Core Collection Formation in Guatemalan Wild Avocado Germplasm with Phenotypic and SSR Data

José Alejandro Ruiz-Chután [1,2,*] , Marie Kalousová [1] , Anna Maňourová [1,3] , Hewan Demissie Degu [4] , Julio Ernesto Berdúo-Sandoval [2] , Carlos Enrique Villanueva-González [1,5] and Bohdan Lojka [1,*]

1   Department of Crop Sciences and Agroforestry, Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, Kamýcká 129, 165 00 Prague, Czech Republic; kalousovam@ftz.czu.cz (M.K.); anna.manourova@slu.se (A.M.); villanueva@ftz.czu.cz (C.E.V.-G.)

2   Facultad de Agronomía, Universidad de San Carlos de Guatemala, Guatemala City 010012, Guatemala; berduoj@hotmail.com

3   Department of Plant Protection Biology, Swedish University of Agricultural Sciences, SE-230 53 Alnarp, Sweden

4   School of Plant and Horticulture Science, College of Agriculture, Hawassa University, Hawassa P.O. Box 05, Ethiopia; hewand@hu.edu.et

5   Facultad de Ciencias Ambientales y Agrícolas, Universidad Rafael Landívar, Campus San Pedro Claver S.J., San Juan Chamelco 16010, Guatemala

*   Correspondence: ruiz_chutan@ftz.czu.cz (J.A.R.-C.); lojka@ftz.czu.cz (B.L.); Tel.: +502-5580-1474 (J.A.R.-C.); +420-734-170-763 (B.L.)

Abstract: Guatemala's wild avocado germplasm holds vital genetic value, but lacking conservation strategies imperils it. Studying its diversity is pivotal for conservation and breeding. The study aimed to comprehensively assess the wild avocado germplasm in Guatemala by combining phenotypic and genotypic data and to create a core collection for conservation and future breeding programs. A total of 189 mature avocado trees were sampled across Guatemala's northern, southern, and western regions. Morphological characteristics were documented, and genetic diversity was assessed using 12 SSR loci. The investigated germplasm revealed three distinct genetic clusters, exhibiting an average gene diversity of 0.796 and a 7.74% molecular variation among them. The samples showed various morphological characteristics that indicate the presence of three avocado races in Guatemala. The weak correlation between phenotypic and genotypic distances highlighted their independence and complementary nature. The joint matrix effectively integrated and captured genotypic and phenotypic data for comprehensive genetic diversity analysis. A core collection comprising 20% of total accessions that captured maximum genetic diversity was formed. This study exposed wild Guatemalan avocados' genetic diversity, morphological traits, and conservation significance. Integrated data capture via clustering validates holistic genetic insight for conservation and breeding strategies.

Keywords: *Persea americana*; clustering analysis; allelic richness; genetic resources conservation; genetic-phenotypic joint analysis

1. Introduction

The avocado (*Persea americana* Mill.) is a prominent fruit crop in Mesoamerica, featuring three distinct horticultural races, each with its unique ecological preferences and fruit characteristics [1]. The Mexican race bears cold-tolerant, early maturing fruit with thin skin, while the Guatemalan race, originating from tropical highlands, displays slight cold-tolerance and produces thick-skinned fruit. The West Indian race, adapted to humid tropical conditions, produces thin-skinned fruits with higher sugar content and lower oil levels [1–3]. With 24 chromosomes (2n = 24) and a genome size of 907 Mbp, avocado is a highly heterozygous diploid species [4]. It exhibits cross-pollination with outcrossing rates ranging from 74% to 96% [5]. The three avocado races are cross-compatible, allowing

hybridization when grown in close proximity, with no sterility barriers among them [3,6,7]. Each race possesses distinctive genetic characteristics that differentiate them from the others [8,9]. Despite its worldwide distribution, the avocado remains a vital fruit tree in the Mesoamerican region, holding both economic and cultural importance [10]. Although avocado production is predominantly concentrated in South America and Mesoamerica, the global consumption of the fruit is increasing steadily [11].

Assessing genetic variation is fundamental for crop breeding, conserving germplasm, and understanding evolutionary forces shaping genotypic variations [12]. Such knowledge facilitates the selection of prioritized genotypes for conservation strategies and the development of new varieties with improved fruit quality and maturation precocity. By leveraging genetic diversity, researchers can optimize avocado cultivation, ensuring sustainable and profitable fruit production [13].

Avocado germplasm has been characterized using diverse methods, encompassing both morphological and genetic markers. Morphological markers have been globally utilized, for example in Tanzania [14], Colombia [15], and Mexico [16], proving their validity and usefulness. However, these markers often face limitations such as low polymorphism and heritability, and susceptibility to environmental factors [17]. On the other hand, genetic markers have proven to overcome these constraints, providing improved characterization and understanding of avocado germplasm [18]. Applied genetic markers include isozymes [19], RFLP (restriction fragment length polymorphism) [20], VNTRs (variable number tandem repeats) [21], RAPD (random amplified polymorphic DNA) [22], ISSR (inter simple sequence repeats) [23], SSR (simple sequence repeats) [6,8,24], and SNPs (single nucleotide polymoprhism) [9,25,26]. Genetic markers have significantly improved avocado germplasm characterization, aiding in conservation and breeding efforts.

The presence of wild avocado genotypes in Guatemala that do not exclusively belong to the Guatemalan race can be attributed to a combination of historical, ecological, and biological factors. The Mesoamerican topographical conditions, climatic barriers, and the large size of the avocado seed historically contributed to the limited mobility of genetic material among regions, allowing the three races to remain distinct. However, the arrival of Spanish explorers facilitated greater movement and contact among these races, leading to increased genetic intermingling [27]. Avocado trees' protogynous dichogamy flowering model, favoring cross-pollination, along with the absence of sterility barriers between races, further accelerated the mixing of genetic material [3,7]. As a result, contemporary avocado populations in various regions of the Americas demonstrate significant racial introgression [28]. This intricate interplay of historical events, ecological conditions, and reproductive traits has shaped the genetic diversity of wild avocado germplasm in Guatemala, reflecting both native heritage and the broader genetic legacy of avocado migration and cultivation.

For conservation and reference purposes, so called core collections are established, which refers to a subset of a larger germplasm collection that represents the genetic diversity present in the entire collection [29,30]. This subset is carefully selected to encompass a wide range of genetic variations while maintaining a manageable size. An optimal core collection should possess the following key attributes: representativeness, minimal redundancy, practical manageability, comprehensive data completeness, and high usability [31]. By capturing the essential genetic diversity of a species, core collections provide a valuable resource for researchers, breeders, and conservationists. Core collections facilitate efficient utilization of genetic diversity, enabling focused investigations and breeding efforts [32,33]. Core collections also help conserve unique and rare traits within a species, which is particularly crucial in the context of diminishing biodiversity [34].

Despite being an invaluable genetic resource, the wild Guatemalan avocado faces constant threats from land use changes and deforestation [35], as well as the introduction of improved varieties that are displacing wild genotypes [16]. Therefore, studying the diversity of this genetic resource is essential for guiding conservation and breeding efforts. To date, only two studies have characterized the wild germplasm, one utilizing molecular markers [36] and the other employing morphological markers [37]. This present study aims

to conduct a comprehensive assessment, combining both phenotypic and genotypic data, to gain a deeper understanding of the wild avocado germplasm in Guatemala. Additionally, the study aims to construct a core collection based on this data to conserve the germplasm and ensure its availability for future breeding programs.

2. Materials and Methods

2.1. Study Site and Sampling

In order to identify wild avocado trees, a field survey was conducted in collaboration with the staff of Rafael Landívar University Herbarium (Guatemala), local experts, and using information from the Guatemalan atlas of wild relatives of cultivated plants [38] A total of 189 distinct avocado trees were sampled and phenotyped, representing eight geographic populations across three physiographic regions: Sacatepéquez, Chimaltenango, Sololá, Totonicapán, Quiché, Huehuetenango, Alta Verapaz, and Baja Verapaz departments located in the central, western, and northern regions (Figure 1). Table S1 provides ecological characteristics of the study site.
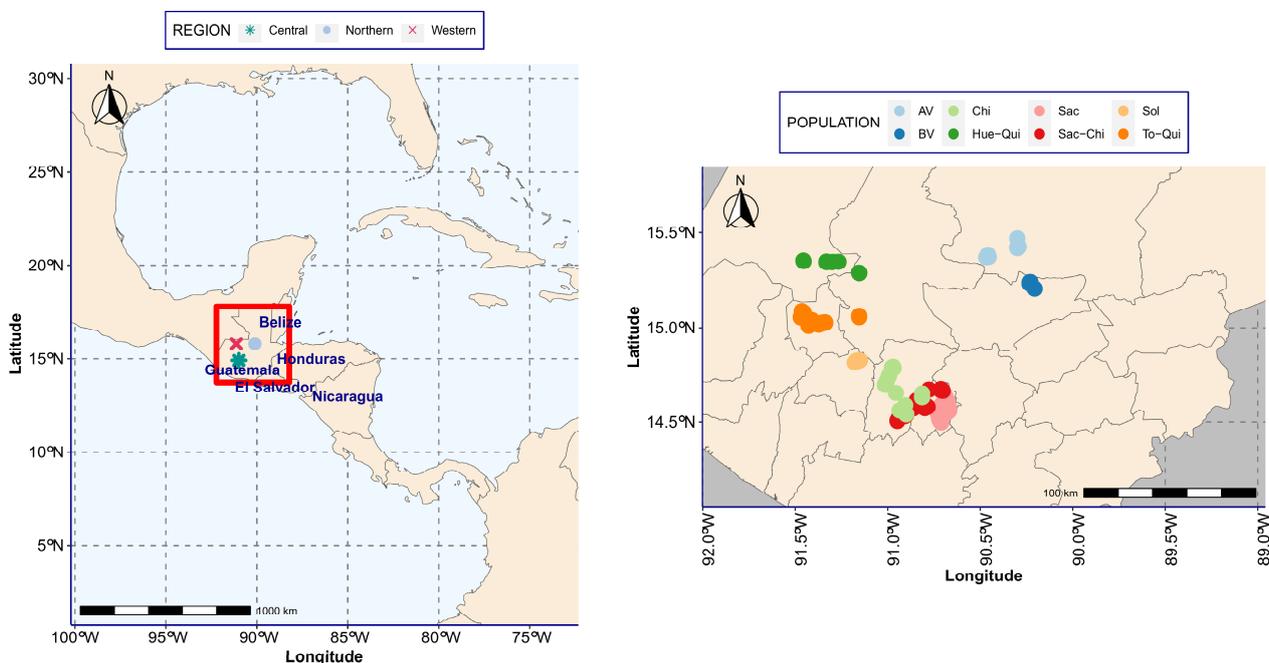
Figure 1. Map depicting the geographical locations of the sampled avocado trees in Guatemala.

The sampling strategy aimed to ensure diversity while avoiding closely related trees. The criteria for collecting 8 to 36 individuals per population were primarily based on accessibility and availability, considering factors such as the distribution of wild avocado trees, terrain, and local conditions. To determine the boundaries between populations and to minimize kinship between sampled trees, a distance of more than 30 m was maintained between the selected trees [39]. This distance was chosen as a practical guideline to reduce the likelihood of sampling closely related individuals within the same population.

During the survey, the latitude, longitude, and elevation of each tree site were recorded. For molecular analysis, three fresh leaves were collected from each individual tree. The collected leaves were carefully dried using silica gel, packed in labeled plastic bags, and transported to the Czech University of Life Sciences Prague's (CZU) Molecular Genetics Laboratory in Prague, Czech Republic.

2.2. DNA Isolation, SSR Amplification and Genotyping

The cetrimonium bromide (CTAB) technique was used to extract DNA [40]. A Nanodrop^{TM} (Thermofisher Scientific, Waltham, 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}.

We utilized twelve microsatellite primer pairs previously designed for *P. americana* [21,24] to amplify DNA samples. To enable fluorescent detection, forward primers were marked with four different colors. Three multiplex PCRs were conducted, each with specific annealing temperatures and optimized primer concentrations (Table S2). The PCR reaction mixtures were prepared in a total volume of 10 µL, consisting of 1 µL of DNA (25 ng µL^{-1}), primers at the concentrations specified in Table S2, and Multiplex PCR Plus (1 X) (QIAGEN®, Hilden, Germany). PCR amplification was performed using the Thermal Cycler T 100 (BIO-RAD, Hercules, CA, USA) with the following profile: an initial denaturation at 95 °C for 15 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at either 63.4 °C (M1), 57.6 °C (M2), or 65 °C (M3) for 1 min, extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. The PCR products were separated by electrophoresis using the Genetic Analyzer 3500 (Applied Biosystems, Waltham, MA, USA). For analysis, a mixture of 1 µL of PCR products, 0.2 µL of GeneScan-500 LIZ (Applied Biosystems, Waltham, MA, USA), and 12 µL of Hi-Di formamide (Applied Biosystems, Waltham, MA, USA) was prepared. The microsatellite alleles were scored using GeneMarker® v.2.4.0 software (Softgenetics, State College, PA, USA).

2.3. Measurement of Quantitative and Qualitative Morphological Traits

The selected avocado trees were characterized using the IPGRI field guide for avocado crops [41] which provided 21 plant descriptors to assess various parts of the tree. The methodology outlined by Juma et al. [14] was followed, evaluating each tree using a specific number of twigs, leaves, fruits, flowers, and seeds. These descriptors were chosen based on their ability to differentiate between phenotypes, exhibiting high heritability and consistent manifestation across different environments [41]. The plant descriptors were the trunk circumference and surface, leaf length, width, shape, anise smell, and color of young twig and mature leaf. The flower descriptors assessed were sepal length, petal pubescent, and pedicel shape. The fruit descriptors included the weight, length, skin surface, mature skin color, shape, and flesh texture. For the seed, the descriptors evaluated were weight, shape, and cotyledon surface. Figures S2 and S3 present the recorded descriptors and their potential variations. This standardized approach enabled a comprehensive characterization of the avocado trees, facilitating further analysis of their genetic diversity and characteristics.

The evaluation of avocado traits followed a systematic sequence as typically done in the field. Dendrometric features: Trunk circumference was measured using a tape measure, and the trunk surface texture was assessed by touch. Observations included young twig color. Fruit Traits: We conducted haptic testing to assess fruit texture and used portable semi-analytical balances to measure fruit and seed weight. Visual observations determined mature fruit skin color. We compared seed shapes to reference pictures for shape determination and examined cotyledon surface texture by tactile examination. Leaf traits: Leaf measurements (length, width, and sepal length) were taken with vernier calipers. We compared leaf and pedicel to reference pictures for shape determination, and mature leaf color was identified by observation. Additional observations: Petal pubescence and an anise odor in crushed leaves were noted.

2.4. Data Analysis

2.4.1. Population Structure Analysis

The allele dataset was utilized to investigate the genetic clusters (subpopulations) and analyze the population structure of the sampled trees through the application of discriminant analysis of principal components (DAPC) implemented in the adegenet package v.2.1.6 [42] in R v.4.2.0 [43] (The R Foundation for Statistical Computing, Vienna, Austria). The procedure involved employing the "find.clusters" function to identify the optimal number of genetic clusters (K) and subsequently selecting the best number of

genetic clusters through the Bayesian Information Criterion (BIC) using the elbow method. DAPC was utilized for characterizing the identified clusters. To determine the appropriate number of principal components and discriminant functions to retain, the "optim.a.score" function was employed. These genetic clusters, derived from the population structure analysis, were considered as populations, enabling the assessment and comparison of tree clustering in both the microsatellite and morphology-based multivariate analyses, as well as hierarchical cluster analysis.

2.4.2. Genetic Diversity

A genotype accumulation curve was created using the function "genotype_curve" in the poppr package v.2.9.4 [44] to verify the number of markers that were adequate for evaluating the genetic diversity of avocado trees. The poppr and hierfstat v.0.5 [45] packages were used to estimate the number of alleles (Na), number of private alleles (Pa), Shannon diversity index (H), Simpson's index (λ), and evenness. The allelic richness (ar), observed heterozygosity (Ho), expected heterozygosity (He), unbiased expected heterozygosity (uHe), fixation index (F_{ST}), inbreeding coefficient (F_{IS}), and Hardy–Weinberg equilibrium test were carried out in the diveRsity v.1.9.90 [46] and PopGenReport v.3.0.7 [47,48] packages. Linkage disequilibrium between loci was examined with the poppr package using 10,000 permutations. Genetic differentiation of clusters was determined through analysis of molecular variance (AMOVA) implemented in the poppr library. Covariance components were used to calculate fixation indices. A randomization test with 10,000 permutations determined significance. Population divergence was assessed by comparing pairwise population F_{ST} in hierfstat.

2.4.3. Phenotypic Variability

The statistical analysis was conducted using the compareGroups package v.4.5.1 [49] in R software. One-way analysis of variance (ANOVA) and Tukey's tests were performed to assess the significance of the cluster factor on the measured morphological values at a significance threshold of 0.05. The coefficient of variation was calculated to determine the variability across clusters for each quantitative attribute. For the qualitative morphological attributes, a cross-tabulation statistical approach was employed to examine the frequency distribution among clusters. The Pearson Chi-square (χ^2) test was used to determine the relationship between cross-tabulation variables using the "chisq.test" function in R. The Shannon diversity index was computed using EvaluateCore R package v.0.1.3 [50].

2.4.4. Joint Analysis of Phenotypic and Molecular Data

To examine the relationship among the morphological and genetic data, we employed a tanglegram analysis. This widely used 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 [51]. This allowed us to assess the correspondence and relationships between the two-clustering generated from both kinds of data. To conduct a tanglegram analysis, first genetic distances were calculated with the SSR data set. The pairwise distances were then hierarchically clustered using Ward's method with the ape package v.5.6 [52] 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 v.2.4 [53]. FAMD, which combines principles of principal component analysis (PCA) and multiple correspondence analysis (MCA), was chosen for its ability to analyze datasets containing both types of variables and balance their influence [54]. Prior to the FAMD analysis, variables were standardized to ensure equal contribution from different scales, optimizing variance explained in each dimension [55]. Hierarchical Clustering on Principal Components (HCPC), based on FAMD results, was applied to create a dendrogram using Ward's method and to identify clustering among the sampled trees. HCPC combines prin-

cipal component methods, hierarchical clustering, and partitioning clustering, including the k-means method [56].

The tanglegram was carried out using the dendextend package v.1.15.2 [57] 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 [58] and Ward's method was 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 v.1.7 [59].

2.4.5. 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 [60] using the 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 [61] 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 [62,63] 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. [31]. 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. [64].

I.   maximizing E-NE distances (CC 01)
II. maximizing A-NE distance (CC 02)
III. maximizing both E-NE and A-NE with equal weightage of 1:1 (CC 03)
IV. E-NE and A-NE with unequal weightage of 0.3:0.7 (CC 04)
V.   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 [30].

2.4.6. Evaluation of the Core Collection

A comprehensive comparison of the seven core sets was conducted, utilizing genetic distance criteria as outlined by Odong et al. [31]. 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 [65], were calculated. For qualitative traits, the coverage criteria were applied [66]. To evaluate the correlation between the trait correlation matrices of the core collection and the entire collection, the Mantel test [67] 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 [68,69] 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 [70]. Furthermore, the Wilcoxon rank test [71] 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 [72] and computed Kullback–Leibler distances [73]. The assessment of phenotypic diversity included the calculation of the Shannon–Weaver diversity index (H′) and evenness using the frequencies of qualitative traits [74]. 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.

3. Results

3.1. Genetic Characterization

#### 3.1.1. Identification of Genetic Subpopulations (Clusters) and Description of Population Structure

The genotypic resolution of the SSR markers was very sufficient as indicted by an almost complete discrimination of individuals at *n* = 4 (Figure 2).
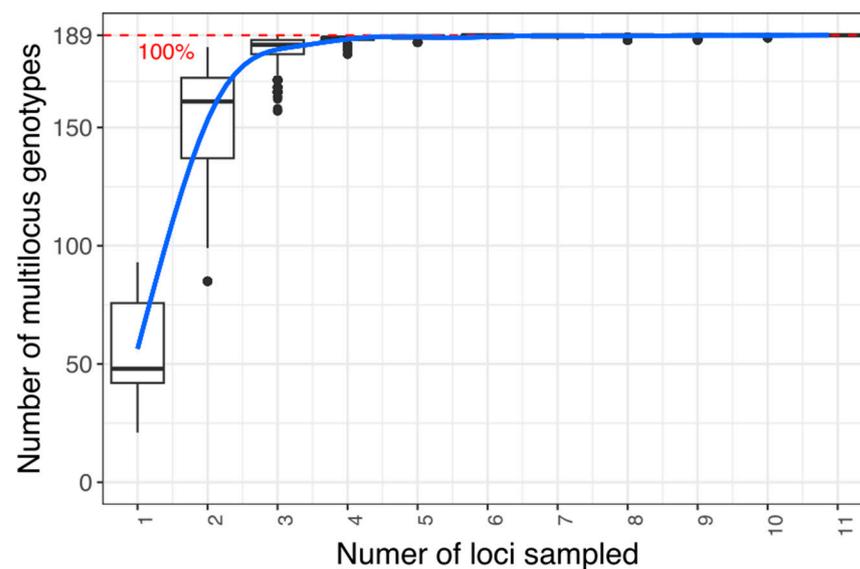
Figure 2. Genotype accumulation curve to assess avocado genotype differentiation using increasing cumulative SSR markers.

The optimal number of principal components (PCs) in the principal component analysis (PCA) step of discriminant analysis of principal components (DAPC) was determined to be 21 based on the a-score value, as shown in Figure S1. The 21 PCs of the PCA, amounting to 83.7% of the total variance, and three discriminant functions were retained. The clustering analysis using the find.cluster function on the SSR data resulted in the identification of three clusters based on the lowest BIC value, as depicted in Figure 3A. The DAPC plot in Figure 3B displayed three distinct clusters, with clusters 1 and 3 positioned to the right and cluster 2 to the left. The separation between clusters 1 and 3 was primarily driven by the second discriminant function. Among the clusters, cluster 1 had the largest number of individuals (67), followed by cluster 3 (66) and cluster 2 (56). When assigning ancestry

for K = 3, there was no clear separation of geographic populations into genetic clusters observed (Figure 3C).
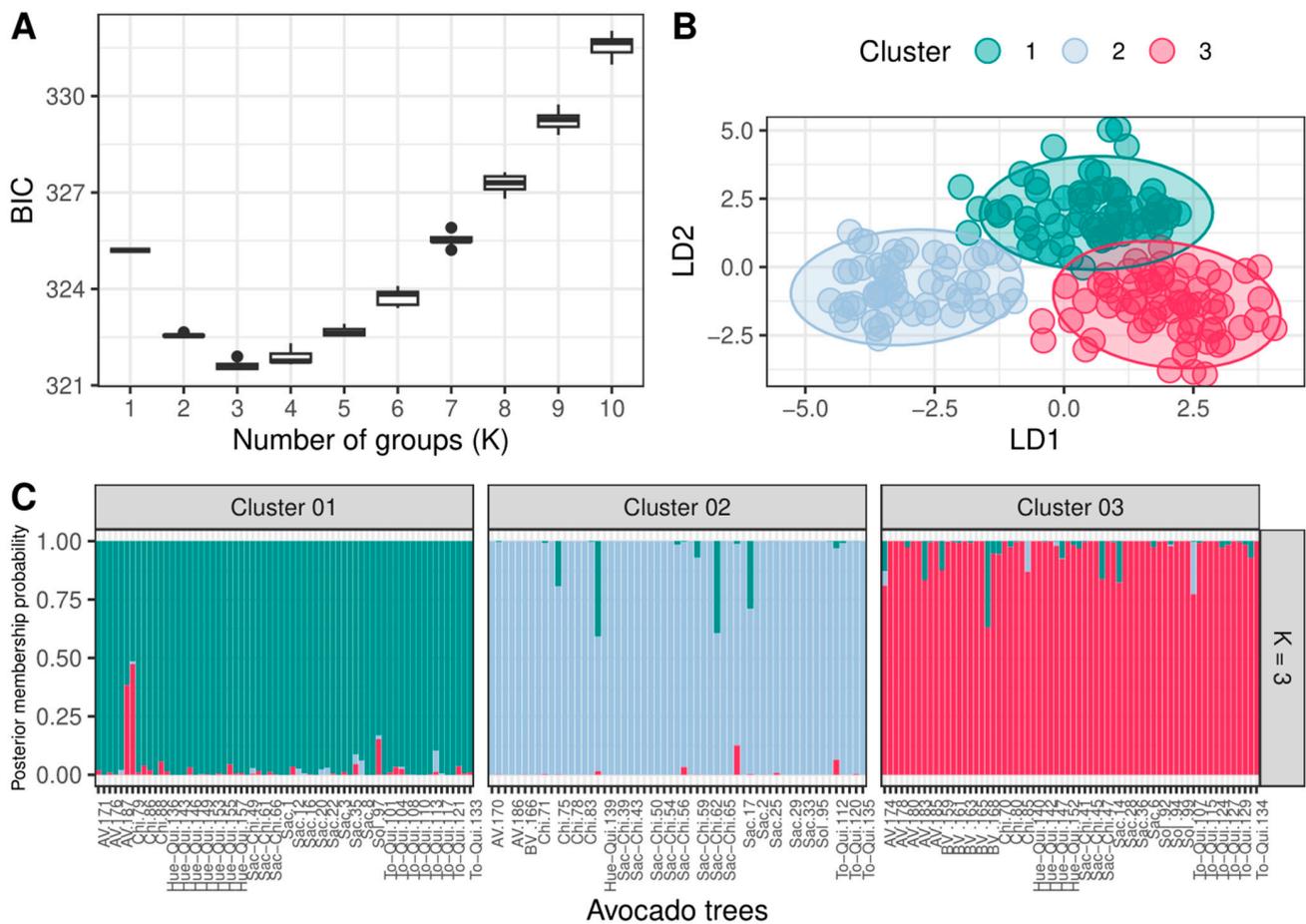
Figure 3. Discriminant analysis of principal components. In (**A**), Bayesian Information Criterion (BIC) values identify optimal clusters. (**B**) is a scatterplot showing three distinct clusters among 189 individuals. (**C**) exhibits a barplot displaying the assignment probability of each individual into one of the inferred genetic clusters (Central region: Sac, Sac-Chi, Chi; West region: Hue-Qui, To-Qui; Central region: BV, AV).

3.1.2. Genetic Diversity among Genetic Clusters

Table 1 provides detailed information of the genetic diversity parameters. The 189 avocado trees were divided into three clusters based on DAPC, with significant genetic diversity levels and deviations from Hardy–Weinberg equilibrium (HWE). Cluster 1 had 67 individuals with an average allelic richness of 13.48 alleles per locus (ar) and 2.50 private alleles (Pa), while Cluster 2 and Cluster 3 had 56 and 66 individuals, with an average allelic richness of 13.83 (ar) (1.17 Pa) and 18.83 (5.08 Pa), respectively. The Shannon diversity index (H) ranged from 4.04 to 4.19, and Simpson's index (λ) was 0.98 for all clusters, indicating high genetic diversity within the clusters. Observed heterozygosity (Ho) ranged from 0.53 to 0.59 and expected heterozygosity (He) varied from 0.77 to 0.81. The inbreeding coefficients (F_{IS}) values ranged from 0.24 to 0.35. The results demonstrate the presence of distinct genetic groups and highlight the importance of genetic variation in the studied population.

Table 1. Genetic diversity analysis of 189 avocado trees across three genetic clusters.

Group	Size	Na	ar	Pa	H	λ	Ho	He	uHe	F_{IS}	HWE
Cluster 1	67	16.25	13.49	2.50	4.19	0.98	0.59	0.81	0.81	0.28	**
Cluster 2	56	13.83	11.87	1.17	4.04	0.98	0.58	0.77	0.78	0.24	**
Cluster 3	66	18.83	15.08	5.08	4.19	0.98	0.53	0.81	0.82	0.35	**
mean	63.00	16.30	13.48	2.92	4.14	0.98	0.56	0.80	0.80	0.288	

Na: observed number of alleles per locus; ar: allelic richness; Pa: number of private alleles per locus; H: Shannon diversity index; λ: Simpson's index; Ho: observed heterozygosity; He: expected heterozygosity; uHe: unbiased expected heterozygosity; FIS: inbreeding coefficient; HWE: Hardy–Weinberg equilibrium test. ** indicates significance of *p* value at ≤0.01.

The allelic variation of the SSR loci used was wide-ranging, Na: 9 to 32, ar: 3.93 to 8.66. Genetic diversity was substantial (He: 0.60 to 0.92), inbreeding variable (F_{IS}: 0.15 to 0.51), and gene flow (Nm: 1.74 to 124.75, mean: 12.25) indicated population connectivity (Table S3). Most of the loci showed significance deviation in the Hardy–Weinberg equilibrium test. 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.

3.1.3. Analysis of Molecular Variance and Population Differentiation

The analysis of molecular variance (AMOVA) unveiled significant variation at distinct levels. Among clusters, 7.74% of the total variation (Φ_{CT} = 0.18) was observed, indicating limited differentiation among clusters. Within clusters, 26.5% of the total variation (Φ_{SC} = 0.28) was attributed to tree differences, signifying a moderate level of genetic variation. Predominantly, 66.06% of the variation (Φ_{ST} = 0.34) was detected within samples, underscoring the high genetic diversity among avocado trees (Table 2). These findings suggest that genetic variation in avocados is primarily driven by distinctions within individual trees rather than among clusters.

Table 2. Analysis of Molecular Variance (AMOVA) is conducted by grouping trees into their respective genetic clusters. Sigma signifies the variance (σ) within each cluster and the respective contribution of each source of variance to the total. Phi (Φ) serves as a metric for population differentiation. A greater Phi value is indicative of a more substantial degree of differentiation.

Variation	Sigma	%	Φ Statistics	*p*-Value
Among clusters	0.46	7.74	Φ_{CT} = 0.18	<0.01
Among samples within clusters	1.55	26.20	Φ_{SC} = 0.28	<0.01
Within samples	3.90	66.06	Φ_{ST} = 0.34	<0.01
Total	5.91	100		

3.2. Morphological Characterization

3.2.1. Quantitative Traits among Genetic Clusters

Cluster 3 showed higher fruit weight (FW: 341.61 g) than Clusters 2 and 3. Cluster 1 had a mean fruit weight of 336.11 g while Cluster 2 had 250.40 g. Cluster 3 also had the highest leaf length (LL: 37.39 cm). Traits like fruit weight and leaf length exhibited moderate variability (CV: 0.26–0.36), while others like pedicel length (PL) showed low variability (CV: 0.08) (Table 3). All traits, except leaf width and petal length, significantly differed between clusters ($p < 0.05$). This indicates unique phenotypic traits in distinct avocado genetic clusters, valuable for targeted breeding and variety selection.

Table 3. Summary of quantitative trait descriptions across genetic clusters, including mean values, standard deviations (SD), coefficients of variation (CV), and results from Tukey post-hoc test.

Trait	Cluster 1			Cluster 2			Cluster 3			Overall		
	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV	Mean	SD	CV
FW	336.11 [a]	88.56	0.26	250.40 [b]	95.44	0.38	341.61 [a]	90.54	0.27	313.14	99.40	0.32
FL	13.79 [a]	2.79	0.20	12.73 [b]	1.74	0.14	11.72 [b]	2.94	0.25	12.73	2.72	0.21
SW	92.78 [a]	18.85	0.20	88.68 [b]	15.18	0.17	89.74 [b]	15.99	0.18	90.49	16.83	0.19
LL	22.52 [b]	6.02	0.27	21.23 [b]	6.17	0.29	37.39 [a]	7.64	0.20	27.49	9.99	0.36
LW	13.06	3.53	0.27	12.38	3.41	0.28	12.89	3.77	0.29	12.80	3.57	0.28
SL	3.67 [a]	0.66	0.18	3.65 [a]	0.87	0.24	3.26 [b]	0.85	0.26	3.52	0.81	0.23
PL	3.52	0.32	0.09	3.42	0.39	0.11	3.49	0.29	0.08	3.48	0.33	0.10
TC	107.12 [a]	20.54	0.19	111.14 [a]	19.60	0.18	96.23 [b]	25.92	0.27	104.37	23.15	0.22

FW: fruit weight; FL: fruit length; SW: seed weight; LL: leaf length; LW: leaf width; SL: sepal length; PL: pedicel length; TC: trunk circumference; SD: standard deviation; CV: coefficient of variation. Different letters indicate significant differences ($p < 0.05$).

3.2.2. Qualitative Traits among Genetic Clusters

Table 4 displays diversity indices (Shannon's H and Simpson's λ) for qualitative traits in wild Guatemalan avocado germplasm, alongside chi-squared test values. Diversity indices varied notably among clusters for different traits. Trunk surface (TS) displayed similar values across clusters, with Clusters 1 and 2 having the highest diversity indices (H = 1.58 and H = 1.57, respectively) and Simpson's indices (λ = 0.66 and λ = 0.66). The color of young twigs (CYT) showed differences, with Clusters 1 and 2 having higher diversity indices (H = 2.3) compared to Cluster 3 (H = 2.21), though Cluster 3 displayed the highest Simpson's index (λ = 0.77). Leaf shape (LS) exhibited variations, with Clusters 1 and 3 having higher diversity indices (H = 2.97 and H = 2.99) compared to Cluster 2 (H = 2.86). The chi-squared test indicated significant differences in trait frequencies among clusters for specific traits like leaf anise smell (LAS), mature fruit skin color (MFSC), fruit shape (FSh), fruit texture (FT), seed shape (SS), and cotyledon surface (CS). These results align with the observed frequencies in Figure S2, indicating diverse trait profiles across genetic clusters and the entire germplasm.

Table 4. Diversity analysis of qualitative traits and chi-squared test in the wild Guatemalan avocado.

Trait	Cluster 1		Cluster 2		Cluster 3		Overall		
	λ	H	λ	H	λ	H	λ	H	χ^2
TS	0.66	1.58	0.66	1.57	0.65	1.56	0.67	1.58	3.672 [ns]
CYT	0.79	2.30	0.79	2.29	0.77	2.21	0.79	2.28	5.32 [ns]
CML	0.50	0.99	0.50	1.00	0.50	0.99	0.50	0.99	0.45 [ns]
LS	0.85	2.97	0.86	2.99	0.86	2.99	0.87	3.06	20.58 [ns]
LAS	0.47	0.96	0.49	0.98	0.39	0.83	0.48	0.97	12.91 [***]
PP	0.66	1.57	0.64	1.52	0.63	1.51	0.66	1.56	5.52 [ns]
PS	0.67	1.58	0.66	1.58	0.58	1.40	0.65	1.56	10.34 [*]
FSS	0.58	1.41	0.61	1.47	0.54	1.33	0.66	1.57	49.63 [***]
MFSC	0.73	2.33	0.86	2.80	0.85	2.79	0.83	2.71	21.50 [*]
FSh	0.72	2.44	0.88	3.11	0.88	3.09	0.87	3.07	52.68 [***]
FT	0.73	1.94	0.51	1.41	0.75	1.99	0.71	1.90	25.10 [***]
SS	0.87	2.95	0.67	2.22	0.86	2.90	0.84	2.85	36.46 [***]
CS	0.56	1.38	0.53	1.29	0.47	1.18	0.66	1.56	79.20 [***]

TS: trunk surface; CYT: color of young twig; CML: color mature leaf; LS: leaf shape; LAS: leaf anis smell; PP: petal pubescent; PS: pedicel shape; FSS: fruit skin surface; MFSC: mature color skin color; FSh: fruit shape; FT: fruit texture; SS: seed shape; CS: cotyledon surface. ns indicates not significant an *, and *** indicate $p \leq 0.05$, and 0.001, respectively.
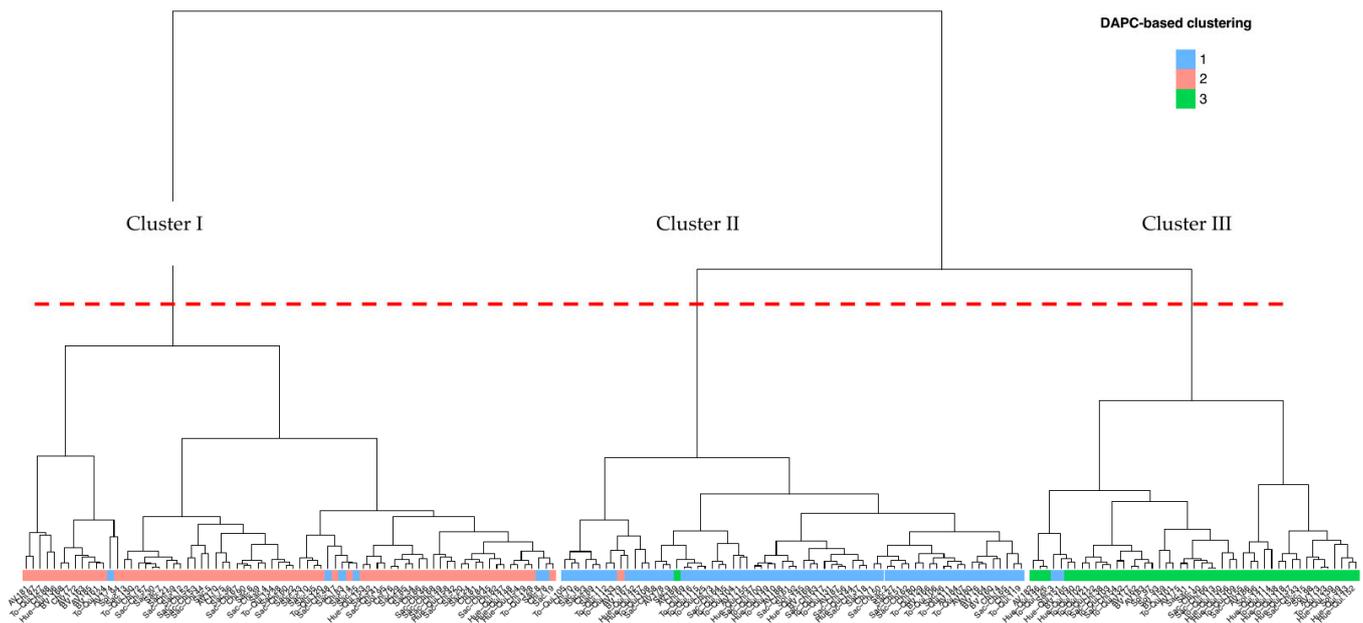
3.3. Joint Analysis of Phenotypic and Molecular Data

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 4). Using SSR markers to assess genetic diversity among wild avocado genotypes, we identified the presence of three distinct groups (Figure 4). 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. The joint matrix revealed three similarly sized clusters among the genotypes (Figure 5).
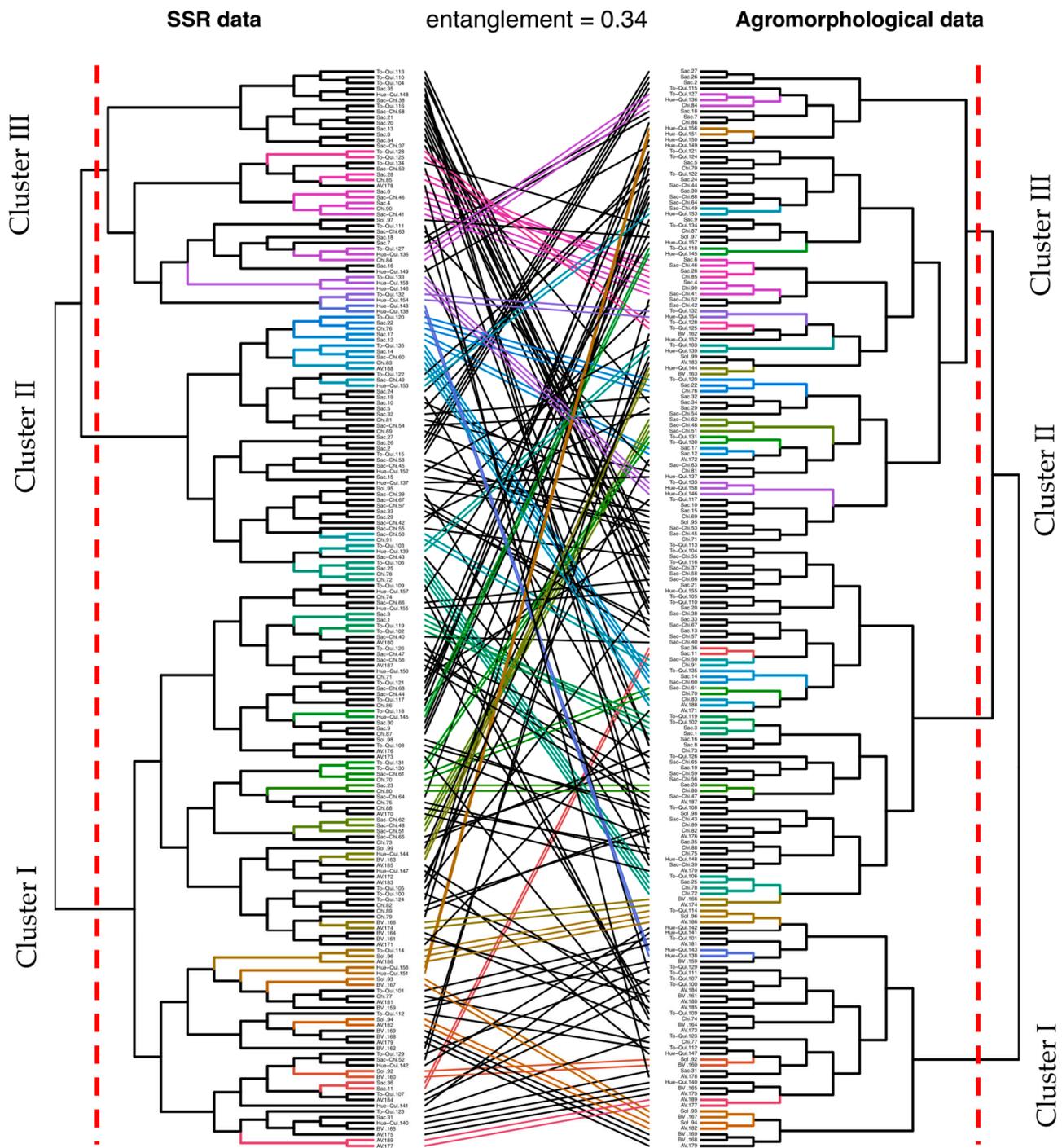
Figure 4. Tanglegram comparison of hierarchical clusters of 189 wild avocado trees based on SSR (**left**) and phenotypic (**right**) data. Colored lines connect the subtrees with identical topology in both trees.

Figure 5. Hierarchical cluster analysis showing relatedness among the 189 wild avocado genotypes based on phenotypic and molecular joint matrix.

The hierarchical dendrogram (Figure 5) and DAPC method (Figure 3) 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 4). Concurrently, the cophenetic coefficient, measuring the degree of similarity between both dendrograms, was calculated at 0.65. The discrepancy between both hierarchical clusters (Figure 4) 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 moderate 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.

3.4. Selection of a Core Collection of Avocado Genotypes Based on Phenotypic Traits and Molecular Markers

3.4.1. 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 5). 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 (95.56%), CR (92.06%), and VR (108.71%) values that exceeded the threshold CR (80%), and VR (100%), as well as a high Shannon–Weaver diversity index (H′) (Table 5), 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 5) compared to the complete germplasm sample (Ho = 0.56, Table 1). This core set includes samples from the three genetic clusters revealed by the DAPC analysis (Table S5). 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 wild Guatemalan avocado germplasm and will be used for comparative analysis with the entire germplasm collection.

Table 5. 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	0.05	0.06	0.05	0.06	0.09	0.07	0.01
E-NE	0.22	0.26	0.24	0.23	0.24	0.23	0.22
E-E	0.12	0.13	0.13	0.13	0.12	0.12	0.13
MD%	46.34	55.26	37.56	25.13	22.95	37.69	22.37
VD%	75.34	63.93	82.40	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	0.99	0.92	1.45	1.33	1.46	1.31
Mantel	0.91 **	0.87 **	0.80 **	0.80 **	0.82 **	0.90 **	0.75 **
Ho	0.54	0.51	0.58	0.54	0.58	0.52	0.55

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 $p \leq 0.01$.

3.4.2. Comparative Evaluation of the Core Collection with the Entire Wild Guatemalan Avocado Germplasm Collection

Descriptive statistics, encompassing means, ranges, coefficient of variation, interquartile range, and frequency distribution, were scrutinized for various quantitative traits within both the chosen core set and the complete collection (Table 6). Notably, CC 03 demonstrated a heightened coefficient of variation (CV) for all traits when compared to the entire germplasm, indicating a more comprehensive capture of variability within the core set. While statistical examinations, including the Newman–Keuls test and *t*-test, revealed no significant differences in means between the core set and the entire collection for all traits, it is worth noting that Levene's test did identify significant differences in sepal length (SL), although the remaining traits showed no significant disparities.

Frequency distribution plots (Figure S4) convincingly illustrated the comprehensive representation of all classes from the entire collection within the core set, affirming the capture of quantitative trait variability. Notably, the interquartile range remained largely consistent across traits, encompassing SW, FL, PL, LL, LW, and SL, except for FW and TC (Table 6). These two traits displayed symmetrical distributions of accessions between the core collection and the entire germplasm. To scrutinize the distribution patterns of the eight qualitative traits, QQ plots, and Kullback–Leibler distance calculations were meticulously conducted for both the core set and the entire collection (Figure S5). The Kullback distances (Figure S5), falling within the range of 0.04 to 0.08 for all traits, unequivocally indicated that the distribution of traits in the core collections mirrored that of the entire collection.

The results obtained from calculating the Shannon–Weaver diversity index (H′) and evenness for qualitative or categorical data in both the entire avocado germplasm and the core collection demonstrate the successful maximization of existing diversity by the extracted core sets. This is evident in the increased values of H′ for all traits, except for a minimal difference observed in FS and SS (Table 7). Notably, 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). Higher evenness values indicate a more equitable

representation of trait categories, while lower values suggest a skewed distribution with some traits being more predominant. Consequently, the increase in evenness values across all traits in the core collection implies the effective representation of the available diversity in the entire avocado germplasm.

Table 6. Comparison between the entire avocado germplasm and the core collection for various quantitative descriptors used in the formation of the core. Descriptors include range, mean, coefficient of variation, interquartile range, and frequency distribution.

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	0.60	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. $\bar{x}^a$ 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 Wilcoxon rank test. ns indicate not significant; ** indicates significant differences at 1% probability level.

Table 7. Shannon diversity index of qualitative traits in the entire germplasm and core collections of wild 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	1.16	1.1	1.1	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	2.2	2.2	0.91	0.96
LAS	0.63	0.69	0.69	0.69	0.9	1
PP	0.93	1.1	1.1	1.1	0.93	1
PS	0.99	1.08	1.1	1.1	0.9	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
FSh	2.18	2.17	2.2	2.2	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	1.1	1.1	1.1	0.99	0.99

TS: trunk surface; CYT: color of young twig; CML: color mature leaf; LS: leaf shape; LAS: leaf anis smell; PP: petal pubescent; PS: pedicel shape; FSS: fruit skin surface; MFSC: mature color skin color; FSh: fruit shape; FT: fruit 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 moderate 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 S6). 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. This conservation of trait associations and their magnitudes after sampling the core set indicates the reliability and representativeness of the selected core collection.

Principal component analysis (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 S4).

4. Discussion

4.1. Genetic Characterization

When compared to relevant studies, our analysis reveals intriguing patterns in allele diversity. For instance, Juma et al. [75] found an average of 9.4 alleles among 226 avocado trees, while Boza et al. [8] reported 9 alleles across three horticultural groups. Gross-German and Viruel [3], with 41 avocado trees, observed a lower average of 5.6 alleles. Schnell et al. [76], studying six populations with 221 samples, found a higher average of 10.3 alleles. Notably, Cañas-Gutiérrez et al. [77] reported 4.3 alleles across 90 Colombian avocado cultivars.

In terms of genetic diversity, our analysis yielded an average observed heterozygosity (Ho) of 0.56 and expected heterozygosity (He) of 0.80 across the three clusters. Comparatively, Boza et al. [8] reported Ho: 0.53 and He: 0.64 for three avocado horticultural races. Gross-German and Viruel [3], as well as Schnell et al. [76], recorded higher values of Ho: 0.66, He: 0.71 (four populations), and Ho: 0.71, He: 0.77 (six populations), respectively. Juma et al. [75] reported Ho: 0.65 and He: 0.74. Our overall average gene diversity value (0.80) emphasizes the importance of preserving genetic variability within avocado populations for conservation and breeding purposes.

The number of private alleles per locus ranged from 1.17 (cluster 2) to 5.08 (cluster 3) with mean value of 2.92 across all clusters. Private alleles could potentially be associated with the adaptation of each genetic cluster to specific environmental conditions [78,79]. Wild avocado populations in different regions may encounter varying environmental challenges, potentially resulting in the selection of distinct alleles that confer adaptive advantages within their respective habitats.

The AMOVA revealed significant genetic differentiation among the three avocado genetic clusters (F_{CT} = 0.18, p < 0.001), indicating substantial diversity and distinctiveness among the clusters. Compared to previous studies (F_{CT} = 0.02) [36], our observed population differentiation (F_{ST}) was higher when grouping was based on genetic origin, while it was lower when based on geographical origin. Gross-German and Viruel [3], as well as Boza et al. [8], reported higher overall population differentiation values (0.25 and 0.19, respectively) compared to our study. Notably, these studies employed racial origin as the basis for population grouping. In contrast, Juma et al. [6] observed a lower overall population differentiation (0.02) when considering district-based populations. This suggests that genetic origin may have a stronger impact on population differentiation in avocado groups. The results emphasize the importance of considering the grouping criteria when studying genetic diversity in avocado populations.

The remarkable genetic diversity found in wild Guatemalan avocado populations can be attributed to Guatemala's role as the species' center of origin [80–82] and domestication [2,83]. This distinction implies that the avocado first evolved and diversified in this region due to diverse ecological factors and historical processes. This extended natural evolution, combined with early human cultivation practices, fostered the accumulation and preservation of genetic diversity. Localized adaptation within distinct ecological niches led to the development of unique genetic traits, enhancing the species' resilience and adaptability [84]. Although domestication led to the propagation of selected traits in cultivated varieties [85], the wild avocado populations in Guatemala remained relatively untouched by intensive breeding. Unlike cultivated varieties, which underwent deliberate selection for specific traits through breeding practices, the wild populations in Guatemala have evolved naturally over an extended period. Consequently, wild avocado populations in Guatemala continue to harbor substantial genetic variability, offering a rich source for future breeding endeavors. This diversity is crucial for developing avocado varieties resilient to diseases, climate fluctuations, and changing agricultural practices.

4.2. Morphological Characterization

4.2.1. Quantitative Traits

The quantitative morphological traits assessed 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 [86]. This high level of variation suggests that each cluster possesses unique morphological characteristics, contributing to the overall diversity of wild avocados. These findings are consistent with previous research conducted on Mexican avocado germplasm [16,87], further validating the importance of understanding and preserving the genetic diversity of wild avocado populations. Based on Tukey's test, Clusters 1 and 3 displayed the highest average values for fruit weight and length (Table 4). These results hold significant implications for various aspects of the avocado industry. Firstly, they offer targeted breeding opportunities to enhance desirable traits, such as fruit size, aligning with market demands and consumer preferences. Secondly, these findings inform market segmentation strategies based on fruit size, aiding in tailored marketing approaches. Thirdly, they guide orchard management practices to optimize fruit yield, increasing avocado growers' profitability.

The significant variation in fruit weight (FW) among the avocado clusters can be attributed to multiple factors. Primarily, the different genetic basis of each cluster contributes to the diversity in fruit size and weight. Different genetic backgrounds may result in variations in fruit development and maturation processes [88–90]. 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. It is important to note that while these factors can lead to variation in fruit characteristics among different populations, within a specific population or region, the observed strong correlation between fruit weight and length can still serve as a valuable indicator for yield estimation and monitoring changes in avocado production [91]. As fruit weight and length are closely related within a given context, 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 fruit weight, including genetic and environmental factors, 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.

4.2.2. Qualitative Traits

The prevalence of smooth trunk surfaces in our avocado study aligns with the study area's characteristics. Most sampled trees were in medium to highlands, reflecting findings in Guatemalan and Mexican races at elevations above 1500 m above sea level, which have smoother bark. In contrast, the lowland-adapted West Indian race tends to have rougher bark [92]. Examining the aromatic aspect, the distinctive anise-scented fragrance pervading avocado groves can be ascribed to the high abundance of estragole—a unique organic compound identified exclusively in the Mexican race of avocados [93,94]. This intriguing olfactory signature sets avocados derived from Mexican race apart from their counterparts, adding to the sensory allure of these fruits.

Avocado leaf pubescence affects photosynthesis by reducing light absorption and slowing down photosynthetic activity during the growth season [95]. Additionally, it enhances water use efficiency through condensation [96], making it adaptive in drier environments and areas vulnerable to climate change.

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. [14], who explored the association between fruit shapes and avocado cultivars originating from different races. These findings indicate that the

avocado trees in Guatemala possess genetic diversity encompassing all three avocado races. The prized buttery texture of Guatemalan avocados is economically valuable. Numerous studies have established a correlation between buttery flesh and Mexican and Guatemalan avocado varieties, typically associated with moderate to high oil content [94,97,98]. Our study indicates Mexican and Guatemalan race presence, while watery flesh suggests the West Indian race. Seasonal variations reported by producers highlight the environmental impact on fruit quality, echoing Juma et al.'s findings [14]. Considering climate change effects on fruit quality is crucial, particularly in vulnerable regions like Guatemala.

Our study's diverse seed forms align with Tanzanian avocado morphological reports [14], showcasing about 17 forms. In contrast, India [99] and Colombia [15] reports noted six and three forms, respectively. Guatemala's greater genetic diversity and larger sample size likely account for the discrepancy. Popenoe [27] linked spheroid, obovate, and oblong-conic shapes to Guatemalan, West Indian, and Mexican races, respectively. These shapes in our study suggest Guatemala's wild avocado germplasm is a mix of all three races.

Fruit shape, skin color, and texture, all highlighted, aid selection for improved cultivars by farmers and breeders. Barrett et al. [100] emphasized external factors' influence on buyer attraction and impulsive purchases. Once tasted, attributes like texture and freshness impact consumer satisfaction. Visual cues shape perceptions of freshness and flavor quality during purchase, albeit sometimes misleading [100,101].

4.3. 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 [102]. A correlation coefficient exceeding 80% indicates a robust alignment between these matrices [103,104]. 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. 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. 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 noncoding regions and may not necessarily be associated with features [102,103], 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 [104]. 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 [105]. 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 [106,107]. 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.

Previous studies of other crops such as cowpea [108], yam [109], and common bean [110] 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 [111,112]. 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 [107,111–113].

4.4. Core Collection

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 [30,114,115].

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 [62,63,114].

4.4.1. Quality Assessment of 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 5). Prior studies have indicated that optimizing the mean genetic distance within a core collection is considered a favorable quality criterion, especially for core collections designed for plant breeding purposes [114,115].

Furthermore, the assessment of mean difference (MD%), variance difference (VD%), co-efficient 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%) [65,116]. Similar parameters were also employed in the evaluation of core sets in avocado [10], and other crops such as Indian mustard [117], rice [116,118], and wheat [119]. The core set's ability to accurately mirror the geographical distribution of both indigenous and exotic germplasm across the entire collection is apparent.

4.4.2. Comparative Evaluation of the Core Collections with the Whole Wild 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. While some traits showed similar interquartile ranges between the core set and the entire collection, variations were observed for others.

Additionally, we used relative frequency plots for qualitative traits and box plots (Figure S7) for quantitative traits, finding consistent patterns in both the germplasm and

the core collection. Quantile-quantile (QQ) plots and Kullback–Leibler distance analysis confirmed the core set's representation of trait distribution, with values ranging between 0.037 and 0.08. 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 for various crop species, including avocado, to examine the inter-relationships among different traits [75,90,120]. 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 S6). 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 [121,122].

The results presented here demonstrate the presence of a broad range of variability in phenotypic traits within the wild 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.

5. Conclusions

This study revealed a rich diversity in wild Guatemalan avocado germplasm, emphasizing high phenotypic and genetic variation. While phenotypic and molecular data align individually, subtle discordance surfaces in entanglement analysis. Environmental influences and non-coding markers are potential contributors to this genetic-phenotypic discordance. The joint analysis of combined data offers a holistic perspective on genetic diversity.

Amidst ongoing land use changes and logging, urgent conservation and preservation strategies are vital. We recommend establishing a core germplasm collection, exemplified by core set CC 03, balancing diversity and utility, valuable for breeding and conservation. We endorse proactive measures such as targeted habitat conservation, in situ preservation, and robust logging regulations to address shifting land use and logging. Collaboration among researchers, local communities, and policymakers is crucial.

Future research should target unrepresented sampling locations, especially in southern Guatemala's eastern and lowland regions, to enhance germplasm representation. Expanding the range of assessed morphological traits, with a specific emphasis on those pertinent to breeding programs, is crucial. A pivotal focus for future investigations lies in implementing marker-trait association analysis as an initial step in molecular-assisted selection, a promising tool for expediting avocado variety development.

Addressing these aspects will advance our understanding and facilitate the development of effective conservation and breeding strategies crucial for ensuring the continued success and sustainable utilization of wild Guatemalan avocados.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy13092385/s1, Figure S1: Principal components that should be kept in a discriminant analysis of principal components (DAPC) analysis based on 12 SSR loci; Figure S2: Barplot depicting the distribution of qualitative traits that exhibited significant differences among the three genetic clusters, as determined by a chi-squared test; Figure S3: Barplot illustrating the distribution of qualitative traits that showed no significant differences among the three genetic clusters, as determined by a chi-squared test; Figure S4: Frequency distribution plots showing the comparison of variability of quantitative traits in the entire germplasm (EC) and core collection (CS) of avocado; Figure S5: Quantile-Quantile (QQ) plots and Kullback–Leibler distance for the entire collection and core set for quantitative traits; Figure S6: Correlogram of quantitative variables; Figure S7: Boxplots showing the distribution of eight quantitative traits in the entire germplasm (EC) and avocado core collection (CS); Table S1: Ecological characteristics of the studied locations; Table S2:

Conditions for each multiplex PCR; Table S3: Characterization of 12 simple sequence repeat (SSR) loci in wild Guatemalan avocado germplasm; Table S4: Comparison of the first five principal components in entire germplasm and core collections of wild Guatemalan avocado; Table S5: Selected genotypes for avocado core collection.

References

1. Schaffer, B.; Wolstenholme, B.; Whiley, A. *The Avocado: Botany, Production and Uses*, 2nd ed.; CPI Group (UK) Ltd.: Croydon, UK, 2012.
2. Galindo-Tovar, M.E.; Ogata-Aguilar, N.; Arzate-Fernández, A. Some Aspects of Avocado (*Persea americana* Mill.) Diversity and Domestication in Mesoamerica. *Genet. Resour. Crop Evol.* **2008**, *55*, 441–450. [CrossRef]
3. Gross-German, E.; Viruel, M.A. Molecular Characterization of Avocado Germplasm with a New Set of SSR and EST-SSR Markers: Genetic Diversity, Population Structure, and Identification of Race-Specific Markers in a Group of Cultivated Genotypes. *Tree Genet. Genomes* **2013**, *9*, 539–555. [CrossRef]
4. Sánchez-González, E.I.; Gutiérrez-Díez, A.; Mayek-Pérez, N. Outcrossing Rate and Genetic Variability in Mexican Race Avocado. *J. Am. Soc. Hortic. Sci.* **2020**, *145*, 53–59. [CrossRef]
5. Borrone, J.; Olano, C.; Kuhn, D.; Brown, J.; Schnell, R.J.; Violi, H. Outcrossing in Florida Avocados as Measured Using Microsatellite Markers. *J. Am. Soc. Hortic. Sci.* **2008**, *133*, 255–261. [CrossRef]
6. Juma, I.; Geleta, M.; Nyomora, A.; Saripella, G.V.; Hovmalm, H.P.; Carlsson, A.S.; Fatih, M.; Ortiz, R. Genetic Diversity of Avocado from the Southern Highlands of Tanzania as Revealed by Microsatellite Markers. *Hereditas* **2020**, *157*, 40. [CrossRef] [PubMed]
7. Alcaraz, M.L.; Hormaza, J.I. Influence of Physical Distance between Cultivars on Yield, Outcrossing Rate and Selective Fruit Drop in Avocado (*Persea americana*, Lauraceae). *Ann. Appl. Biol.* **2011**, *158*, 354–361. [CrossRef]
8. Boza, E.J.; Tondo, C.L.; Ledesma, N.; Campbell, R.J.; Bost, J.; Schnell, R.J.; Gutiérrez, O.A. Genetic Differentiation, Races and Interracial Admixture in Avocado (*Persea americana* Mill.), and Persea Spp. Evaluated Using SSR Markers. *Genet. Resour. Crop Evol.* **2018**, *65*, 1195–1215. [CrossRef]
9. Talavera, A.; Soorni, A.; Bombarely, A.; Matas, A.J.; Hormaza, J.I. Genome-Wide SNP Discovery and Genomic Characterization in Avocado (*Persea americana* Mill.). *Sci. Rep.* **2019**, *9*, 20137. [CrossRef]
10. Guzmán, L.F.; Machida-hirano, R.; Borrayo, E.; Cortés-cruz, M.; Jarret, R.L. Genetic Structure and Selection of a Core Collection for Long Term Conservation of Avocado in Mexico. *Front. Plant Sci.* **2017**, *8*, 243. [CrossRef]
11. Yasir, M.; Das, S.; Kharya, M.D. The Phytochemical and Pharmacological Profile of *Persea americana* Mill. *Pharmacogn. Rev.* **2010**, *4*, 77–84. [CrossRef]
12. Thormann, C.E.; Ferreira, M.E.; Camargo, L.E.A.; Tivang, J.G.; Osborn, T.C. Comparison of RFLP and RAPD Markers to Estimating Genetic Relationships within and among Cruciferous Species. *Theor. Appl. Genet.* **1994**, *88*, 973–980. [CrossRef] [PubMed]
13. Ramirez-Guerrero, T.; Hernandez-Perez, M.I.; Tabares, M.S.; Marulanda-Tobon, A.; Villanueva, E.; Peña, A. Agroclimatic and Phytosanitary Events and Emerging Technologies for Their Identification in Avocado Crops: A Systematic Literature Review. *Agronomy* **2023**, *13*, 1976. [CrossRef]
14. Juma, I.; Nyomora, A.; Hovmalm, H.P.; Fatih, M.; Geleta, M.; Carlsson, A.S.; Ortiz, R.O. Characterization of Tanzanian Avocado Using Morphological Traits. *Diversity* **2020**, *12*, 64. [CrossRef]
15. López-Galé, Y.; Murcia-Riaño, N.; Romero-Barrera, Y.; Fernando Martínez, M. Morphological Characterization of Seed-Donor Creole Avocado Trees from Three Areas in Colombia. *Rev. Chapingo Ser. Hortic.* **2022**, *28*, 93–108. [CrossRef]
16. Rincón-Hernández, C.A.; Sánchez-Pérez, J.; Espinosa-García, F.J. Caracterización Química Foliar de Los Árboles de Aguacate Criollo (*Persea americana* Var. Drymifolia) En Los Bancos de Germoplasma de Michoacán, México. *Rev. Mex. Biodivers.* **2011**, *82*, 395–412.
17. Alcaraz, M.L.; Hormaza, J.I. Molecular Characterization and Genetic Diversity in an Avocado Collection of Cultivars and Local Spanish Genotypes Using SSRs. *Hereditas* **2007**, *144*, 244–253. [CrossRef]

18. Bekele, A.; Bekele, E. Overview: Morphological and Molecular Markers Role in Crop Improvement Programs. *Int. J. Curr. Res. Life Sci.* **2014**, *3*, 35–42.

19. Torres, A.M.; Bergh, B. Isozymes as Indicator of Outcrossing among "Pinkerton" Seedlings. *Calif. Avocado Soc. Yearb.* **1978**, *62*, 103–110.

20. Davis, J.; Henderson, D.; Kobayashi, M.; Clegg, M.T.; Michael, T.; Allen, P.C.K. Genealogical Relationships among Cultivated Avocado as Revealed through RFLP Analyses. *J. Hered.* **1998**, *89*, 319–323. [CrossRef]

21. Mhameed, S.; Sharon, D.; Kaufman, D.; Lahav, E.; Hillel, J.; Lavi, U. Genetic Relationships within Avocado (*Persea americana* Mill) Cultivars and between Persea Species. *Theor. Appl. Genet.* **1997**, *94*, 279–286. [CrossRef]

22. Fiedler, J.; Bufler, G.; Bangerth, F. Genetic Relationships of Avocado (*Persea americana* Mill.) Using RAPD Markers. *Euphytica* **1998**, *101*, 249–255. [CrossRef]

23. López-Guzmán, G.G.; Palomino-Hermosillo, Y.A.; Balois-Morales, R.; Bautista-Rosales, P.U.; Jiménez-Zurita, J.O. Genetic Diversity of Native Avocado in Nayarit, Mexico, Determined by ISSRs. *Cienc. Tecnol. Agropecu.* **2021**, *22*, e1686. [CrossRef]

24. Ashworth, V.E.T.M.; Kobayashi, M.C.; De La Cruz, M.; Clegg, M.T. Microsatellite Markers in Avocado (*Persea americana* Mill.): Development of Dinucleotide and Trinucleotide Markers. *Sci. Hortic.* **2004**, *101*, 255–267. [CrossRef]

25. Ge, Y.; Zhang, T.; Wu, B.; Tan, L.; Ma, F.; Zou, M.; Chen, H.; Pei, J.; Liu, Y.; Chen, Z.; et al. Genome-Wide Assessment of Avocado Germplasm Determined from Specific Length Amplified Fragment Sequencing and Transcriptomes: Population Structure, Genetic Diversity, Identification, and Application of Race-Specific Markers. *Genes* **2019**, *10*, 215. [CrossRef] [PubMed]

26. Rubinstein, M.; Eshed, R.; Rozen, A.; Zviran, T.; Kuhn, D.N.; Irihimovitch, V.; Sherman, A.; Ophir, R. Genetic Diversity of Avocado (*Persea americana* Mill.) Germplasm Using Pooled Sequencing. *BMC Genom.* **2019**, *20*, 379. [CrossRef]

27. Popenoe, W.; Zentmyer, G.A. Early History of the Avocado. *Calif. Avocado Assoc. Yearb.* **1997**, *81*, 163–171.

28. Reyes-Alemán, J.C.; Valadez-Moctezuma, E.; Simuta-Velázco, L.; Barrientos-Priego, A.; Gallegos-Vázquez, C. Distinción de Especies Del Género Persea Mediante RAPD e ISSR de ADN. *Rev. Mex. Cienc. Agrícolas* **2013**, *4*, 517–529. [CrossRef]

29. Brown, A.H.D. The Use of Plant Genetic Resources. In *The Case for Core Collections*; Brown, A.H.D., Frankel, O.H., Marshall, D.R., Williams, J.T., Eds.; Cambridge University Press: Cambridge, UK, 1989; pp. 136–156.

30. Brown, A.H.D. Core Collections: A Practical Approach to Genetic Resources Management. *Genome* **1989**, *31*, 818–824. [CrossRef]

31. Odong, T.L.; Jansen, J.; van Eeuwijk, F.A.; van Hintum, T.J.L. Quality of Core Collections for Effective Utilisation of Genetic Resources Review, Discussion and Interpretation. *Theor. Appl. Genet.* **2013**, *126*, 289–305. [CrossRef]

32. Mahmoodi, R.; Dadpour, M.R.; Hassani, D.; Zeinalabedini, M.; Vendramin, E.; Micali, S.; Nahandi, F.Z. Development of a Core Collection in Iranian Walnut (*Juglans regia* L.) Germplasm Using the Phenotypic Diversity. *Sci. Hortic.* **2019**, *249*, 439–448. [CrossRef]

33. Richards, C.M.; Volk, G.M.; Reeves, P.A.; Reilley, A.A.; Henk, A.D.; Forsline, P.L.; Aldwinckle, H.S. Selection of Stratified Core Sets Representing Wild Apple (*Malus sieversii*). *J. Am. Soc. Hortic. Sci.* **2009**, *134*, 228–235. [CrossRef]

34. Salgotra, R.K.; Chauhan, B.S. Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. *Genes* **2023**, *14*, 174. [CrossRef]

35. Bullock, E.L.; Nolte, C.; Segovia, A.R.; Woodcock, C.E. Ongoing Forest Disturbance in Guatemala's Protected Areas. *Remote Sens. Ecol. Conserv.* **2020**, *6*, 141–152. [CrossRef] [PubMed]

36. Ruiz-Chután, J.A.; Berdúo-Sandoval, J.E.; Kalousová, M.; Fernández, E.; Žiarovská, J.; Sánchéz-Pérez, A.; Lojka, B. SSRs Markers Reveal High Genetic Diversity and Limited Differentiation among Populations of Native Guatemalan Avocado. *J. Microbiol. Biotechnol. Food Sci.* **2022**, *12*, e6134. [CrossRef]

37. Ruiz-Chután, J.A.; Berdúo-Sandoval, J.E.; Maňourová, A.; Kalousová, M.; Villanueva-González, C.E.; Fernández, E.; Žiarovská, J.; Sánchez-Pérez, A.; Lojka, B. Variability Analysis of Wild Guatemalan Avocado Germplasm Based on Agro-Morphological Traits. *Trop. Subtrop. Agroecosyst.* **2023**, *26*, 52. [CrossRef]

38. Azurdia, C.; Williams, K.; Williams, D.; Van Damme, V.; Jarvis, A.; Castaño, S. Guatemalan Atlas of Crop Wild Relatives. Available online: http://www.ars.usda.gov/Services/docs.html?docid=22225 (accessed on 2 May 2023).

39. Gutierrez Caro, B. Consideraciones Para El Muestreo y Colecta de Germoplasma En La Conservación Ex Situ de Recursos Genéticos Forestales. In *Conservacion de Recursos Genéticos Forestales: Principios y Prácticas*; Gutierrez, B., Ipoinza, R., Barros, S., Eds.; Instituto Forestal: Concepción, Chile, 2015; pp. 179–196. ISBN 978 956 318 108 1.

40. Doyle, J.J.; Doyle, J.L. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.

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

42. Jombart, T. Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* **2008**, *24*, 1403–1405. [CrossRef]

43. R Core Team. *R: A Language and Environment for Statistical Computing*; GBIF: Vienna, Austria, 2022.

44. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. Poppr: An R Package for Genetic Analysis of Populations with Clonal, Partially Clonal, and/or Sexual Reproduction. *PeerJ* **2014**, *2*, e281. [CrossRef]

45. Goudet, J.; Jombart, T. Hierfstat: Estimation and Tests of Hierarchical F-Statistics. R Package Version 0.5-11. Available online: https://CRAN.R-project.org/package=hierfstat/ (accessed on 24 May 2023).

46. Keenan, K.; McGinnity, P.; Cross, T.F.; Crozier, W.W.; Prodöhl, P.A. DiveRsity: An R Package for the Estimation and Exploration of Population Genetics Parameters and Their Associated Errors. *Methods Ecol. Evol.* **2013**, *4*, 782–788. [CrossRef]

47. Adamack, A.T.; Gruber, B. PopGenReport: Simplifying Basic Population Genetic Analyses in R. *Methods Ecol. Evol.* **2014**, *5*, 384–387. [CrossRef]

48. Gruber, B.; Adamack, A.T. Landgenreport: A New r Function to Simplify Landscape Genetic Analysis Using Resistance Surface Layers. *Mol. Ecol. Resour.* **2015**, *15*, 1172–1178. [CrossRef] [PubMed]

49. Subirana, I.; Sanz, H.; Vila, J. Building Bivariate Tables: The CompareGroups Package for R. *J. Stat. Softw.* **2014**, *57*, 1–16. [CrossRef]

50. Aravind, J.; Kaur, V.; Wankhede, D.P.; Nanjundan, J. EvaluateCore: Quality Evaluation of Core Collections. R Package Version 0.1.3. Available online: https://aravind-j.github.io/EvaluateCore/ (accessed on 21 June 2023).

51. Wotzlaw, A.; Speckenmeyer, E.; Porschen, S. Generalized K-Ary Tanglegrams on Level Graphs: A Satisfiability-Based Approach and Its Evaluation. *Discret. Appl. Math.* **2012**, *160*, 2349–2363. [CrossRef]

52. Paradis, E.; Schliep, K. Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R. *Bioinformatics* **2019**, *35*, 526–528. [CrossRef]

53. Lê, S.; Josse, J.; Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25*. [CrossRef]

54. Pages, J. Analyse Factorielle de Données Mixtes. *Rev. Stat. Appliquée* **2004**, *52*, 93–111.

55. Kenkel, N. On Selecting an Appropriate Multivariate Analysis. *Can. J. Plant Sci.* **2006**, *86*, 663–676. [CrossRef]

56. Husson, F.; Josse, J.; Pages, J. Principal Component Methods—Hierarchical Clustering—Partitional Clustering: Why Would We Need to Choose for Visualizing Data? *Appl. Math. Dep.* **2010**, *17*, 1–17.

57. Galili, T. Dendextend: An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering. *Bioinformatics* **2015**, *31*, 3718–3720. [CrossRef]

58. Muñoz-Pajares, A.J. SIDIER: Substitution and Indel Distances to Infer Evolutionary Relationships. *Methods Ecol. Evol.* **2013**, *4*, 1195–1200. [CrossRef]

59. Bougeard, S.; Dray, S. Supervised Multiblock Analysis in R with the Ade4 Package. *J. Stat. Softw.* **2018**, *86*, 1–17. [CrossRef]

60. Graebner, R.; Cuesta-Marcos, A. GeneticSubsetter: Identify Favorable Subsets of Germplasm Collections. R Package Version 0.8. Available online: https://CRAN.R-project.org/package=GeneticSubsetter/ (accessed on 10 June 2021).

61. Brouwer, M.; de Blok, R. CoreCollection: Creating a Core Collection. R Package Version 0.9.5. Available online: https://github.com/PBR/coreCollection/ (accessed on 13 June 2023).

62. De Beukelaer, H.; Davenport, G. Corehunter: Multi-Purpose Core Subset Selection. R Package Version 3.2.2. Available online: https://CRAN.R-project.org/package=corehunter/ (accessed on 30 June 2023).

63. De Beukelaer, H.; Davenport, G.F.; Fack, V. Core Hunter 3: Flexible Core Subset Selection. *BMC Bioinform.* **2018**, *19*, 203. [CrossRef]

64. Kaur, V.; Aravind, J.; Manju; Jacob, S.R.; Kumari, J.; Panwar, B.S.; Pal, N.; Rana, J.C.; Pandey, A.; Kumar, A. Phenotypic Characterization, Genetic Diversity Assessment in 6778 Accessions of Barley (*Hordeum vulgare* L. ssp. *Vulgare*) Germplasm Conserved in National Genebank of India and Development of a Core Set. *Front. Plant Sci.* **2022**, *13*, 771920. [CrossRef] [PubMed]

65. Hu, J.; Zhu, J.; Xu, H.M. Methods of Constructing Core Collections by Stepwise Clustering with Three Sampling Strategies Based on the Genotypic Values of Crops. *Theor. Appl. Genet.* **2000**, *101*, 264–268. [CrossRef]

66. 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. Relationship between Phenolic Compounds, Anthocyanins Content and Antioxidant Activity in Colored Barley Germplasm. *J. Agric. Food Chem.* **2007**, *55*, 4802–4809. [CrossRef]

67. Mantel, N. The Detection of Disease Clustering and Generalized Regression Approach. *Cancer Res.* **1967**, *27*, 209–220.

68. Newman, D. The Distribution of Range in Samples from a Normal Population, Expressed in Terms of an Independent Estimate of Standard Deviation. *Biometrika* **1939**, *31*, 20–30. [CrossRef]

69. Keuls, M. The Use of the "Studentized Range" in Connection with an Analysis of Variance. *Euphytica* **1952**, *1*, 112–122. [CrossRef]

70. Levene, H. Robust Tests for Equality of Variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*; Olkin, I., Ed.; Stanford Studies in Mathematics and Statistics; Stanford University Press: Redwood City, CA, USA, 1960; pp. 278–292. ISBN 9780804705967.

71. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1*, 80–83. [CrossRef]

72. Wilk, M.B.; Gnanadesikan, R. Probability Plotting Methods for the Analysis of Data. *Biometrika* **1968**, *55*, 1–17. [CrossRef]

73. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

74. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]

75. Juma, I.; Geleta, M.; Hovmalm, H.P.; Nyomora, A.; Saripella, G.V.; Carlsson, A.S.; Fatih, M.; Ortiz, R. Comparison of Morphological and Genetic Characteristics of Avocados Grown in Tanzania. *Genes* **2021**, *12*, 63. [CrossRef]

76. Schnell, R.J.; Brown, J.S.; Olano, C.T.; Power, E.J.; Krol, C.A.; Kuhn, D.N.; Motamayor, J.C. Evaluation of Avocado Germplasm Using Microsatellite Markers. *J. Am. Soc. Hortic. Sci.* **2003**, *128*, 881–889. [CrossRef]

77. Cañas-Gutiérrez, G.P.; Alcaraz, L.; Hormaza, J.I.; Arango-Isaza, R.E.; Saldamando-Benjumea, C.I. Diversity of Avocado (*Persea americana* Mill.) Cultivars from Antioquia (*Northeast colombia*) and Comparison with a Worldwide Germplasm Collection. *Turkish J. Agric. For.* **2019**, *43*, 437–449. [CrossRef]

78. Sjöstrand, A.E.; Sjödin, P.; Jakobsson, M. Private Haplotypes Can Reveal Local Adaptation. *BMC Genet.* **2014**, *15*, 61. [CrossRef]

79. Turner, T.L.; Bourne, E.C.; Von Wettberg, E.J.; Hu, T.T.; Nuzhdin, S.V. Population Resequencing Reveals Local Adaptation of *Arabidopsis lyrata* to Serpentine Soils. *Nat. Genet.* **2010**, *42*, 260–263. [CrossRef]

80. Ashworth, V.E.T.M.; Chen, H.; Clegg, M.T. Wild Crop Relatives: Genomic and Breeding Resources: Tropical and Subtropical Fruits. In *Persea*; Kole, C., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 173–189. ISBN 978-3-642-20447-0.

81. Storey, W.; Bergh, B.; Zentmyer, G. The Origin, Indigenous Range and Dissemination of the Avocado. *Calif. Avocado Soc.* **1986**, *70*, 127–133.
82. Bergh, B. The Origin, Nature, and Genetic Improvement of the Avocado. *Calif. Avocado Soc.* **1992**, *76*, 61–75.
83. Colunga, P.; Zizumboo, D. Domestication of Plants in Maya Lowlands. *Econ. Bot.* **2004**, *58*, 101–110. [CrossRef]
84. Chung, M.Y.; Merilä, J.; Li, J.; Mao, K.; López-Pujol, J.; Tsumura, Y.; Chung, M.G. Neutral and Adaptive Genetic Diversity in Plants: An Overview. *Front. Ecol. Evol.* **2023**, *11*, 1116814. [CrossRef]
85. Purugganan, M.D. Evolutionary Insights into the Nature of Plant Domestication. *Curr. Biol.* **2019**, *29*, R705–R714. [CrossRef] [PubMed]
86. Hidalgo, R. Variabilidad Genética y Caracterización de Especiesvegetales. In *Análisis Estadístico de Datos de Caracterización Morfológica de Recursos Fitogenéticos*; Franco, T.L., Hidalgo, R., Eds.; Instituto Internacional de Recursos Fitogenéticos (IPGRI): Cali, Colombia, 2003; p. 226.
87. López-Guzmán, G.; Medina-Torres, R.; Guillén-Andrade, H.; Ramírez-Guerrero, L.G.; Juárez-López, P.; Ruelas-Hernández, P. Caracterización Morfológica En Genotipos Nativos de Aguacate (*Persea americana* Mill.) de Clima Tropical En Nayarit, México. *Rev. Mex. Cienc. Agrícolas* **2015**, *6*, 2157–2163. [CrossRef]
88. Chen, H.; Ashworth, V.; Xu, S.; Clegg, M. Quantitative Genetic Analysis of Growth Rate in Avocado. *J. Am. Soc. Hortic. Sci.* **2007**, *132*, 691–696. [CrossRef]
89. Henao-Rojas, J.C.; Lopez, J.H.; Osorio, N.W.; Ramírez-Gil, J.G. Fruit Quality in Hass Avocado and Its Relationships with Different Growing Areas under Tropical Zones. *Rev. Ceres* **2019**, *66*. [CrossRef]
90. Cañas-Gutiérrez, G.P.; Sepulveda-Ortega, S.; López-Hernández, F.; Navas-Arboleda, A.A.; Cortés, A.J. Inheritance of Yield Components and Morphological Traits in Avocado Cv. Hass from "Criollo" "Elite Trees" via Half-Sib Seedling Rootstocks. *Front. Plant Sci.* **2022**, *13*, 843099. [CrossRef]
91. Mokria, M.; Gebrekirstos, A.; Said, H.; Hadgu, K.; Hagazi, N.; Dubale, W.; Bräuning, A. Fruit Weight and Yield Estimation Models for Five Avocado Cultivars in Ethiopia. *Environ. Res. Commun.* **2022**, *4*, 075013. [CrossRef]
92. Scora, R.; Bergh, B. Origin of the Taxonomic Relationships within the Genus Persea. In Proceedings of the II World Avocado Congress, Orange, CA, USA, 21–26 April 1991; pp. 505–514.
93. Pino, J.A.; Marbot, R.; Martí, M.P. Leaf Oil of *Persea americana* Mill. Var. Drymifolia Cv. Duke Grown in Cuba. *J. Essent. Oil Res.* **2006**, *18*, 440–442. [CrossRef]
94. Pereira, M.E.C.; Tieman, D.M.; Sargent, S.A.; Klee, H.J.; Huber, D.J. Volatile Profiles of Ripening West Indian and Guatemalan-West Indian Avocado Cultivars as Affected by Aqueous 1-Methylcyclopropene. *Postharvest Biol. Technol.* **2013**, *80*, 37–46. [CrossRef]
95. Ehleringer, J.; Björkman, O.; Mooney, H.A. Leaf Pubescence: Effects on Absorptance and Photosynthesis in a Desert Shrub. *Science* **1976**, *192*, 376–377. [CrossRef]
96. Konrad, W.; Burkhardt, J.; Ebner, M.; Roth-Nebelsick, A. Leaf Pubescence as a Possibility to Increase Water Use Efficiency by Promoting Condensation. *Ecohydrology* **2015**, *8*, 480–492. [CrossRef]
97. Bost, J.B.; Smith, N.J.; Crane, J. History, Distribution and Uses. In *The Avocado: Botany, Production and Uses*; Schaffer, B., Wolstenholme, B., Whiley, A.W., Eds.; CAB International: Wallingford, UK, 2013; pp. 10–30.
98. Espinosa-Alonso, L.G.; Paredes-López, O.; Valdez-Morales, M.; Oomah, B.D. Avocado Oil Characteristics of Mexican Creole Genotypes. *Eur. J. Lipid Sci. Technol.* **2017**, *119*, 1600406. [CrossRef]
99. Ranjitha, V.; Chaitanya, H.; Ravi, C.; Shivakumar, B.; Naveen, N. Morphological Characterization of Avocado (*Persea americana* Mill.) Accessions Explored from Hill Zone Taluks of Chikkamagaluru District, Karnataka State. *J. Pharmacogn. Phytochem.* **2021**, *10*, 310–318.
100. Barrett, D.M.; Beaulieu, J.C.; Shewfelt, R. Color, Flavor, Texture, and Nutritional Quality of Fresh-Cut Fruits and Vegetables: Desirable Levels, Instrumental and Sensory Measurement, and the Effects of Processing. *Crit. Rev. Food Sci. Nutr.* **2010**, *50*, 369–389. [CrossRef]
101. Shewfelt, R.L. Fruit and Vegetable Quality. In *Fruit and Vegetable Quality: An Integrated View*; Shewfelt, R.L., Bruckner, B., Eds.; CRC Press: Boca Raton, FL, USA, 2000; pp. 160–173.
102. Allendorf, F.W.; Funk, W.C.; Aitken, S.N.; Byrne, M.; Luikart, G. Phenotypic Variation in Natural Populations. In *Conservation and the Genomics of Populations*; Allendorf, F.W., Funk, W.C., Aitken, S.N., Byrne, M., Luikart, G., Antunes, A., Eds.; Oxford University Press: New York, NY, USA, 2022; ISBN 9780198856566.
103. Vieira, M.L.C.; Santini, L.; Diniz, A.L.; de Freitas Munhoz, C. Microsatellite Markers: What They Mean and Why They Are So Useful. *Genet. Mol. Biol.* **2016**, *39*, 312–328. [CrossRef] [PubMed]
104. Singh, S.P.; Nodari, R.; Gepts, P. Genetic Diversity in Cultivated Common Bean: I. Allozymes. *Crop Sci.* **1991**, *31*, 19–23. [CrossRef]
105. Alves, A.A.; Bhering, L.L.; Rosado, T.B.; Laviola, B.G.; Formighieri, E.F.; Cruz, C.D. Joint Analysis of Phenotypic and Molecular Diversity Provides New Insights on the Genetic Variability of the Brazilian Physic Nut Germplasm Bank. *Genet. Mol. Biol.* **2013**, *36*, 371–381. [CrossRef]
106. Collard, B.C.Y.; Jahufer, M.Z.Z.; Brouwer, J.B.; Pang, E.C.K. An Introduction to Markers, Quantitative Trait Loci (QTL) Mapping and Marker-Assisted Selection for Crop Improvement: The Basic Concepts. *Euphytica* **2005**, *142*, 169–196. [CrossRef]
107. Sunil, N.; Sujatha, M.; Kumar, V.; Vanaja, M.; Basha, S.D.; Varaprasad, K.S. Correlating the Phenotypic and Molecular Diversity in *Jatropha curcas* L. *Biomass Bioenergy* **2011**, *35*, 1085–1096. [CrossRef]
108. Nkhoma, N.; Shimelis, H.; Laing, M.D.; Shayanowako, A.; Mathew, I. Assessing the Genetic Diversity of Cowpea [*Vigna unguiculata* (L.) Walp.] Germplasm Collections Using Phenotypic Traits and SNP Markers. *BMC Genet.* **2020**, *21*, 110. [CrossRef]

109. Agre, P.; Asibe, F.; Darkwa, K.; Edemodu, A.; Bauchet, G.; Asiedu, R.; Adebola, P.; Asfaw, A. Phenotypic and Molecular Assessment of Genetic Structure and Diversity in a Panel of Winged Yam (*Dioscorea alata*) Clones and Cultivars. *Sci. Rep.* **2019**, *9*, 18221. [CrossRef]

110. Guidoti, D.T.; Gonela, A.; Vidigal, M.C.G.; Conrado, T.V.; Romani, I. Interrelationship between Morphological, Agronomic and Molecular Characteristics in the Analysis of Common Bean Genetic Diversity. *Acta Sci.—Agron.* **2018**, *40*, 1–9. [CrossRef]

111. Vinu, V.; Singh, N.; Vasudev, S.; Yadava, D.K.; Kumar, S.; Naresh, S.; Bhat, S.R.; Prabhu, K.V. Assessment of Genetic Diversity in *Brassica juncea* (*Brassicaceae*) Genotypes Using Phenotypic Differences and SSR Markers. *Rev. Biol. Trop.* **2013**, *61*, 1919–1934. [PubMed]

112. Sartie, A.; Asiedu, R.; Franco, J. Genetic and Phenotypic Diversity in a Germplasm Working Collection of Cultivated Tropical Yams (*Dioscorea* Spp.). *Genet. Resour. Crop Evol.* **2012**, *59*, 1753–1765. [CrossRef]

113. De Andrade, E.K.V.; de Andrade Júnior, V.C.; de Laia, M.L.; Fernandes, J.S.C.; Oliveira, A.J.M.; Azevedo, A.M. Genetic Dissimilarity among Sweet Potato Genotypes Using Morphological and Molecular Descriptors. *Acta Sci.—Agron.* **2017**, *39*, 447–455. [CrossRef]

114. Thachuk, C.; Crossa, J.; Franco, J.; Dreisigacker, S.; Warburton, M.; Davenport, G.F. Core Hunter: An Algorithm for Sampling Genetic Resources Based on Multiple Genetic Measures. *BMC Bioinform.* **2009**, *10*, 243. [CrossRef]

115. Franco, J.; Crossa, J.; Warburton, M.L.; Taba, S. Sampling Strategies for Conserving Maize Diversity When Forming Core Subsets Using Genetic Markers. *Crop Sci.* **2006**, *46*, 854–864. [CrossRef]

116. Agrama, H.A.; Yan, W.; Lee, F.; Fjellstrom, R.; Chen, M.-H.; Jia, M.; McClung, A. Genetic Assessment of a Mini-Core Subset Developed from the USDA Rice Genebank. *Crop Sci.* **2009**, *49*, 1336–1346. [CrossRef]

117. Nanjundan, J.; Aravind, J.; Radhamani, J.; Singh, K.H.; Kumar, A.; Thakur, A.K.; Singh, K.; Meena, K.N.; Tyagi, R.K.; Singh, D. Development of Indian Mustard [*Brassica juncea* (L.) Czern.] Core Collection Based on Agro-Morphological Traits. *Genet. Resour. Crop Evol.* **2022**, *69*, 145–162. [CrossRef]

118. Ndjiondjop, M.N.; Gouda, A.C.; Eizenga, G.C.; Warburton, M.L.; Kpeki, S.B.; Wambugu, P.W.; Gnikoua, K.; Tia, D.D.; Bachabi, F. Genetic Variation and Population Structure of Oryza Sativa Accessions in the AfricaRice Collection and Development of the AfricaRice *O. sativa* Core Collection. *Crop Sci.* **2023**, *63*, 724–739. [CrossRef]

119. Phogat, B.S.; Kumar, S.; Kumari, J.; Kumar, N.; Pandey, A.C.; Singh, T.P.; Kumar, S.; Tyagi, R.K.; Jacob, S.R.; Singh, A.K.; et al. Characterization of Wheat Germplasm Conserved in the Indian National Genebank and Establishment of a Composite Core Collection. *Crop Sci.* **2021**, *61*, 604–620. [CrossRef]

120. Awachare, C.; Karunakaran, G.; Madhavi, M.; Sakthivel, T. Studies on Morphological Characterization of 72 Avocado (*Persea americana* Mill.) Accessions. *Pharma Innov. J.* **2023**, *12*, 1970–1975.

121. Mahajan, R.; Bisht, I.; Dhillon, B. Establishment of a Core Collection of World Sesame (*Sesamum indicum* L.) Germplasm Accessions. *SABRAO J. Breed. Genet.* **2007**, *39*, 53–64.

122. Reddy, L.J.; Upadhyaya, H.D.; Gowda, C.L.L.; Singh, S. Development of Core Collection in Pigeonpea [*Cajanus cajan* (L.) Millspaugh] Using Geographic and Qualitative Morphological Descriptors. *Genet. Resour. Crop Evol.* **2005**, *52*, 1049–1056. [CrossRef]