J, Zhu J, Xu HM. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops.

Theoretical and Applied Genetics **101**:264–268.

Hubisz MJ, Falush D, Stephen M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology **9**:1322–1332.

IPGRI. Descriptors for Avocado *(Persea* spp.*)*; International Plant Genetic Resources Institute: Rome, Italy, 1995; 106p.

Ish-Am G, Eisikowitch D. 1993. The behaviour of honey bees (*Apis mellifera*) visiting avocado (*Persea americana*) flowers and their contribution to its pollination. Journal of Apicultural Research **32**:175–186.

Janick J. 2005. The origins of fruits, fruit growing, and fruit breeding. Plant Breeding Reviews **25**:255–321.

Jarvis P, Lister C, Szabo V, Dean C. 1994. Integration of CAPS markers into the RFLP map generated using recombinant inbred lines of *Arabidopsis thaliana*. Plant molecular biology **24**:685–687.

Jeffreys AJ, Neumann R, Wilson V. 1990. Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell **60**:473–485.

Jeffreys AJ, Wilson V, Thein SL. 1985. Hypervariable 'minisatellite' regions in human DNA. Nature **314**:67–73.

Jiang G-L. 2013. Molecular markers and marker-assisted breeding in plants. Pages 45–83 in Andersen SB, editor. Plant breeding from laboratories to fields. IntechOpen, Rijeka.

Jombart T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics **24**:1403–1405.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics **1**:94.

Juma I, Geleta M, Hovmalm HP, Nyomora A, Saripella GV, Carlsson AS, Fatih M, Ortiz R. 2021. Comparison of morphological and genetic characteristics of avocados grown in Tanzania. Genes **12**:63.

Juma I, Geleta M, Nyomora A, Saripella GV, Hovmalm HP, Carlsson AS, Fatih M, Ortiz R. 2020a. Genetic diversity of avocado from the southern highlands of Tanzania as revealed by microsatellite markers. Hereditas **157**.

Juma I, Nyomora A, Hovmalm HP, Fatih M, Geleta M, Carlsson AS, Ortiz RO. 2020b. Characterization of Tanzanian avocado using morphological traits. Diversity **12**:64.

Kaiser SA, Taylor SA, Chen N, Sillett TS, Bondra ER, Webster MS. 2017. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. Molecular Ecology Resources **17**:183–193.

Kamaraj M, Dhana Rangesh Kumar V, Nithya TG, Danya U. 2020. Assessment of antioxidant, antibacterial activity and phytoactive compounds of aqueous extracts of avocado fruit peel from Ethiopia. International Journal of Peptide Research and Therapeutics **26**:1549–1557.

Kämper W, Ogbourne SM, Hawkes D, Trueman SJ. 2021. SNP markers reveal relationships between fruit paternity, fruit quality and distance from a cross-pollen source in avocado orchards. Scientific Reports **11**:20043.

Kamvar ZN, Tabima JF, Gŕunwald NJ. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ **2**:e281.

Karaköy T, Baloch FS, Toklu F, Özkan H. 2014. Variation for selected morphological and quality-related traits among 178 faba bean landraces collected from Turkey. Plant Genetic Resources **12**:5–13.

Karam D, Westra P, Niessen SJ, Ward SM, Figueiredo JEF. 2006. Assessment of silver-stained AFLP markers for studying DNA polymorphism in proso millet (*Panicum miliaceum* L.). Revista Brasileira de Botanica **29**:609–615.

Karikari B, Lemay M-A, Belzile F. 2023. k-mer-based genome-wide association studies in plants: Advances, challenges, and perspectives. Genes **14**:1439.

Kassambara A, Mundt F. 2020. factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7.

Kaur V, Aravind J, Manju, Jacob SR, Kumari J, Panwar BS, Pal N, Rana JC, Pandey A, Kumar A. 2022. Phenotypic characterization, genetic diversity assessment in 6,778 accessions of barley (*Hordeum vulgare* L. ssp. *vulgare*) germplasm conserved in National Genebank of India and development of a core set. Frontiers in Plant Science **13**:771920.

Kebriyaee D, Kordrostami M, Rezadoost MH, Lahiji HS. 2012. QTL analysis of agronomic traits in rice using SSR and AFLP markers. Notulae Scientia Biologicae **4**:116–123.

Kenkel N. 2006. On selecting an appropriate multivariate analysis. Canadian Journal of Plant Science **86**:663–676.

Keuls M. 1952. The use of the "studentized range” in connection with an analysis of variance. Euphytica **1**:112–122.

Kim M-J, Hyun J-N, Kim J-A, Park J-C, Kim M-Y, Kim J-G, Lee S-J, Chun S-C, Chung I-M. 2007. Relationship between phenolic compounds, anthocyanins content and antioxidant activity in colored barley germplasm. Journal of Agricultural and Food Chemistry **55**:4802–4809.

Knight RJ, Campbell CW. 1999. Ecological adaptation and the evolution of modern avocado cultivars. Revista Chapingo serie horticultura **5**:49–54.

Konrad W, Burkhardt J, Ebner M, Roth-Nebelsick A. 2015. Leaf pubescence as a possibility to increase water use efficiency by promoting condensation. Ecohydrology **8**:480–492.

Kulkarni P, Paul R, Ganesh N. 2010. In vitro evaluation of genotoxicity of avocado (Persea americana) fruit and leaf extracts in human peripheral lymphocytes. Journal of environmental science and health. Part C, Environmental Carcinogenesis & Ecotoxicology Reviews **28**:172–187.

Kullback S, Leibler RA. 1951. On information and sufficiency. The Annals of Mathematical Statistics **22**:79–86.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution **35**:1547–1549.

Labate JA. 2000. Software for population genetic analyses of molecular marker data. Crop Science **40**:1521–1528.

Lachance J. 2009. Detecting selection-induced departures from Hardy-Weinberg proportions. Genetics Selection Evolution **41**:1–6.

Lahav E, Trochoulias T. 1982. The effect of temperature on growth and dry matter production of avocado plants. Australian Journal of Agricultural Research **33**:549–558.

Lander ES. 1996. The new genomics: Global views of biology. Science **274**:536–539.

Landon AJ. 2009. Domestication and significance of *Persea americana*, the avocado , in Mesoamerica. Nebraska Anthropologist **24**:62–79.

Larson BMH, Barret SCH. 2000. A comparative analysis of pollen limitation in flowering plants. Biological Journal of the Linnean Society **69**:503–520.

Lavi U, Sa'ada D, Ragev I, Lahav E. 2003. Avocado genetics and breeding - present and future. World Avocado Congress **42**:134–135.

Lê S, Josse J, Husson F. 2008. FactoMineR: An R package for multivariate analysis. Journal of Statistical Software **25**.

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**:243–246.

Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. Nucleic Acids Research **47**:W256–W259.

Levene H. 1960. Robust tests for equality of variances. Pages 278–292 in Olkin I, editor. Contributions to probability and statistics: Essays in honor of Harold Hotelling. Stanford University Press.

Li G, Quiros CF. 2001. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theoretical and Applied Genetics **103**:455–461.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular Ecology **11**:2453–2465.

Li YL, Liu JX. 2018. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources **18**:176–177.

Litt M, Luty JA. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American journal of human genetics **44**:397–401.

Liu M, Shen Y. 1998. Sierra Nevada uplift; a ductile link to mantle upwelling under the basin and range province. Geology **26**:299–302.

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**.

López-Galé Y, Murcia-Riaño N, Romero-Barrera Y, Martínez MF. 2022. Morphological characterization of seed-donor Creole avocado trees from three areas in Colombia. Revista Chapingo Serie Horticultura **28**:93–108.

López-Guzmán G, Medina-Torres R, Guillén-Andrade H, Ramírez-Guerrero LG, Juárez-López P, Ruelas-Hernández P. 2015. Caracterización morfológica en genotipos nativos de aguacate (*Persea americana* Mill.) de clima tropical en Nayarit, México. Revista Mexicana de Ciencias Agrícolas **11**: 2157–2163.

López-Guzmán GG, Palomino-Hermosillo YA, Balois-Morales R, Bautista-Rosales PU, Jiménez-Zurita JO. 2021. Genetic diversity of native avocado in Nayarit, Mexico,

determined by ISSRs. Ciencia Tecnologia Agropecuaria **22**:e1686.

Loupit G, Brocard L, Ollat N, Cookson SJ. 2023. Grafting in plants: recent discoveries and new applications. Journal of Experimental Botany **74**:2433–2447.

Lozi L, Anjichi V, Letting F, Were J. 2018. Effect of seed size on germination and seedling performance on grafted avocado. African Journal of Education, Science and Technology **4**:95–100.

Macneish RS. 1964. Ancient Mesoamerican civilization. Science **143**:531–537.

Maeda M, Uryu N, Murayama N, Ishii H, Ota M, Tsuji K, Inoko H. 1990. A simple and rapid method for HLA-DP genotyping by digestion of PCR-amplified DNA with allele-specific restriction endonucleases. Human immunology **27**:111–121.

Mahajan R, Bisht I, Dhillon B. 2007. Establishment of a core collection of world sesame (Sesamum indicum L.) germplasm accessions. SABRAO journal of breeding and genetics **39**:53–64.

Mahbou STG, Ntsomboh-Ntsefong G, Fanche AM, Tchio F, Dongmo F, Etoga GO, Youmbi E. 2022. Development of a technical itinerary for the production of avocado (*Persea americana* Mill.) seedlings with biofertilizers. American Journal of Plant Sciences **13**:1209–1226.

Mantel N. 1967. The detection of disease clustering and generalized regression approach. Cancer research **27**:209–220.

Marsh JI, Hu H, Gill M, Batley J, Edwards D. 2021. Crop breeding for a changing climate: integrating phenomics and genomics with bioinformatics. Theoretical and Applied Genetics **134**:1677–1690.

Meirmans PG. 2020. Genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. Molecular Ecology Resources **20**:1126–1131.

Meyer MD, Terry LA. 2010. Fatty acid and sugar composition of avocado, cv. Hass, in response to treatment with an ethylene scavenger or 1-methylcyclopropene to extend storage life. Food Chemistry **121**:1203–1210.

Michaels SD, Amasino RM. 1998. A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. The Plant journal : for cell and molecular biology **14**:381–385.

Millar C. 1996. Tertiary vegetation history. En: Sierra Nevada Ecosystem Proyect: Final report to Congress. Albany, California.

Mokria M, Gebrekirstos A, Said H, Hadgu K, Hagazi N, Dubale W, Bräuning A. 2022.

Fruit weight and yield estimation models for five avocado cultivars in Ethiopia. Environmental Research Communications **4**:0175013.

Morineau A. 1984. Note sur la caractérisation statistique d'une classe et les valeurs-tests. Bulletin Technique du Centre de Statistique et d'Informatique Appliqués **2**:20–27.

Mueller NG, Flachs A. 2022. Domestication, crop breeding, and genetic modification are fundamentally different processes: implications for seed sovereignty and agrobiodiversity. Agriculture and Human Values **39**:455–472.

Muñoz-Pajares AJ. 2013. SIDIER: substitution and indel distances to infer evolutionary relationships. Methods in Ecology and Evolution **4**:1195–1200.

Nadeem MA et al. 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. Biotechnology and Biotechnological Equipment **32**:261–285.

Nanjundan J, Aravind J, Radhamani J, Singh KH, Kumar A, Thakur AK, Singh K, Meena KN, Tyagi RK, Singh D. 2022. Development of Indian mustard [*Brassica juncea* (L.) Czern.] core collection based on agro-morphological traits. Genetic Resources and Crop Evolution **69**:145–162.

Ndjiondjop MN, Gouda AC, Eizenga GC, Warburton ML, Kpeki SB, Wambugu PW, Gnikoua K, Tia DD, Bachabi F. 2023. Genetic variation and population structure of Oryza sativa accessions in the AfricaRice collection and development of the AfricaRice O. sativa core collection. Crop Science **63**:724–739.

Nerdo N et al. 2009. Agro-morphologic traits , isoenzyme and DNA markers for estimating the polymorphism levels , discriminating capacity and informativeness in avocado. Revista CENIC Ciencias Biologicas **40**:63–74.

Newman D. 1939. the distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. Biometrika **31**:20–30.

Ninh TT, Doan HT, Nguyen MT, Tran TT, Dang PDD, Dinh ST, Nguyen HT, Nong HT, Nguyen CX. 2022. Genetic diversity of avocado (*Persea americana* Mill.) germplasm in Vietnam using RAPD and ISSR molecular markers. Australian Journal of Crop Science **16**:856–862.

Nirody BS. 1922. Investigations in avocado breeding. California Avocado Association **7**:65–78.

Nkansah GO, Ofosu-Budu KG, Ayarna AW. 2013. Genetic diversity among local and introduced avocado germplasm based on morpho-agronomic traits. International

Journal of Plant Breeding and Genetics **7**:76–91.

Nkhoma N, Shimelis H, Laing MD, Shayanowako A, Mathew I. 2020. Assessing the genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collections using phenotypic traits and SNP markers. BMC Genetics **21**:110.

Nyakang'i CO, Ebere R, Marete E, Arimi JM. 2023. Avocado production in Kenya in relation to the world, Avocado by-products (seeds and peels) functionality and utilization in food products. Applied Food Research **3**:100275.

Oboh G, Odubanjo VO, Bello F, Ademosun AO, Oyeleye SI, Nwanna EE, Ademiluyi AO. 2016. Aqueous extracts of avocado pear (*Persea americana* Mill.) leaves and seeds exhibit anti-cholinesterases and antioxidant activities in vitro. Journal of basic and clinical physiology and pharmacology **27**:131–140.

Odong TL, Jansen J, van Eeuwijk FA, van Hintum TJL. 2013. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. Theoretical and Applied Genetics **126**:289–305.

Oliveira M, Azevedo L. 2022. Molecular markers: An overview of data published for fungi over the last ten years. Journal of Fungi **8**:803.

Ozdemir F, Topuz A. 2004. Changes in dry matter, oil content and fatty acids composition of avocado during harvesting time and post-harvesting ripening period. Food Chemistry **86**:79–83.

Pages J. 2004. Analyse factorielle de données mixtes. Revue de statistique appliquée **52**:93–111.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics **35**:526–528.

Patil I. 2021. Visualizations with statistical details: The "ggstatsplot" approach. Journal of Open Source Software **6**:3167.

Paull RE, Duarte O. 2011. Avocado. Pages 153–184 Crop Production Science in Horticulture. CABI International.

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.

Pereira MEC, Tieman DM, Sargent SA, Klee HJ, Huber DJ. 2013. Volatile profiles of ripening West Indian and Guatemalan-West Indian avocado cultivars as affected by aqueous 1-methylcyclopropene. Postharvest Biology and Technology **80**:37–46.

Phogat BS et al. 2021. Characterization of wheat germplasm conserved in the Indian National Genebank and establishment of a composite core collection. Crop Science

61:604–620.

Pickersgill B. 2007. Domestication of plants in the Americas: Insights from Mendelian and molecular genetics. Annals of Botany **100**:925–940.

Pino JA, Marbot R, Martí MP. 2006. Leaf oil of P*ersea americana* Mill. var. drymifolia cv. Duke grown in Cuba. Journal of Essential Oil Research **18**:440–442.

Poczai P, Varga I, Laos M, Cseh A, Bell N, Valkonen JPT, Hyvönen J. 2013. Advances in plant gene-targeted and functional markers: a review. Plant Methods **9**:6.

Popenoe W. 1935. Origin of the cultivated races of avocados. California Avocado Association Yearbook **20**:184–194.

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

Puechmaille SJ. 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**:608–627.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramesh P, Mallikarjuna G, Sameena S, Kumar A, Gurulakshmi K, Reddy BV, Reddy PCO, Sekhar AC. 2020. Advancements in molecular marker technologies and their applications in diversity studies. Journal of Biosciences **45**:123.

Ramírez IM, Fuentes JL, Rodríguez NN, Coto O, Cueto J, Becker D, Rohde W. 2005. Diversity analysis of Cuban avocado varieties based on agro-morphological traits and DNA polymorphisms. Journal of Genetics and Breeding **59**:241–252.

Ramos-Aguilar AL, Ornelas-Paz J, Tapia-Vargas LM, Gardea-Bejar AA, Yahia EM, Ornelas-Paz J de J, Perez-Martinez JD, Rios-Velasco C, Escalante-Minakata P. 2021. Metabolomic analysis and physical attributes of ripe fruits from Mexican Creole (Persea americana var. Drymifolia) and "Hass" avocados. Food Chemistry **354**:129571.

Ranjitha V, Chaitanya H, Ravi C, Shivakumar B, Naveen N. 2021. Morphological characterization of avocado (*Persea americana* Mill.) accessions explored from hill zone taluks of Chikkamagaluru district, Karnataka state. Journal of Pharmacognosy and Phytochemistry **10**:310–318.

Razeto B, Longueira J, Fichet T. 1992. Close Planting of Avocado. Proceedings of Second World Avocado Congress:273–279.

Ražná K, Žiarovská J, Gálová Z. 2019. MicroRNA-Based markers in Plant genome

response to abiotic stress and their application in plant genotyping. Tutar L, Aras S, Tutar E, editors. IntechOpen, Rijeka.

Reddy LJ, Upadhyaya HD, Gowda CLL, Singh S. 2005. Development of core collection in Pigeonpea [*Cajanus cajan* (L.) Millspaugh] using geographic and qualitative morphological descriptors. Genetic Resources and Crop Evolution **52**:1049–1056.

Renner SS. 1999. Circumscription and phylogeny of the Laurales : evidence from molecular and morphological data. American Journal of Botany **86**:1301–1315.

Reyes-Alemán JC, Valadez-Moctezuma E, Simuta-Velázco L, Barrientos-Priego A, Gallegos-Vázquez C. 2013. Distinción de especies del género Persea mediante RAPD e ISSR de ADN. Revista Mexicana de Ciencias Agrícolas **4**:517–529.

Reyes-Herrera PH, Muñoz-Baena L, Velásquez-Zapata V, Patiño L, Delgado-Paz OA, Díaz-Diez CA, Navas-Arboleda AA, Cortés AJ. 2020. Inheritance of Rootstock Effects in Avocado (*Persea americana* Mill.) cv. Hass. Frontiers in Plant Science **11**:555071.

Rincón-Hernández CA, Sánchez-Pérez J, Espinosa-García FJ. 2011. Caracterización química foliar de los árboles de aguacate criollo (*Persea americana* var. drymifolia) en los bancos de germoplasma de Michoacán, México. Revista Mexicana de Biodiversidad **82**:395–412.

Rincon-Patino J, Lasso E, Corrales J. 2018. Estimating avocado sales using machine learning algorithms and weather data. Sustainability **10**:3498.

Rubinstein M, Eshed R, Rozen A, Zviran T, Kuhn DN, Irihimovitch V, Sherman A, Ophir R. 2019. Genetic diversity of avocado (*Persea americana* Mill.) germplasm using pooled sequencing. BMC Genomics **20**:1–10.

Ruiz-Chután JA, Berdúo-Sandoval JE, Kalousová M, Fernández E, Žiarovská J, Sánchéz-Pérez A, Lojka B. 2022. SSRs markers reveal high genetic diversity and limited differentiation among populations of native Guatemalan avocado. Journal of Microbiology, Biotechnology and Food Sciences **12**:e6134.

Salazar JA, Rasouli M, Moghaddam RF, Zamani Z, Imani A, Martínez-Gómez P. 2014. Low-cost strategies for development of molecular markers linked to agronomic traits in Prunus. Agricultural Sciences **5**:430–439.

Salgotra RK, Chauhan BS. 2023. Genetic diversity, conservation, and utilization of plant genetic resources.Genes **14**:174

Sandoval-Castro E, Peraza-Magallanes AY, Dodd RS, Ashworth VETM, Cruz-Mendívil A, Calderón-Vázquez CL. 2021. Exploring genetic diversity of lowland avocado

(*Persea americana* Mill.) as a genetic reservoir for breeding. Genetic Resources and Crop Evolution **68**:2757–2766.

Sartie A, Asiedu R, Franco J. 2012. Genetic and phenotypic diversity in a germplasm working collection of cultivated tropical yams (*Dioscorea* spp.). Genetic Resources and Crop Evolution **59**:1753–1765.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. Nature Reviews Genetics **14**:807–820.

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

Schieber E, Bergh B. 1987. *Persea zentmyerii*: A new species from Guatemala. California Avocado Association Yearbook **71**:199–203.

Schnell RJ, Brown JS, Olano CT, Power EJ, Krol CA, Kuhn DN, Motamayor JC. 2003. Evaluation of avocado germplasm using microsatellite markers. Journal of the American Society for Horticultural Science **128**:881–889.

Schnell RJ, Tondo CL, Brown JS, Kuhn DN, Ayala-Silva T, Borrone JW, Davenport TL. 2009. Outcrossing between 'Bacon' pollinizers and adjacent 'Hass' avocado trees and the description of two new lethal mutants. HortScience horts **44**:1522–1526.

Schröeder C. 1968. Prehistoric avocados in California. California Avocado Association Yearbook **52**:29–34.

Scora R, Bergh B. 1990. The origins and taxonomy of avocado (Persea americana) Mill. Lauraceae. Acta Horticulturae **275:** 387-394

Scora R, Bergh B. 1992. Origin of the taxonomic relationships within the genus Persea. Pages 505–514 Proceedings II World Avocado Congress. California.

Shannon CE. 1948. A mathematical theory of communication. The Bell System Technical Journal **27**:379–423.

Sharma S, Schulthess AW, Bassi FM, Badaeva ED, Neumann K, Graner A, Özkan H, Werner P, Knüpffer H, Kilian B. 2021. Introducing beneficial alleles from plant genetic resources into the wheat germplasm. Biology **10**:982.

Sharon D, Cregan PB, Mhameed S, Kusharska M, Hillel J, Lahav E, Lavi U. 1997. An integrated genetic linkage map of avocado. TAG Theoretical and Applied Genetics **95**:911–921.

Shewfelt RL. 2000. Fruit and vegetable quality. Pages 160–173 in Shewfelt RL, Bruckner B, editors. Fruit and Vegetable Quality: An Integrated View; CRC Press.

Simionca Mărcășan LI, Oltean I, Popa S, Plazas M, Vilanova S, Gramazio P, Sestras AF,

Prohens J, Sestras RE. 2023. Comparative analysis of phenotypic and molecular data on response to main pear diseases and pest attack in a germplasm collection. International Journal of Molecular Sciences **24**: 6239.

Singh A, Behera C. 2022. Strategies, opportunities, and challenges in crop genetic diversity conservation: A plant breeder's perspective. Pages 151–169 in Kumar A, Choudhury B, Dayanandan S, Khan ML, editors. Molecular Genetics and Genomics Tools in Biodiversity Conservation. Springer Nature Singapore, Singapore.

Singh SP, Nodari R, Gepts P. 1991. Genetic diversity in cultivated common bean: I. Allozymes. Crop Science **31**:19–31.

Slatkin M, Barton NH. 1989. A comparison of three indirect methods for estimating average levels of gene flow. Evolution **43**:1349–1368.

Sneath PHA, Sokal RR. 1973. Numerical taxonomy: The principles and practice of numerical classification. W. H. Freeman and Co.

Solares E, Morales-Cruz A, Balderas RF, Focht E, Ashworth VETM, Wyant S, Minio A, Cantu D, Arpaia ML, Gaut BS. 2023. Insights into the domestication of avocado and potential genetic contributors to heterodichogamy. G3 Genes|Genomes|Genetics **13**:jkac323.

Sommaruga R, Eldridge HM. 2021. Avocado Production: Water Footprint and Socio-economic Implications. EuroChoices **20**:48–53.

Spaniolas S, May ST, Bennett MJ, Tucker GA. 2006. Authentication of coffee by means of PCR-RFLP analysis and lab-on-a-chip capillary electrophoresis. Journal of Agricultural and Food Chemistry **54**:7466–7470.

Stadhouders R, Pas SD, Anber J, Voermans J, Mes THM, Schutten M. 2010. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5′ nuclease assay. Journal of Molecular Diagnostics **12**:109–117.

Stern RA, Rozen A, Eshed R, Zviran T, Sisai I, Sherman A, Irihimovitch V, Sapir G. 2021. Bumblebees (*Bombus terrestris*) improve "Hass" Avocado (*Persea americana*) pollination. Plants **10**:1372.

Storey W, Bergh B, Zentmyer G. 1986. The origin, indigenous range and dissemination of the avocado. California Avocado Society **70**:127–133.

Stout A. 1923. A study in cross-pollination of avocados in southern California. California Avocado Association Annual Report **8**:29–45.

Subirana I, Sanz H, Vila J. 2014. Building bivariate tables: The compareGroups package for R. Journal of Statistical Software **57**:1–16.

Sunil N, Sujatha M, Kumar V, Vanaja M, Basha SD, Varaprasad KS. 2011. Correlating the phenotypic and molecular diversity in Jatropha curcas L. Biomass and Bioenergy **35**:1085–1096.

Swarup S, Cargill EJ, Crosby K, Flagel L, Kniskern J, Glenn KC. 2021. Genetic diversity is indispensable for plant breeding to improve crops. Crop Science **61**:839–852.

Talavera A, Soorni A, Bombarely A, Matas AJ, Hormaza JI. 2019. Genome-Wide SNP discovery and genomic characterization in avocado (*Persea americana* Mill.). Scientific Reports **9**.

Tesfaye T, Ayele M, Gibril M, Ferede E, Limeneh DY, Kong F. 2022. Beneficiation of avocado processing industry by-product: A review on future prospect. Current Research in Green and Sustainable Chemistry **5**:100253.

Thachuk C, Crossa J, Franco J, Dreisigacker S, Warburton M, Davenport GF. 2009. Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. BMC Bioinformatics **10**:243.

Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to understand local adaptation. Trends in Ecology & Evolution **29**:673–680.

Tomlinson PB. 1980. The biology of trees native to tropical Florida. Systematic Botany **6**:311.

Tsykun T, Rellstab C, Dutech C, Sipos G, Prospero S. 2017. Comparative assessment of SSR and SNP markers for inferring the population genetic structure of the common fungus Armillaria cepistipes. Heredity **119**:371–380.

Unlu NZ, Bohn T, Clinton SK, Schwartz SJ. 2005. Carotenoid absorption from salad and salsa by humans is enhanced by entrye addition of avocado or avocado oil. The Journal of Nutrition **135**:431–436.

Uzun A, Yesiloglu T, Aka-Kacar Y, Tuzcu O, Gulsen O. 2009. Genetic diversity and relationships within Citrus and related genera based on sequence related amplified polymorphism markers (SRAPs). Scientia Horticulturae **121**:306–312.

Van der Heyden H, Dutilleul P, Brodeur L, Carisse O. 2014. Spatial distribution of single-nucleotide polymorphisms related to fungicide resistance and implications for sampling. Phytopathology **104**:604–613.

Van der Werff H. 2002. A synopsis of *Persea* (Lauraceae) in Central America. Novon **12**:575–586.

van Rooyen Z, Bower JP. 2006. Effects of storage temperature, harvest date and fruit origin on post-harvest physiology and the severity of mesocarp discolouration in

"Pinkerton" avocado (*Persea americana* Mill.). Journal of Horticultural Science and Biotechnology **81**:89–98.

Vekemans X, Beauwens M, Lemaire M, Ruiz-Roldán I. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size X. Molecular Ecology **11**:139–151.

Vieira MLC, Santini L, Diniz AL, Munhoz C de F. 2016. Microsatellite markers: What they mean and why they are so useful. Genetics and Molecular Biology **39**:312–328.

Villa-Rodríguez JA, Molina-Corral FJ, Ayala-Zavala JF, Olivas GI, González-Aguilar GA. 2011. Effect of maturity stage on the content of fatty acids and antioxidant activity of 'Hass' avocado. Food Research International **44**:1231–1237.

Vinu V, Singh N, Vasudev S, Yadava DK, Kumar S, Naresh S, Bhat SR, Prabhu KV. 2013. Assessment of genetic diversity in Brassica juncea (Brassicaceae) genotypes using phenotypic differences and SSR markers. Revista de biologia tropical **61**:1919–1934.

Vos P et al. 1995. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research **23**:4407–4414.

Wakabayashi J, Sawyer T. 2001. Stream incision, tectonics, uplift, and evolution of topography of the Sierra Nevada, California. Journal of Geology **109**:539–562.

Wang J. 2017. The computer program structure for assigning individuals to populations: easy to use but easier to misuse. Molecular Ecology Resources **17**:981–990.

Wang L, Tao L, Hao L, Stanley TH, Huang K-H, Lambert JD, Kris-Etherton PM. 2020. A moderate-fat diet with one avocado per day increases plasma antioxidants and decreases the oxidation of small, dense ldl in adults with overweight and obesity: A randomized controlled trial. The Journal of Nutrition **150**:276–284.

Wangithi C, Muriithi BW, Diiro G, Dubois T, Mohamed S, Lattorff MG, Ngowi B V, Abdel-Rahman EM, Adan M, Kassie M. 2022. Synergies of integrated pest and pollinator management in avocado farming in East Africa: An ex-ante economic analysis. PLOS ONE **17**:e0271241.

Wattier R, Engel CR, 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**:1569–1573.

Weiland JJ, Yu MH. 2003. A cleaved amplified polymorphic sequence (CAPS) marker

associated with Root-Knot nematode resistance in sugarbeet. Crop Science **43**:1814–1818.

Welsh J, McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research **18**:7213–7218.

Wienk R, Mostert-O'Neill M, Abeysekara N, Manosalva P, Freeman B, van den Berg N. 2022. Genetic diversity, population structure, and clonal verification in South African avocado cultivars using single nucleotide polymorphism (SNP) markers. Tree Genetics and Genomes **18**.

Wilcoxon F. 1945. Individual comparisons by ranking methods. Biometrics Bulletin **1**:80–83.

Wilk MB, Gnanadesikan R. 1968. Probability plotting methods for the analysis of data. Biometrika **55**:1–17.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey S V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research **18**:6531–6535.

Williams L. 1977. The avocados, a synopsis of the genus *Persea*, subg. *Persea*. Economic Botany **31**:315–320.

Wiradona I. 2017. Using avocado leaves to prevent Streptococcus viridans' inhibition in growing caries in mouth. ARC Journal of Surgery **3**:5–7.

Wolstenholme BN, Whiley AW. 1999. Ecophysiology of the avocado (*Persea americana* Mill.) tree as a basis for pre-harvest management. Revista Chapingo Serie Horticultura **5**:77–78.

Wotzlaw A, Speckenmeyer E, Porschen S. 2012. Generalized k-ary tanglegrams on level graphs: A satisfiability-based approach and its evaluation. Discrete Applied Mathematics **160**:2349–2363.

Wright S. 1931. Evolution in mendelian populations. Genetics **16**:97–149.

Yang W, Kang X, Yang Q, Lin Y, Fang M. 2013. Review on the development of genotyping methods for assessing farm animal diversity. Journal of Animal Science and Biotechnology **4**:2.

Zabeau M, Vos P. 1993. Selective restriction fragment amplification: A general method for DNA fingerprinting. European patent application no. 92402629.7.

Zhao Z, Fu Y-X, Hewett-Emmett D, Boerwinkle E. 2003. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. Gene **312**:207–213.

Zimmerman SJ, Aldridge CL, Oyler-McCance SJ. 2020. An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. BMC Genomics **21**:382.

10 Apendices

List of Apendices

Table A1. Conditions for each multiplex PCR.

Multiplex	Annealing temperature	Final concentration	Primer name	Size range	Fluorescent label
M1	63.4 °C	0.3 µM	AVAG05	83-125	6-FAM
			AVAG13	96-160	VIC
			AVT436	152	NED
		1.2 µM	AVAG11	105-161	PET
M2	57.6 °C	0.25 µM	AVAG21	158-221	VIC
			AVAG22	103-137	NED
			AVAG25	96-140	PET
		0.5 µM	AVAG07	98-114	6-FAM
			AUCR418	379	VIC
		0.7 µM	AVMIX04	160-194	6-FAM
M3	65 °C	0.5 µM	AVD001	208-267	NED
		1.0 µM	AVD022	220-258	PET

Table A2. Table of null allele frequency

Estimator	AVAG05	AVAG11	AVAG13	AVT436	AUCR418	AVAG07	AVAG21	AVAG22	AVAG25	AVD001	AVD022	AVMIX04
Observed frequency	0.226	0.251	0.115	0.218	0.188	0.208	0.102	0.071	0.220	0.163	0.224	0.113
Median frequency	0.222	0.249	0.113	0.216	0.185	0.206	0.101	0.069	0.219	0.161	0.222	0.110
2.5th percentile	0.167	0.187	0.075	0.160	0.134	0.145	0.059	0.037	0.161	0.116	0.163	0.073
97.5th percentile	0.280	0.312	0.155	0.282	0.241	0.273	0.148	0.109	0.276	0.213	0.279	0.156

Table A3. Selected genotypes for avocado core collection.

Genotype	Population	Region	Genetic Cluster	Genotype	Population	Region	Genetic Cluster
Chi.71	Chi	Central	02	To-Qui.101	To-Qui	Western	03
Chi.74	Chi	Central	01	To-Qui.105	To-Qui	Western	01
Chi.83	Chi	Central	02	To-Qui.121	To-Qui	Western	01
Chi.88	Chi	Central	01	To-Qui.128	To-Qui	Western	03
Sac.11	Sac	Central	02	To-Qui.131	To-Qui	Western	02
Sac.13	Sac	Central	01	To-Qui.134	To-Qui	Western	03
Sac.18	Sac	Central	01	Hue-Qui.137	Hue-Qui	Western	02
Sac.23	Sac	Central	02	Hue-Qui.143	Hue-Qui	Western	03
Sac.26	Sac	Central	02	Hue-Qui.147	Hue-Qui	Western	03
Sac.29	Sac	Central	02	Hue-Qui.152	Hue-Qui	Western	03
Sac.31	Sac	Central	03	Hue-Qui.157	Hue-Qui	Western	01
Sac.33	Sac	Central	03	AV.174	AV	Northern	03
Sac-Chi.39	Sac-Chi	Central	02	AV.177	AV	Northern	02
Sac-Chi.41	Sac-Chi	Central	03	AV.176	AV	Northern	01
Sac-Chi.46	Sac-Chi	Central	03	AV.182	AV	Northern	03
Sac-Chi.56	Sac-Chi	Central	01	AV.185	AV	Northern	03
Sac-Chi.64	Sac-Chi	Central	01	BV.159	BV	Northern	03
Sol.91	Sol	Central	03	BV.161	BV	Northern	02
Sol.97	Sol	Central	01	BV.164	BV	Northern	01

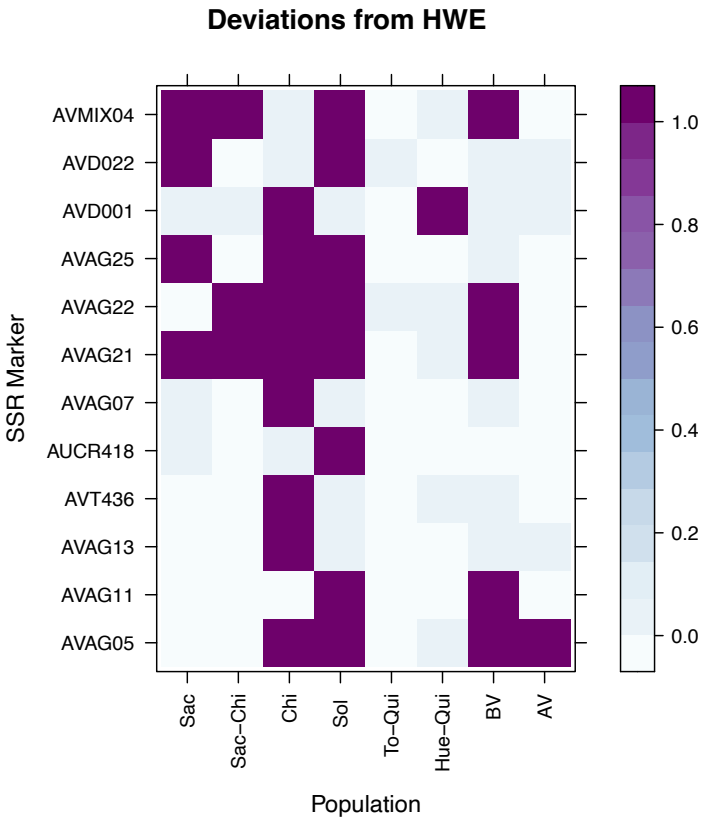


Figure A1. Deviations from HWE across populations and 12 SSR markers.

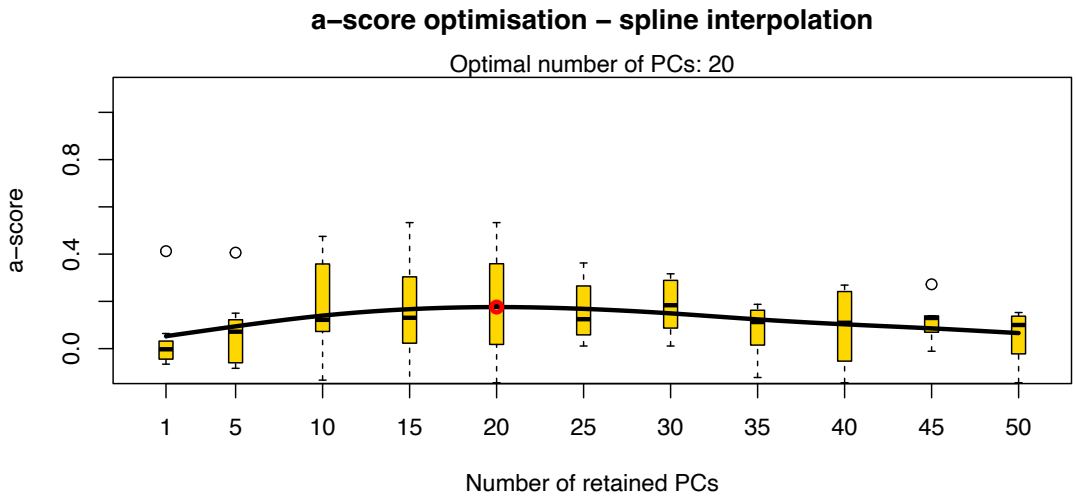


Figure A2. Principal components that should be kept in a discriminant analysis of principal components (DAPC) analysis based on 72 AFLP loci.

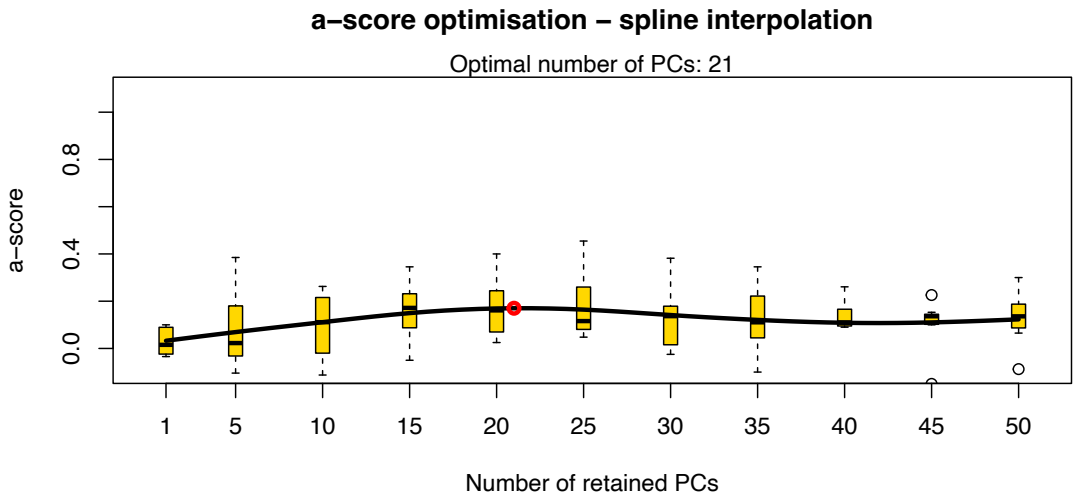


Figure A3. Principal components that should be kept in a discriminant analysis of principal components (DAPC) analysis based on 12 SSR loci.

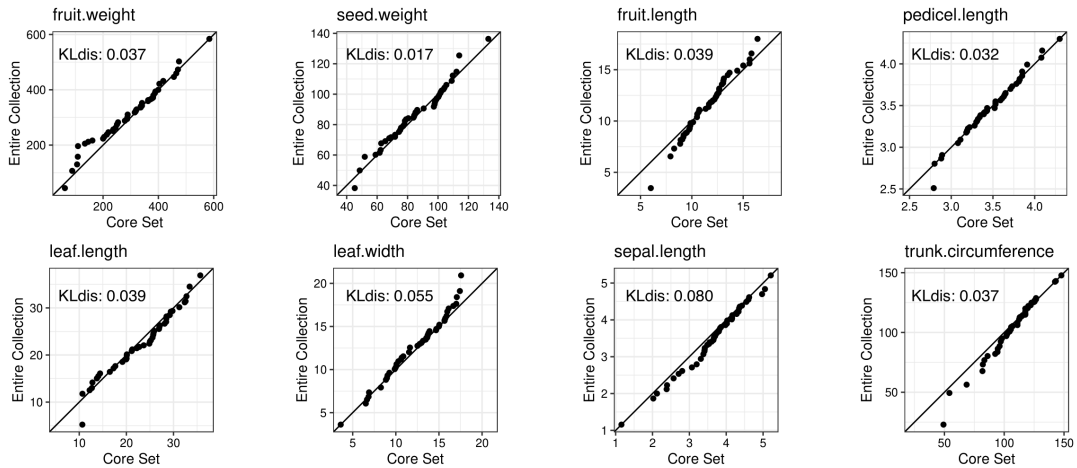


Figure A4. Quantile-Quantile (QQ) plots and Kullback-Leibler distance for the entire collection and core set for quantitative traits

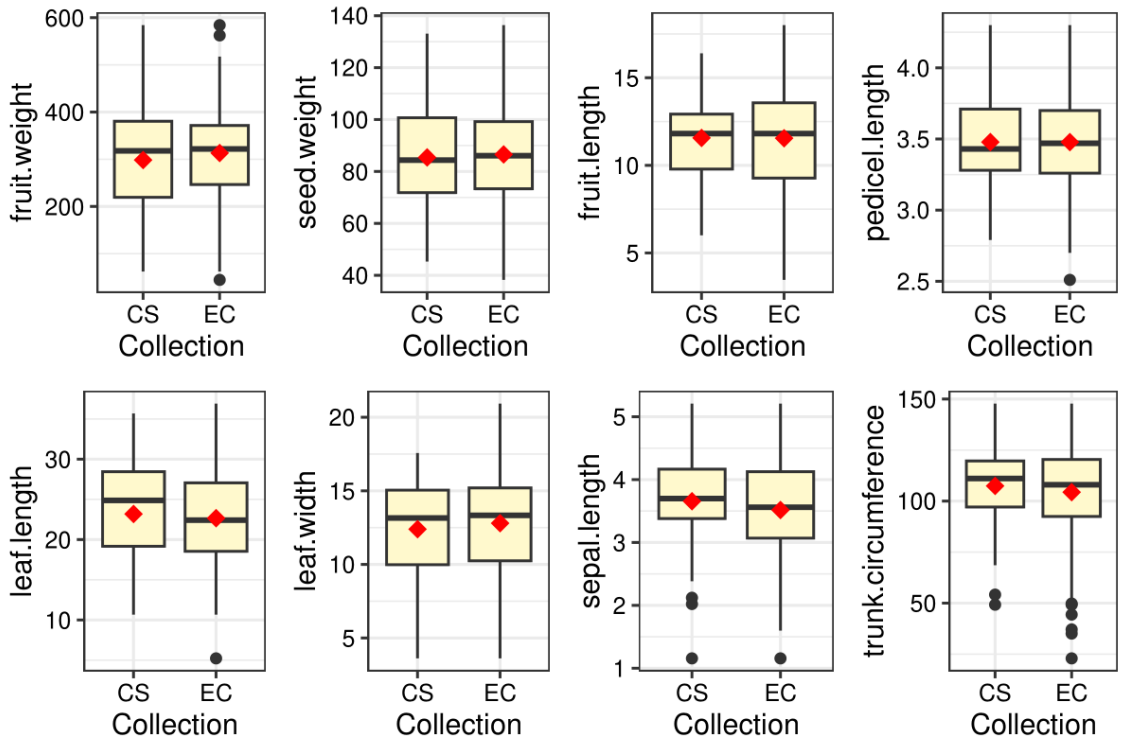


Figure A5. Boxplots showing the distribution of 8 quantitative traits in the entire germplasm (EC) and avocado core collection (CS).

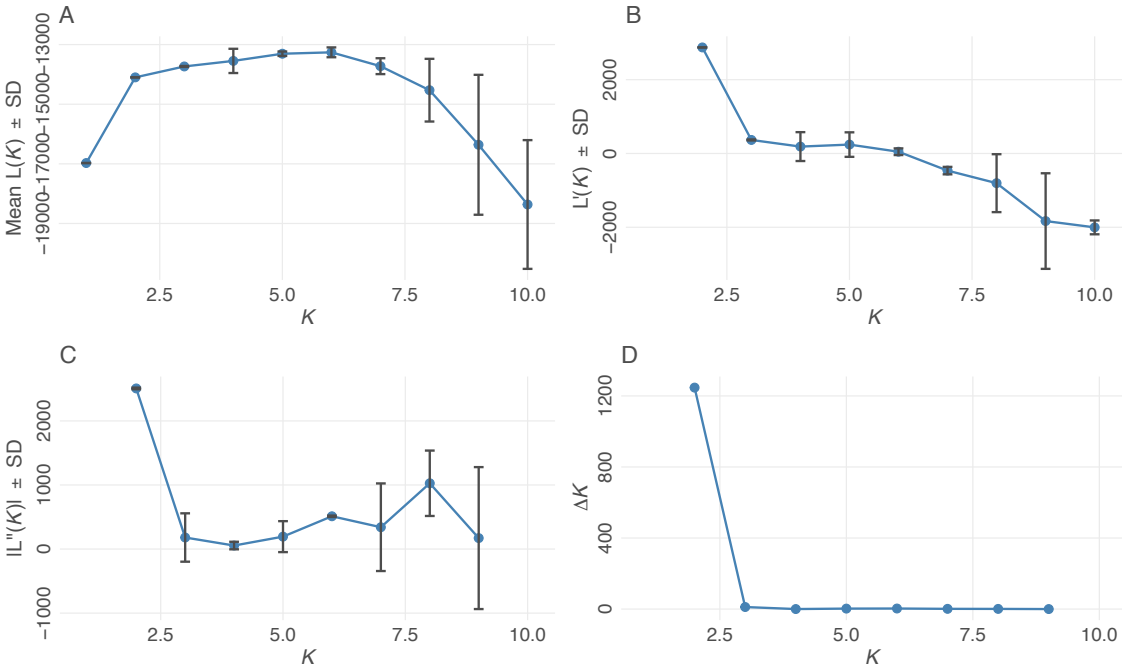


Figure A6. The choice of the most likely number of clusters (K) inferred from the clustering based on the STRUCTURE model: the values of lnP(X|K) values (A, B, C) for each of the ten independent runs for each K and ΔK values (D) for each K based on the second order rate of change of the likelihood function concerning K.

10.1 Curriculum vitae

Name: José Alejandro Ruiz Chután

Address: 15th street B 16-61 01006, Guatemala, Guatemala

Phone number: +502 55801474

Email: ruiz_chutan@ftz.czu.cz; josealendro.ruiz@icloud.com

EDUCATION

PhD. Czech University of Life Sciences, Prague (CZU), Faculty of Tropical AgriSciences (FTA)
Study programme: Tropical Agroecology and Bioresource Management
Thesis title: Assessment of morphological and genetic diversity of avocado (*Persea americana* Mill.) in Guatemala
2017 –2024

MSc. CZU, FTA
Study programme: Tropical Crop Management and Ecology
Thesis title: Assessment of genetic diversity in sorghum bicolor using RAPD markers
2014 -2016

BSc. Universidad de San Carlos de Guatemala (USAC), Facultad de Agronomía (FAUSAC)
Study programme: Ingeniería agrícola en sistemas de producción agrícola
Thesis title: Evaluation of the real-time polymerase chain reaction, Q – PCR for the detection of *Ralstonia solanacearum*, sequevar 1.
2007 –2013

WORK EXPERIENCE, INTERNSHIPS

University of the Isthmus, Guatemala: 2024 - present
Position: Professor

Tropical Agricultural Research and Higher Education Center 2023 - present
Position: Scientific consultant

United Nations Development Programme: 2020 – present
Position: Scientific researcher

Consejo Superior Universitario Centroamericano: 2021- 2023
Position: Scientific researcher

Universidad Mariano Gálvez de Guatemala: 2021 - present
Position: Professor and scientific researcher

Universidad Rafael Landívar: 2019 - 2021

Position: Professor

Universidad de San Carlos de Guatemala: 2017 - present

Position: Professor and scientific researcher

Universidad Federal do Rio Grande do Norte, Brasil

United Nations University

Biotechnology Programme for Latin America and the Caribbean: 2019

Conservation Genetics programme

Position: student's internship

RESEARCH PROJECTS

Title: **Integral project to optimize production, sustainability and strengthening of cocoa associations in Alta Verapaz, Guatemala**

Donor: Inter-American Institute for Cooperation on Agriculture, IICA. Implementation period: 2024. Budget: 150, 000 USD.

Principal investigator: José Alejandro Ruiz Chután, FTZ

Responsibilities: project management, reporting, sampling, molecular genetics lab work, field schools, data analysis, manuscript preparation.

Title: **Genetic and haplotypic diversity of *Hemileia vastatrix* in Chimaltenango, Escuintla and Sacatepéquez: basis for the use of coffee genotypes, phase II**

Donor: Direción General de Investigación (Digi), USAC. Implementation period: 2023. Budget: 49, 000 USD. Project ID: 4.8.63.4.51

Principal investigator: Amílcar Sánchez, FAUSAC

Responsibilities: sampling, molecular genetics and microbiology lab work, data analysis, manuscript preparation

Title: **Genetic diversity of *Dalbergia stevensonii* in the Franja Transversal del Norte and Petén: basis for its conservation and improvement**

Donor: Digi, USAC. Implementation period: 2022. Budget: 45,000 USD. Project ID: 4.8.63.0.29

Principal investigator: José Alejandro Ruiz Chután, CZU, FAUSAC

Responsibilities: project management, reporting, sampling, molecular genetics and tissue culture lab work, data analysis, manuscript preparation

Title: **Establishment of the scientific viability of the commercial use of fir seed (*Abies guatemalensis* Rehder) to contribute to the survival, rescue or safeguarding of the species**

Donor: United Nations Development Programme (UNDP); Global Environment Fund (GEF).

Implementation period: 2020 – 2024. Budget: 167,000 USD. Project ID: 2020-000024

Principal investigator: José Alejandro Ruiz Chután, CZU, FAUSAC

Responsibilities: project management, events organization, reporting, sampling, molecular genetics and tissue culture lab work, data analysis, manuscript preparation

Title: **Molecular characterization of tree species of the *Dalbergia* genus in Guatemala and elucidation of taxonomic aspects related to the genus**

Donor: Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Implementation period: 2021 – 2022. Budget: 50,000 USD. Project ID: (CTSP) S-566

Principal investigator: José Alejandro Ruiz Chután, CZU, FAUSAC

Responsibilities: project management, events organization, reporting, sampling, molecular genetics, data analysis, manuscript preparation

Title: **Genetic and haplotypic diversity of *Hemileia vastatrix* in Santa Rosa, Jutiapa and Jalapa: basis for the use of coffee genotypes.**

Donor: Digi, USAC. Implementation period: 2022. Budget: 50,000 USD. Project ID: 4.8.63.0.30

Principal investigator: Amílcar Sánchez, FAUSAC

Responsibilities: Responsibilities: sampling, molecular genetics and microbiology lab work, data analysis, manuscript preparation

Title: **Characterization and genetic diversity of allspice (*Pimenta dioica*, Myrtaceae) as a basis for management and exploitation plans**

Donor: Digi, USAC. Implementation period: 2021. Budget: 50,000 USD. Project ID: B4CU2021

Principal investigator: Amílcar Sánchez, FAUSAC

Responsibilities: sampling, molecular genetics, data analysis, manuscript preparation

Title: **Resistance of commercial and criollo cacao genotypes to *Moniliophothora roreri* and sensitivity to fungicides**

Donor: Digi, USAC. Implementation period: 2020. Budget: 58,000 USD. ID project: 4.8.63.4.45.

Principal investigator: Luis Montes, FAUSAC

Responsibilities: Responsibilities: sampling, field experiment, molecular genetics and microbiology lab work, data analysis, manuscript preparation

Title: **Assessment of genetic diversity of native Guatemalan cedar and establishment of an in vitro core collection**

Donor: Digi, USAC. Implementation period: 2020. Budget: 55,000 USD. ID project: B21-2020

Principal investigator: Juan Herrera, FAUSAC

Responsibilities: sampling, molecular genetics and tissue culture lab work, data analysis, manuscript preparation

Title: **Population study of *Rhizoctonia solani* and biological control strategy in potato-producing areas of the western highlands of Guatemala**

Donor: Digi, USAC. Implementation period: 2020. Budget: 58,000 USD. Project ID: B18-2020

Principal investigator: Juan Herrera, FAUSAC

Responsibilities: sampling, field experiment, molecular genetics and microbiology lab work, data analysis, manuscript preparation

Title: **Evaluation of the phosphorus solubilization efficiency of seven species of microorganisms from Andisol soils of Guatemala**

Donor: Digi, USAC. Implementation period: 2019. Budget: 65,000 USD. Project ID: 4.8.63.4.41

Principal investigator: Amilcar Sánchez, FAUSAC

Responsibilities: sampling, field experiment, molecular genetics and microbiology lab work, data analysis, manuscript preparation

Title: Assessment of five species of microorganisms, as a biological control strategy, on *Globodera rostochiensis* and *G. pallida* cysts.

Donor: Digi, USAC. Implementation period: 2019. Budget: 53,000 USD. Project ID: 4.8.63.4.48
Principal investigator: Amilcar Sánchez, FAUSAC
Responsibilities: sampling, field experiment, molecular genetics and microbiology lab work, data analysis, manuscript preparation

Title: Assessment of genetic diversity of the Guatemalan avocado (*Persea americana* Mill) using morphological characterization and molecular markers.

Donor: Digi, USAC. Implementation period: 2018. Budget: 60,000 USD. Project ID: 4.8.63.4.41
Principal investigator: Amilcar Sánchez, FAUSAC
Responsibilities: sampling, field experiment, molecular genetics lab work, data analysis, manuscript preparation

Title: Tomato breeding for resistance to late blight caused by *Phytophthora infestans* genotypes collected in Guatemala

Donor: Digi, USAC. Implementation period: 2017. Budget: 60,000 USD. Project ID: 4.8.63.4.09
Principal investigator: Amilcar Sánchez, FAUSAC
Responsibilities: Responsibilities: sampling, field experiment, molecular genetics and microbiology lab work, data analysis, manuscript preparation

SCIENTIFIC PUBLICATIONS

Degu HD, **Ruiz-Chután JA**, Kalousová M. 2024. Genetic diversity and population structure analysis of avocados (*Persea americana* Mill) from southern Ethiopia and Guatemala using polymorphic SSR markers. Genetic Resources and Crop Evolution. https:// 10.1007/s10722-023-01831-1

Ruiz-Chután JA, Kalousová M, Maňourová A, Degu HD, Berdúo-Sandoval JE, Villanueva-González C E, Lojka B. 2023. Core collection formation in Guatemalan wild avocado germplasm with phenotypic and SSR data. Agronomy **13**:2385. https://doi.org/10.3390/agronomy13092385

Maňourová A, Zbyněk P, **Ruiz-Chután JA**, Sillam-Dussès D, Tsafack S, Tchoudjeu Z, Potgieter L, Lojka B. 2023. Identification of plus trees for domestication: phenotypical description of Garcinia kola populations in Cameroon. Genetic Resources and Crop Evolution. http://dx.doi.org/10.1007/s10722-023-01750-1

Villanueva-González CE., **Ruiz-Chután JA**, Kalousova M, Moya Fernandez, RW, Villanueva C, Lojka B. 2023. Botanical diversity, structure and composition in cocoa agroforest systems in Alta Verapaz, Guatemala. Scientia Agropecuaria **14** 223-234. https://doi.org/10.17268/sci.agropecu.2023.020

Ruiz-Chután JA, Berdúo-Sandoval JE, Maňourová A, Kalousová M, Villanueva-González CE, Fernández E, Žiarovská J, Sánchez-Pérez A, Lojka B. 2023. Variability analysis of wild Guatemalan avocado

germplasm based on agro-morphological traits. Tropical and Subtropical Agroecosystems **26**. https://doi.org/10.56369/tsaes.4663

Maňourová, A., Chinheya, P., Kalousová, M., **Ruiz-Chután, J.A.,** Okafor, U., Tchoundjeu, Z., Tsobeng, A., Van Damme, P., & Lojka, B. (2023). Domestication potential of Garcinia kola Heckel (Clusiaceae): searching for diversity in South Cameroon. *Plants*,*12*, 742. https://doi.org/10.3390/plants12040742

Ruiz-Chutan, J. A., Berduo-Sandoval, J. E.., Kalousova, M., Fernandez, E., Ziarovska, J., Sanchez-Perez, A., & Lojka, B. (2022). SSRs markers reveal high genetic diversity and limited differentiation among populations of native Guatemalan avocado. *Journal of Microbiology, Biotechnology and Food Sciences*, *12*(2), e6134. https://doi.org/10.55251/jmbfs.6134

Ruiz-Chután, J.A., Berdúo-Sandoval, J.E., Melgar, S., Kalousová, M., Lojka, B., Fernández, E., Žiarovská, J., Sánchez-Pérez, A. (2022). Identificación molecular y diversidad genética de las especies de nematodos Globodera rostochiensis y G. pallida en regiones productoras de papa de Guatemala. *Ciencia, Tecnología y Salud, 9*(2), 166–181. https://doi.org/10.36829/63CTS.v9i2.930

Ruiz-Chután, J. A., Berdúo-Sandoval, J. E., Sacbajá, A., Kalousová, M., Lojka, B., Fernandez, E., Žiarovská, J., & Sanchez-Perez, A. (2021). Eficiencia de solubilización de fósforo de aislados nativos guatemaltecos de Pseudomonas fluorescens. *Ciencia, Tecnología Y Salud*, *8*(2), 184–201. https://doi.org/10.36829/63CTS.v8i2.923

Ruiz-Chutan, J. A., Berdúo-Sandoval, J. E., Kalousova, M., Lojka, B., Fernández, E., Žiarovská, J., & Sánchez-Pérez, A. (2020). Diversidad genética de materiales nativos de aguacate guatemalteco a través del marcador molecular AFLP. *Ciencia, Tecnología Y Salud*, *7*(2), 155–169. https://doi.org/10.36829/63CTS.v7i2.746

Berdúo-Sandoval, J.E., **Ruiz-Chután, J.A.,** Méndez, L., Mejía, L., Maxwell, D., Sánchez-Pérez, A. (2020). Identificación de los agentes causales asociados a la enfermedad punta morada en los cultivos de papa y tomate en Guatemala. *Ciencia, Tecnología y Salud, 7(2),* 205 - 217. doi: 10.36829/63CTS.v7i2.794

Ruiz-Chután, J.A., Salava, J., Janovská, D., Žiarovská, J., Kalousová., M, Fernández, E. (2019). Assessment of genetic diversity in *Sorghum bicolor* using RAPD markers. *Genetika, 51* (3), *789-803.* doi: 10.2298/GENSR1903789R

Berdúo-Sandoval, J.E., **Ruiz-Chután, J.A.,** Sánchez-Pérez, A. (2019). Evaluación de la resistencia de genotipos de tomate frente a aislados de *Phytophthora infestans* provenientes de Guatemala. *Ciencia, Tecnología y Salud, 6* (1), 36-47. doi: 10.36829/63CTS.v6i1.%25

Ruiz-Chután, J.A., Berdúo-Sandoval, J.E., Sánchez-Pérez, A. (2018). Diversidad genética de aislados de *Phytophthora infestans* colectados en zonas productoras de papa y tomate de Guatemala. *Ciencia, Tecnología y Salud, 5(2)*, 151-161. doi: 10.36829/63CTS.v5i2.%25

Ruiz-Chután, J.A., Herrera, M., Kalousová, M., Linares, J., Lojka, B., Barrios, S., Choxom-Chamorro, P.E., Velásquez-Méndez, L.E., Sánchez-Pérez, A., Berdúo-Sandoval, J.E. New discoveries in the taxonomy of the *Dalbergia* genus in Guatemala revealed by molecular approaches: implications for management and conservation. Submitted for consideration in Tropical and Subtropical Agroecosystems (April 2023)

Ruiz-Chután, J.A., Maňourová, A., Degu H. D., Berdúo-Sandoval, J.E, Kalousová, M., Barrios S., Villanueva-González, C.E., Herrera, M., Lojka, Bohdan. Assessing the genetic diversity of endangered species using molecular markers: The case of *Dalbergia stevensonii* in Guatemala, Submitted for consideration in Bosque (May 2023)

Degu HD, Mohammed H, Adamu M, Baw A, **Ruiz-Chután JA**, Kalousová M. Genetic diversity and population structure of Okra (Abelmoschus esculentus) from Ethiopia using SSR markers. Submitted for consideration in PLOS ONE (August 2023)

Villanueva-González CE, **Ruiz-Chután JA**, Polesný Z, Kalousova M, Villanueva C, Vaides López E, Bol Kloth HE, Chavarría C, and Lojka B. Submitted for consideration in Journal of the Faculty of Agronomy at the University of Zulia (October 2023)

Ruiz-Chután, J.A., Berdúo-Sandoval, J.E., Montes, L., Kalousová, M., Lojka, B., Sánchez-Pérez, A. Genetic diversity and population structure of *Moniliophthora roreri* in cocoa producing areas in Guatemala. Submitted for consideration in Journal of Microbiology, Biotechnology and Food Sciences (April 2022)

Ruiz-Chután, J.A., Berdúo-Sandoval, J.E., Barrios, S., Kalousová, M., Lojka, B., Žiarovská, J., Fernández, E., Sánchez-Pérez, A., (2022). Genetic diversity and population structure of *Pimienta dioica* from Guatemala revealed by RAPD markers. Submitted for consideration in Journal of Microbiology, Biotechnology and Food Sciences (May 2022)

SCIENTIFIC CONFERENCES

Ruiz-Chután, J.A., Kalousová, M., Berdúo-Sandoval, J.E, Villanueva-González, C., Sánchez-Pérez, A. Lojka, B. (2023). Uncovering the genetic diversity of hemileia vastatrix in three coffee-producing areas in Guatemala and its implications for resistance of coffee varieties. Tropentag: Competing pathways for equitable food systems transformation: trade-offs and synergies.

Villanueva-González, C., **Ruiz-Chután, J.A**., Kalousová, M., Villanueva, C., A. Lojka, B. (2023). Analysis of the diversity and timber potential in cocoa agroforestry systems in alta verapaz, Guatemala. Tropentag: Competing pathways for equitable food systems transformation: trade-offs and synergies.

Ruiz-Chután, J.A., Kalousová, M., Berdúo-Sandoval, J.E., Sánchez-Pérez, A., Lojka, B. (2022). Diversity and genetic structure of natural populations of *Cedrela odorata* in Sierra del Lacandón, Guatemala. Tropentag: Can agroecological farming feed the world? Farmers' and academia's views.

Ruiz-Chután, J.A. (2021). Molecular data analysis using R. Congreso Nacional de Biotecnología, Guatemala.

Ruiz-Chután, J.A., Berdúo-Sandoval, J.E., Kalousová, M., Lojka, B., Sánchez-Pérez, A. (2021). in vitro antagonism of native Guatemalan isolates of *Trichoderma harzianum* and *T. viride* for biological control of *Rhizoctonia solani*. Tropentag: Towards shifting paradigms in agriculture for a healthy and sustainable future.

Ruiz-Chután, J.A., Berdúo-Sandoval, J.E., Montes, L., Kalousová, M., Lojka, B., Villanueva-González, C., Sánchez-Pérez, A. (2021). Genetic diversity and population structure of *Moniliophthora roreri* in cocoa producing areas in Guatemala. Tropentag: Towards shifting paradigms in agriculture for a healthy and sustainable future.

Ruiz-Chután, J.A., Berdúo-Sandoval, J.E., Sacbajá, A., Lojka, B., Kalousová, M., Žiarovská, J… Sánchez-Pérez, A. (2020). Phosphate solubilizing activity of native Guatemalan isolates of Pseudomonas fluorescens. Tropentag: Food and nutrition security and its resilience to global crises.

Villanueva-González, C., Lojka, B., Archila, C. E & **Ruiz-Chutan, J. A.** (2020) Diversity of tree and their use in cocoa agroforestry systems in Alta Verapaz, Guatemala. Tropentag: Food and nutrition security and its resilience to global crises.

Ruiz-Chután, J.A., Kalousová, M., Berdúo-Sandoval, J.E., Sánchez-Pérez, A., & Lojka, B. (2019). Genetic diversity of native Guatemalan avocado using SSR molecular markers. Tropentag: International Research on Food Security, Natural Resource Management and Rural Development.

ACADEMIC MERITS

Recognition for an outstanding career in scientific research

Universidad de San Carlos de Guatemala. 2021

European Union Latin America Academic Links Scholarship

European Union: 2014

Second place in grade point average of the promotion of bachelor's degree

Facultad de Agronomía, Universidad de San Carlos de Guatemala. 2013