International Journal of
Molecular Sciences

MDPI

Article

Genomic Diversity of Avocado in the Morogoro Region and Southern Highlands of Tanzania

Andrés J. Cortés [1,*,†], Juma M. Hussein [2] and Ibrahim Juma [3]

[1] Department of Plant Breeding, Swedish University of Agricultural Sciences, 23436 Lomma, Sweden
[2] Department of Molecular Biology and Biotechnology, University of Dar es Salaam, P.O. Box 35179 Dar es Salaam, Tanzania; jmhussein@udsm.ac.tz
[3] Department of Food Science and Technology, University of Dar es Salaam, P.O. Box 35134 Dar es Salaam, Tanzania; vuga.ibrahim@udsm.ac.tz
* Correspondence: ancortesv@unal.edu.co
† Current address: Facultad de Ciencias Agrarias—Departamento de Ciencias Forestales, Universidad Nacional de Colombia—Sede Medellín, Medellín 050034, Colombia.

Abstract

Avocado (*Persea americana* Mill.) is one of the most widely consumed fruit tree crops worldwide, with cultivation expanding rapidly beyond its Mesoamerican and northwest South America center of origin. In emerging secondary diversity centers such as East Africa, farmers have long propagated seedling naturalized populations that may hold valuable reservoirs of genetic diversity, yet these resources remain underexplored. To help fill this gap, this study developed the first genomic resources for avocados in Tanzania, where avocado has a long history of introduction and diversification dating to the first Arab incursions and Catholic missionary missions. Low-coverage whole-genome resequencing (lcWGS) data were obtained from 95 trees sampled in Tanzania across the low- to mid-altitude Morogoro region (n = 25) and the Southern Highlands—i.e., the Iringa (n = 20), Mbeya (n = 30) and Ruvuma (n = 20) regions. In order to guide racial assignation, sequences were merged with NCBI-available lcWGS data from 205 avocado trees, including 42 commercial varieties, with reported ancestry. Population stratification as inferred via maximum likelihood phylogenetic inference, genetic principal component analysis, and ADMIXTURE unsupervised clustering suggested that the sampled Tanzanian avocado trees were genetically closer to the West Indian race and more distant from the northwest South American Caribbean and Andean groups. Additionally, while the trees from the low- to mid-altitude region of Morogoro were almost exclusively West Indian type, some trees from the Southern Highlands aligned more closely with West Indian × Guatemalan and West Indian × Mexican hybrids. These trends were equally supported by a subset of 10,460 high-coverage (10×) SNP markers. Together these findings clarify the dynamics of avocado diversification in a secondary center in East Africa, spanning recent introductions from a single Mesoamerican race, adaptation to a wide range of locally geographic conditions, and farmer-driven selection matching local tribal preferences. Characterizing these locally adapted resources is key for identifying underrepresented yet promising provenances, developing resilient and sustainable horticultural production systems, and safeguarding the species' global genetic heritage.

Keywords: *Persea americana* L.; *in situ* conservation; East Africa; 'plus trees'; low-coverage whole-genome resequencing (lcWGS)

Check for updates

Academic Editors: Karel Doležal and Hee-Jae Cha

Received: 22 February 2026
Revised: 19 March 2026
Accepted: 24 March 2026
Published: 28 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Avocado (*Persea americana* Mill.) is among the most economically important and nutritionally valuable fruit tree crops worldwide [1], with global demand steadily increasing due to its high content of monounsaturated fatty acids, vitamins, and bioactive compounds [2–6]. Native to Mesoamerica and northwest South America [7,8], avocado diversity has historically been structured in three classical botanical races [7,9–11]—Mexican, Guatemalan, and West Indian—the first two adapted to high-altitude cooler environments, the later more associated with warmer lowlands [12,13]. More recently, genomic evidence has further distinguished unique, Pleistocene-dating, Caribbean and Andean linages in northwest South America [14,15], deepening our understanding of avocado's evolutionary history.

Beyond its primary center of origin, avocado cultivation has expanded extensively into secondary regions across the tropics and subtropics. In East Africa, and particularly in Tanzania, avocado has a long yet fragmented history of introduction, beginning with early trade routes associated with Arab incursions and later reinforced by Catholic missionary settlers and colonial agricultural programs [16]. Over generations, farmer-mediated seed propagation, local trading networks, and adaptation to heterogeneous micro-agroecological conditions have uniquely shaped these naturalized populations [17]. Contrasting environments such as the low- to mid-altitude region of Morogoro and the Southern Highlands—including Iringa, Mbeya and Ruvuma regions—may therefore harbor unique genetic diversity and locally adaptive alleles relevant for tolerance to abiotic stresses including drought, heat, cooler temperatures and variable rainfall regimes.

Despite avocado's growing economic importance in East Africa, genomic resources for Tanzanian germplasm are absent. Most population genomic studies have focused on materials from Mesoamerica [18–21], northwest South America [22–25], secondary centers in Europe [26,27] and South Africa [28], or commercial cultivars [29] like 'Hass' [30] and 'Gwen' [31], leaving secondary diversification in East Africa poorly characterized. Thus, the racial ancestry, degree of introgression, and internal population stratification of Tanzanian avocado populations are largely unknown. It remains unclear whether these naturalized trees derive predominantly from a Mesoamerican race, represent admixed lineages introduced through multiple routes, or constitute emerging, locally differentiated genepools shaped by environmental selection and farmer preferences.

This knowledge gap limits the conservation and utilization of Tanzanian germplasm, jeopardizing efforts to leverage local diversity as novel avocado varieties or within ongoing rootstock breeding programs. In particular, contrasting agroecological conditions between the warm, low- to mid-altitude Morogoro region, and the cooler, higher-altitude Southern Highlands offer a natural gradient to explore how introduction history, gene flow, and local adaptation have interplayed to structure genomic diversity in a secondary diversity center. Afterall, in perennial fruit trees crops such as avocado—characterized by allogamy, long generation times and incipient domestication [32–34]—*in situ* farmer-conserved seedling populations may harbor vital reservoirs of genetic novelty [35,36].

Integrating low-coverage whole-genome resequencing (lcWGS) [37], multivariate population genomics, and admixture modeling would enable the bridging of this gap by offering fine-scale reconstructions of ancestry, introgression, and divergence patterns, even in recently established secondary diversity centers. These last-generation genomic tools can further inform marker-assisted selection, genomic prediction, and pre-breeding pipelines aimed at enhancing avocado productivity, resilience and sustainability. Hence, the study goals were to (*i*) generate the first genome-wide SNP dataset for Tanzanian avocado germplasm using lcWGS, (*ii*) reconstruct the racial ancestry of trees from Morogoro and the Southern Highlands by integrating racial controls' lcWGS, and (*iii*) assess the population stratification and extent of introgression in avocados from Tanzania.

To help fill the research gap, this study investigated which classical Mesoamerican avocado races or newly defined northwest South American genetic groups are Tanzanian populations genetically closely related to, and whether the agroecologically distinct regions of low- to mid-altitude Morogoro and the Southern Highlands harbor genomically differentiated or admixed avocado lineages. We hypothesized that (*i*) Tanzanian avocado trees would cluster predominantly with the West Indian race, consistent with adaptation to tropical climates, (*ii*) trees from the Southern Highlands would exhibit greater introgression from Guatemalan and/or Mexican races due to allelic filtering for adaptation to cooler, higher-elevation environments, and (*iii*) regional-specific genomic signatures would reflect emerging differentiation within this secondary center of diversity. Revealing the genomic diversity of Tanzanian avocados would assist conservation and molecular breeding efforts targeting resilient and sustainable horticultural production under changing climatic conditions in an understudied yet promising hub in East Africa.

2. Results

With the aim of reconstructing the population stratification and racial ancestry of avocado germplasm from the low- to mid-altitude Morogoro region and the Southern Highlands of Tanzania, low-coverage whole-genome resequencing (lcWGS) was carried out on 95 avocado trees, and the sequences obtained were merged with the NCBI-available lcWGS data from 205 avocados with reported racial ancestry [14]. This full lcWGS-based SNP panel was analyzed through phylogenetic, multivariate and admixture tools. lcWGS genotyping placed the Tanzania avocados closer to the West Indian race, with introgression from the Guatemalan and Mexican races for some trees from the Southern Highlands. The ancestry, local divergence and introgression patterns of the analyzed Tanzanian germplasm highlights prospects for diversifying breeding by eventually testing novel provenances in East Africa, particularly from underrepresented racial groups.

2.1. lcWGS-Based SNP Markers

Low-coverage whole-genome resequencing (lcWGS) of 95 Tanzanian avocado trees (Table S1), spanning the low- to mid-altitude Morogoro region and the Southern Highlands (i.e., Iringa, Mbeya and Ruvuma regions), yielded an average of 30,648,133 reads and 4,627,868,040 bp per sample with a mean G:C content of 40.04%, Q20 of 97.55% and Q30 of 93.46% (Table S2). An average of 29,628,649 (96.67%) reads passed quality filters at a mean depth of 4.83× (Table S3), 29,439,508 (99.35%) of which mapped against the avocado var. Hass reference genome [30] with an average duplicate rate of 13.11% (Table S4). An average of 81.72% mapped reads per sample had a minimum depth $\geq 1\times$, 46.28% $\geq 4\times$, 5.28% $\geq 10\times$, and 1.15% $\geq 20\times$ (Table S5).

After merging this dataset and the 205 controls with reported racial ancestry, SNP calling yielded an average transition/transversion (Ts/Tv) ratio of 2.19 (Table S6). A total of 40,994,428 and 14,748,136 SNPs mapped to non-coding inter-genic and intronic regions, respectively, while 1,868,065 were exonic, 745,468 synonymous and 1,082,915 non-synonymous. Additionally, 34,027 and 4162 SNPs caused a gain or a loss of a stop codon, respectively, and 12,019 altered the splicing pattern. Finally, 2,882,670 SNPs fell in the 1 kb window upstream of the transcription start site, while 2,415,218 SNPs fell in the 1 kb window downstream of the transcription termination site.

Meanwhile, more stringent marker filtering (i.e., sequencing depth of at least 10× per sample per position, missing data per variant not exceeding 20%, and *maf* $\geq$ 5%) led to a smaller dataset of 10,460 high-coverage SNP markers, explored for corroborative phylogenetic inference and unsupervised multivariate clustering analyses, as well as a resource for downstream applications by the community.

2.2. Phylogenetic, Multivariate and Admixture Reconstructions

Phylogenetic inference grouped the majority of Tanzanian samples as a sister clade of the West Indian race (Figure 1), separated from the Mexican and Guatemalan racial controls, as well as the Andean and Caribbean groups from northwest South America.

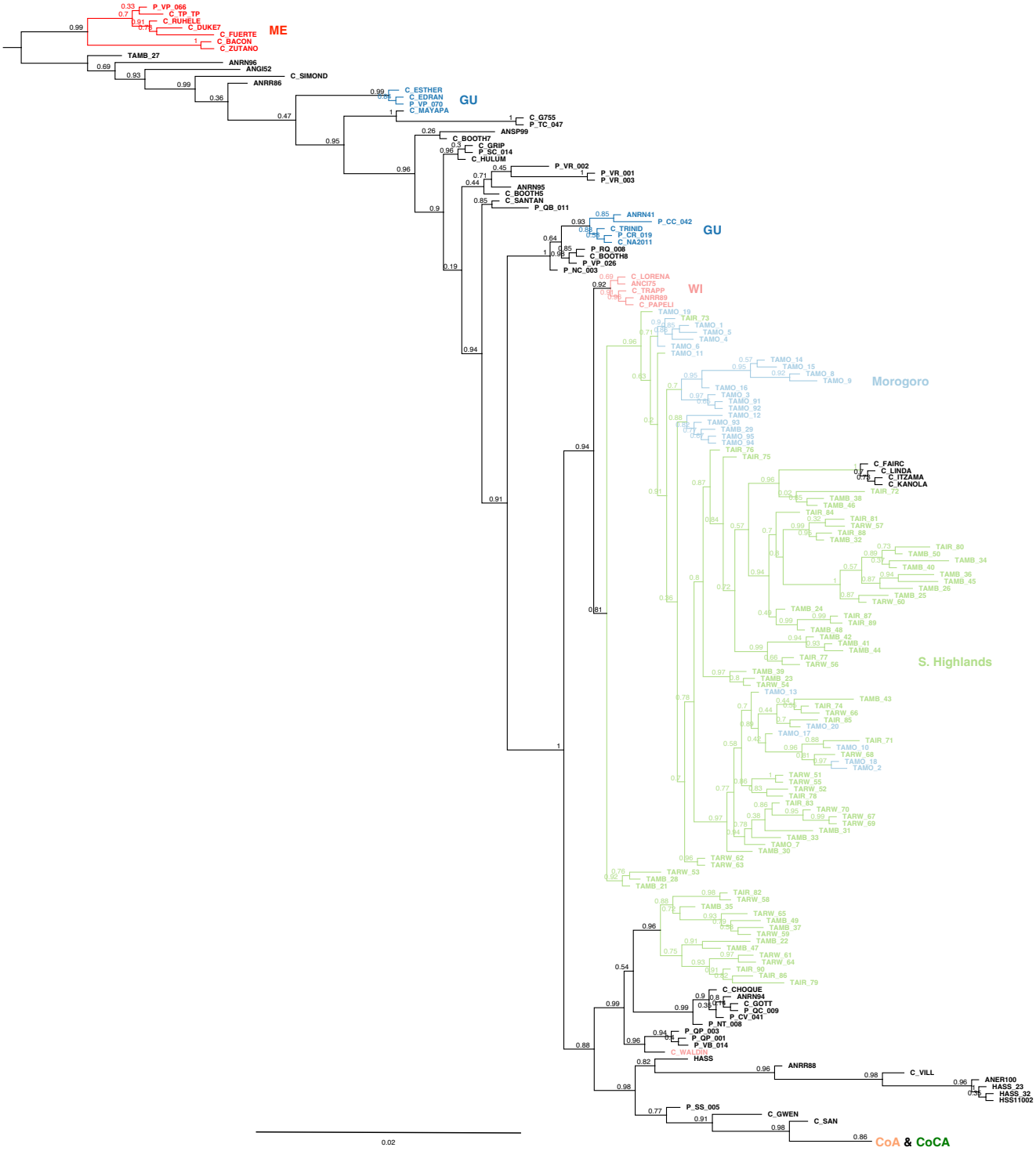


Figure 1. Rooted phylogenetic inference in 95 Tanzanian avocado trees from the low- to mid-altitude Morogoro region (light blue) and the Southern Highlands (i.e., Iringa, Mbeya and Ruvuma regions) (light green), together with 205 NCBI-available racial controls [14], all genotyped with high-coverage lcWGS-derived SNPs (full set in Figure S1). Controls colored as: Mexican (ME) race in dark red, Guatemalan (GU) race in dark blue, and West Indian (WI) in light red. The last branch contains the northwest South American Andean (CoA) and Caribbean (CoCA) clades (Figure S2).

On the other hand, principal component analysis (PCA) recovered seven groups (Figure 2, Table S7), with PC1 and PC2 accounting for 47.18% and 14.31% of variance, respectively. Data projection towards the main component positioned the Tanzanian samples from the low- to mid-altitude Morogoro region and the Southern Highlands closer to the Mesoamerican races (i.e., Mexican ME, Guatemalan GU, and West Indian WI) than to the South American groups (i.e., Caribbean CoCA, and Andean CoA). Projection to the second component confirmed that the Tanzanian trees were genetically closer to the West Indian racial controls.

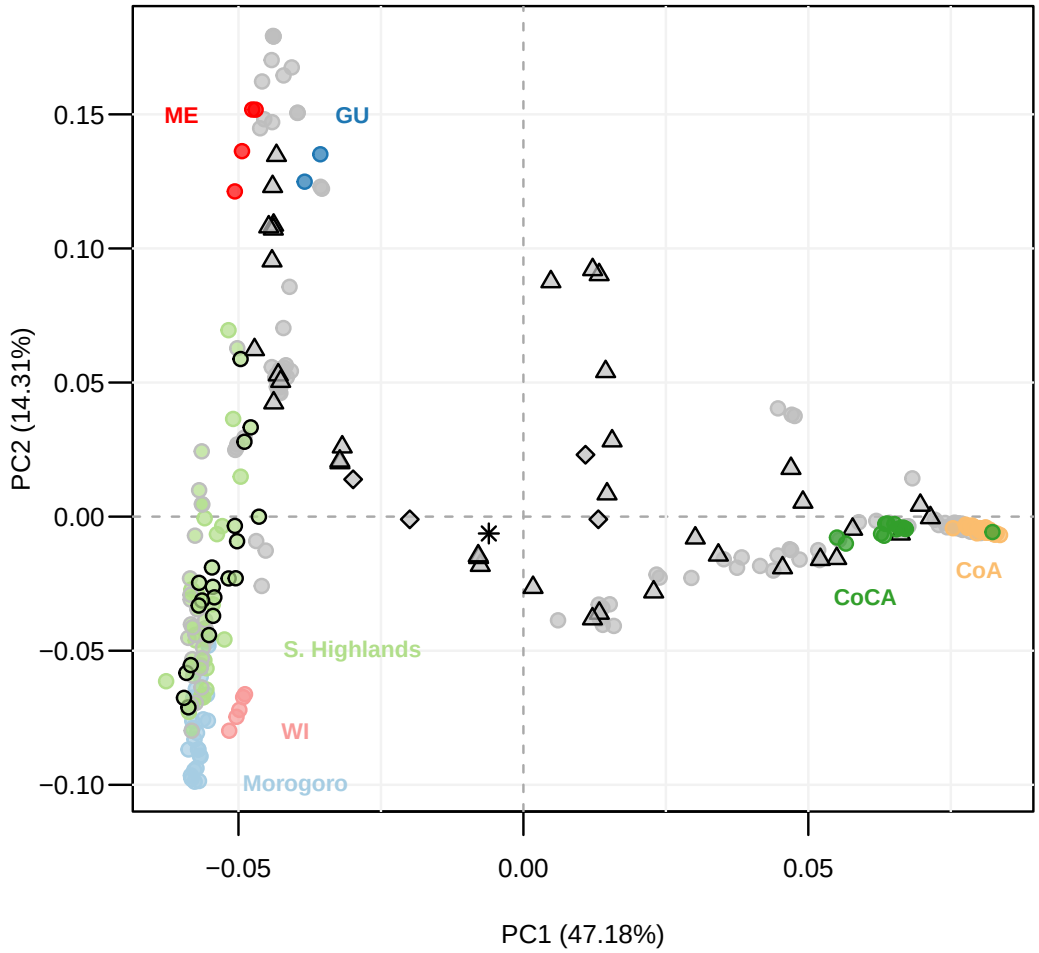


Figure 2. Principal component analysis (PCA) displaying unsupervised genetic clustering of 95 Tanzanian avocado trees from the low- to mid-altitude Morogoro region (light blue) and the Southern Highlands (light green), with 205 NCBI-available racial controls [14]—all genotyped with lcWGS-derived SNPs. Trees from the Southern Highlands in Tanzania are colored as follows: light green dots with green edge are from the Iringa region, while those with gray and black edges are from the Mbeya and Ruvuma regions, respectively. Racial controls are labeled and colored top down from left to right, as follows: dark red for the Mexican (ME) race, dark blue for the Guatemalan (GU) race, light red for the West Indian (WI) race, and dark green and light orange for the Caribbean (CoCA) and Andean (CoA) groups from northwest South America. Gray dots mark double hybrids between these last five racial control groups, while triangles, diamonds and the asterisk, respectively, mark the triple, tetra and penta hybrids. The detailed assignation of the admixed genotypes (i.e., gray dots, triangles and diamonds) to the corresponding putative parental racial groups is detailed in the next figure. Principal component (PC)-explained variance is given in the axes' labels.

Specifically, the trees from Morogoro were almost exclusively of the West Indian type, while some trees from the Southern Highlands aligned more closely with West Indian × Guatemalan and West Indian × Mexican hybrids (Figure 3). Within the Southern Highlands, trees from the regions of Iringa, Mbeya and Ruvuma were panmictic.

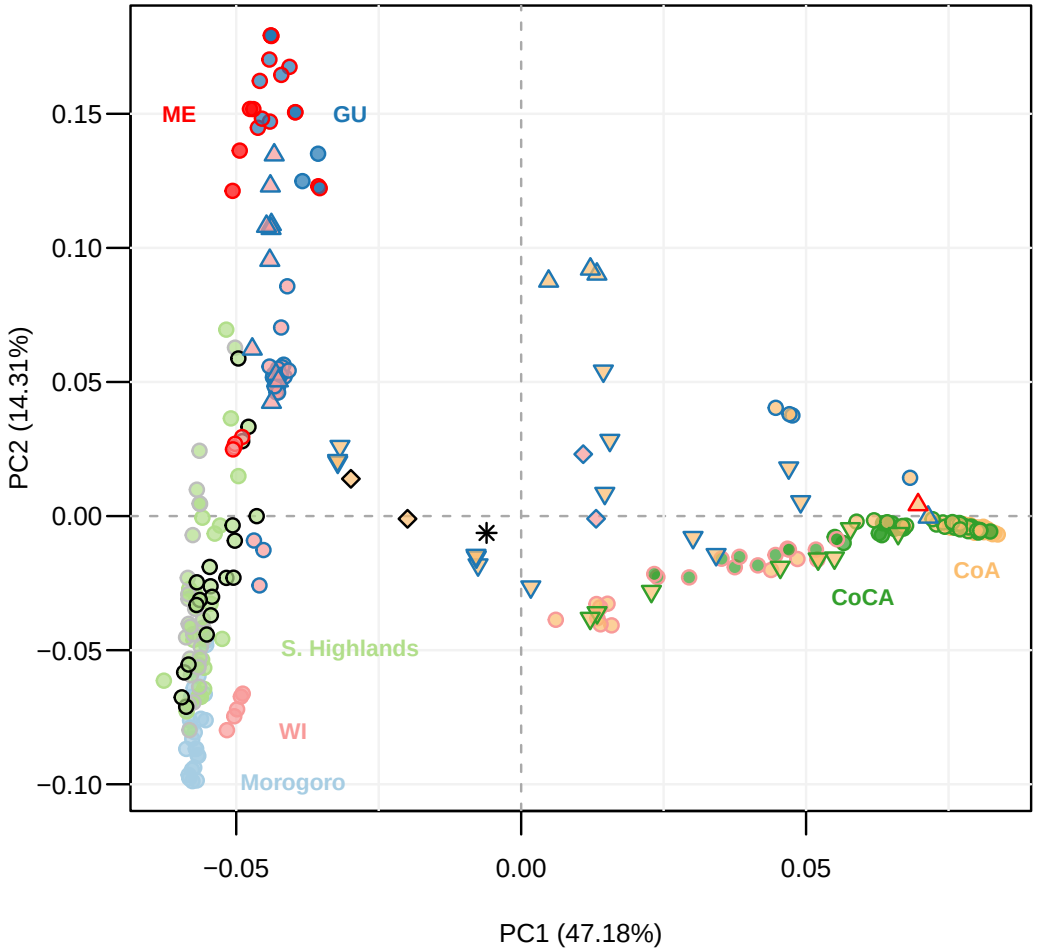


Figure 3. Principal component analysis (PCA) unsupervised genetic clustering of 95 Tanzanian avocado trees from the Morogoro low- to mid-altitude region (light blue) and the Southern Highlands (light green), together with 205 NCBI-available racial controls, all genotyped with lcWGS-derived SNPs. Variance explained by the two main principal components (PCs) is given in the axes' labels. Trees from the Southern Highlands in Tanzania are discriminated as follows: light green dots with green edges are from the Iringa region, while those with gray and black edges are from the Mbeya and Ruvuma regions, respectively. Racial controls are labeled and colored top down from left to right, following Figure 2: dark red for the Mexican (ME) race, dark blue for the Guatemalan (GU) race, light red for the West Indian (WI) race, and dark green and light orange for the Caribbean (CoCA) and Andean (CoA) groups from northwest South America. The remaining dots mark double hybrids between these last five groups—the interior and the edges are colored according to the putative parental groups. Triangles mark tripe hybrids, with the interior and the edges colored according to two of the putative parental groups, and its direction pointing towards the third contributing group. Diamonds mark tetra hybrids, with the interior and the edges colored according to the two most distant putative parental groups, while the other two groups can be resolved without ambiguity based on the proximity to the five racial control groups. Finally, the asterisk marks the penta hybrid among all five main groups (i.e., Mexican race, Guatemalan race, West Indian race, and Caribbean and Andean groups from northwest South America).

These population stratification patterns were replicable when exploring the high-coverage dataset of 10,460 SNPs (Figure S3). Finally, exploring the third main component against the first (Figure S4) and the second (Figure S5) did not offer further discrimination.

ADMIXTURE analysis showed the lowest cross-validation error at $K = 7$ (Figure S6), reinforcing the presence of seven genetic ancestries in the joint panel of 300 trees (Figure 4), and without $K > 7$ (Figure S7) offering additional discrimination. The hierarchy of ancestry segmentation from $K = 2$ to $K = 7$ reflected the patterns of genetic differentiation captured by the previous two main principal components. Specifically, at $K = 2$, the major distinction was between the Mesoamerican races (including the sampled trees from Tanzania) and the two northwest South American control groups. Within the former Mesoamerican cluster, the Mexican and Guatemalan races became distinct from the West Indian race and Tanzanian provenances at $K = 3$, while the latter South American cluster split between the Andean and Caribbean groups at $K = 4$. It was not until $K = 5$ that the Tanzanian provenances detached from the West Indian race, and at $K = 6$ that the disjunction between the Mexican and Guatemalan races was evident. Finally, at $K = 7$, the Tanzanian avocado trees from the Morogoro region and those from the Southern Highlands (i.e., Iringa, Mbeya and Ruvuma regions) separated. Most trees from Tanzania showed high ancestry proportions (>80%) to a single group, as did trees from northwest South America, while the Mesoamerican races exhibited more introgression (Table S8).

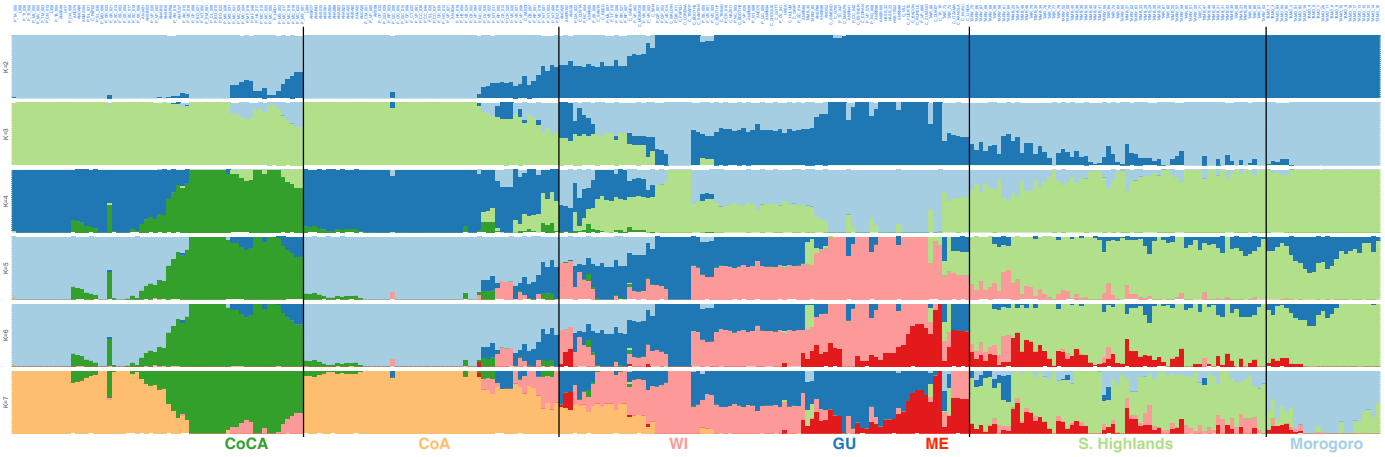


Figure 4. Admixture-based unsupervised genetic clustering (from $K = 2$ to $K = 7$) in 95 Tanzanian avocado tree samples from the low- to mid-altitude Morogoro region and the Southern Highlands, together with 205 NCBI-available racial controls [14], all genotyped with lcWGS-derived SNP markers. Linages at $K = 7$ are colored, from right to left, as in Figure 2: light blue for trees from Morogoro, light green for trees from the Southern Highlands, light red for the West Indian (WI) race, dark red for the Mexican (ME) race, dark blue for the Guatemalan (GU) race, and light orange and dark green for northwest South American Andean (CoA) and Caribbean (CoCA) groups.

3. Discussion

The whole-genome analysis presented here unveils for the first time avocado's diversity and population structure in Tanzania. Tanzanian trees consistently group closer to the West Indian race than to the Mexican or Guatemalan races. Subtle divergence is observed between trees from the low- to mid-altitude Morogoro region and the Southern Highlands, with some West Indian × Guatemalan and West Indian × Mexican introgression in the latter. PCA concordance between the full lcWGS dataset and the more stringent 10,460 high-coverage SNP panel reinforces these inferences. These results offer a whole-genome demographic background landscape against which further evolutionary, adaptative, and breeding signatures may be interpreted to boost local avocados.

3.1. Tanzania as a Secondary Center of Avocado Diversification

This study offers the first genome-wide characterization of avocado germplasm from Tanzania and reveals that populations from the Morogoro region and the Southern Highlands form a predominantly West Indian-derived genepool. The strong West Indian ancestry assignment (>80% in most trees) for the Tanzanian samples further suggests relatively recent population expansion from a narrow founder base population [38], rather than deep, long-term admixture across all major racial groups [39]. Nevertheless, the fact that we were also able to capture subtle genetic divergence between trees from the low- to mid-altitudes in Morogoro and those from the Southern Highlands suggests that Tanzania may be regarded as an emerging secondary center of diversification largely shaped by limited Mesoamerican introductions followed by extensive propagation and *in situ* adaptation to the highly heterogenous local topographic and climatic conditions [40].

Furthermore, as part of this diversification pattern, detectable incipient introgression with the other two Mesoamerican races in trees from the Southern Highlands suggests that local environmental and farmer-mediated selection may have indirectly favored segregating Mexican and Guatemalan alleles, already with standing adaptation to cooler mountainous environments. Specifically, trees from Morogoro are almost exclusively West Indian in ancestry, consistent with adaptation to low-elevation, warmer and high-rainfall environments [41,42]. In contrast, the West Indian × Guatemalan and West Indian × Mexican admixture of several trees in the Southern Highlands aligns with the agroecology of these environments, characterized by higher altitudes and cooler temperatures, conditions more similar to those in which the Guatemalan and Mexican races evolved [43]. Also, trees within the Southern Highlands appear largely panmictic across the Iringa, Mbeya and Ruvuma regions, confirming rampant gene flow across this highland agroecological zone rather than strong isolation at the regional level [16,35].

A non-mutually exclusive alternative to ecological and farmer-driven selection favoring introgressed genotypes better suited to cooler environments in the Southern Highlands is additional introductions of non-West Indian provenances. Mesoamerican introductions of fewer non-West Indian trees could have been followed by *in situ* hybridization and introgression [17]. However, at the moment, this ad hoc hypothesis may seem less parsimonious to make sense of the introgression observed in the Southern Highlands, given that no relatively pure Mexican or Guatemalan types were detected in the sampled areas. As a secondary step in the near future, extending the sampling efforts across other regions in Tanzania will help in assessing to what extend the observed regional differentiation reflects ecological filtering or historical re-introduction patterns.

In any case, from an evolutionary perspective, the patterns of allelic filtering and local introgression are both consistent with early-stage divergence in secondary centers, where ecological heterogeneity begins structuring ancestral variants without yet producing deeply differentiated lineages. Such dynamics are particularly relevant for seedling perennial tree crops [10], for which long generation times and allogamy continuously recombine and segregate variation [31,44,45], broadening the base for further selections.

3.2. Leveraging Avocado Breeding in East Africa

The predominance of West Indian ancestry in the studied Tanzanian avocado trees impacts varietal selection and rootstock breeding efforts [46–48]. West Indian avocados are typically associated with adaptation to lowland torrential tropical climates, higher tolerance to humidity, and larger and more watery fruits [42]. If these provenance's traits were to be extrapolated, the narrow racial base of avocados from the low- to mid-altitude Morogoro region and the Southern Highlands may have limited standing genetic variation for tolerance to emerging biotic [49–56] and abiotic stresses [57,58], not to mention consumer preferences towards more buttery fruits [36].

Given this bottleneck, the introgressed trees from the Southern Highlands may therefore represent key germplasm for more diversifying breeding [59]. For instance, West Indian × Guatemalan and West Indian × Mexican hybrid backgrounds may depict differential fruit phenotypes and boost adaptation to cooler climates. Such differential fruit phenotypes are supported by Juma et al. [16], who documented four flesh textures, i.e., buttery (58.3%), watery (21.8%), pastose (18.5%) and granular (1.4%), among 226 avocado trees from the Southern Highlands of Tanzania. There may also be additional hidden diversity yet to be disclosed in these populations, for which more exhaustive morpho-agronomic evaluations (as in [16]), extended sampling of additional 'plus trees', controlled *ex situ* provenances, progeny and clonal trials (e.g., [22]), and market trade surveys (i.e., [36]) would be highly desirable. Anyhow, the diversity in fruit texture highlights the presence of valuable quality-related traits within Southern Highlands germplasm. Therefore, these genotypes could eventually be utilized as novel clonal varieties by themselves, or alternatively as exotic donor candidates for pre-breeding pipelines, assisted by marker-guided or even genomic-enabled backcross selection schemes [34,60,61], aimed at introgressing differential fruit quality traits and climate change resilience [62].

Similarly, the fact that none of the Tanzanian samples clustered with the Andean or Caribbean groups indicates limited provenance or allelic contribution from northwestern South American germplasm. This recently discovered, yet ancient—originating in the Pleistocene—genepool is starting to be recognized as a cryptic pocket of avocado agrobiodiversity, more genetically distinctive than any Mesoamerican race [14,15]. Since Tanzanian avocado diversification in the studied regions has primarily occurred within the Mesoamerican racial genepool, with no introgression of South American alleles, infusion of elite trees from northwest South America in East Africa may therefore source field trials for rootstock section [63–65], and even boost the discovery of promising non-Hass like varieties for particular climates [58,66,67] and market niches [36].

In smallholder systems, farmer preferences and seed exchange networks often act as strong human-mediated selection forces [68]. Given that most trees sampled were 'plus trees', maintained and selected by local tribes for decades, it is plausible that unconscious or deliberate selection for fruit size, oil content, taste, or phenology [69] has already shaped allele frequency distributions. This dynamic can generate, over time, locally adapted landrace-like populations even in the absence of formal breeding programs [70]. The relatively high racial ancestry purity observed in many Tanzanian trees may therefore reflect both founder effects, as discussed in the previous section, or even directional and background selection matching local community market demands [36,71–73].

Meanwhile, the filtering of a 10,460 high-coverage SNP subset offers a more scalable marker panel for targeted applications [74], such as association mapping of key morpho-agronomic traits [75,76], genomic selection of saplings at the nursery stage [77–79], and germplasm diversity management [80,81]. For instance, once markers are escalated to larger training populations, the accuracy of prediction models for complex polygenic traits will increase [82,83], accelerating early indirect selection in nurseries for seedling orchards or as seedling rootstocks for clonal plantations [84].

As global avocado production increasingly relies on a narrower repertoire of commercial cultivars—notably var. Hass—safeguarding regionally adapted germplasm becomes paramount [85]. Secondary diversity centers such as Tanzania are likely to harbor novel allelic combinations conferring elite phenotypes as well as tolerances to yet unforeseen stresses in the region. The lcWGS screening made available here is the first step to characterizing and utilizing this understudied emerging germplasm. Such genomic resources will enable *in situ* conservation efforts, identifying promising as well as underrepresented racial provenances, and guiding hybrid and introgressive breeding.

3.3. Perspectives

Future work must envision extending the sampling effort across other Tanzanian regions and neighboring countries in order to improve our understanding of gene flow and diversification dynamics in the secondary diversity center of East Africa. Expanding the base population across local climatic gradients would in turn enable environmental association analyses capable of reconstructing the genomic bases of local adaptation [86], clarifying whether the observed regional stratification corresponds to adaptive divergent selection [87] rather than neutral demographic drift [88] and linage sorting [89]. Controlled provenance and progeny trials would also assist early forward selection of the identified 'plus trees' for yield [90], fruit quality [91,92], pathogen resistance, and adaptive traits. Last, moving beyond lcWGS, *de novo* depth-coverage sequencing will eventually allow for pangenomic reconstructions [93–95].

4. Materials and Methods

4.1. Tree Sampling

A total of 95 avocado 'plus trees' were sampled in early 2024 across the regions of Morogoro (Mvomero district: 498–1001 m a.s.l., n = 25), representing low- to mid-altitude conditions of eastern Tanzania, and Iringa (Iringa district: 1537–1715 m a.s.l., n = 20), Mbeya (Rungwe district: 931–1245 m a.s.l., n = 30) and Ruvuma (Madaba district: 1219–1450 m a.s.l., n = 20) in the Southern Highlands (Table S1). The overall sampling spanned a total four regions, four districts, and seven wards (Table 1).

Table 1. A summary of the 95 avocado 'plus trees' sampled in the Morogoro region and the Southern Highlands of Tanzania for low-coverage whole-genome resequencing (lcWGS). The table is sorted according to the elevation ranges. Coordinates are listed for each sampled tree under Table S1.

Region	District	Ward	Number of Samples	Elevation Range (m a.s.l.)
Morogoro	Mvomero	Mzumbe	16	498–936
Morogoro	Mvomero	Muhonda	4	504–516
Morogoro	Mvomero	Mgeta	5	995–1001
Mbeya	Rungwe	Masoko	15	931–1145
Mbeya	Rungwe	Lwanga	15	1172–1245
Ruvuma	Madaba	Wino	20	1219–1450
Iringa	Iringa DC	Ifunda	20	1537–1715

Six to eight fresh young leaves were collected from each tree, along with GPS coordinates and key tree dasometric measurements—such as tree height and trunk diameter. The leaves were then placed in porous tea bags (three to four leaves per bag) that were labeled with the corresponding sample ID. Thereafter, the bags were stored in plastic boxes containing moisture-absorbing color-indicating silica gel (Sigma Aldrich, Hamburg, Germany) for drying the leaves. Silica gel was replaced until the leaf samples were totally dry. All samples were collected in duplicate is separate boxes.

4.2. Genomic DNA Extraction and Low-Coverage Whole-Genome Sequencing (lcWGS)

Genomic DNA was extracted from the 95 silica-dried avocado leaf samples as per the DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) protocol. DNA was verified in Agarose gels and through the OD260/OD280 ratio (1.8–2.0, >0.5 µg) using a NanoDrop® 2000 spectrophotometer (Thermo Fisher, Waltham, MA, USA). Genomic libraries were prepared using the Illumina TruSeq™ Nano DNA Sample Prep Kit, and sequenced 150 bp pair-end in Illumina Novaseq at CD Genomics (Shirley, NJ, USA).

4.3. Sequencing Data Proccessing

Base calling with CASAVA software *v.*1.6 (Illumina Inc., San Diego, CA, USA) was performed to migrate sequencing data recorded in image files to sequence reads. Sequencing quality information was stored in FASTQ files, from which 3′ adapter contamination and low-quality reads were removed using Fastp *v.*0.20.0. Specifically, a 5 bp 3′-5′ sliding window was used to retrieve average *Q* scores. Bases within the window were removed for average *Q* scores < 20. Additionally, paired-end sequences were discarded if their length was <50 bp or the number of N bases was ≥5. Additionally, in order to guide racial ancestry assignment for the Tanzanian trees, FASTQ files from 205 avocados, including 42 commercial varieties, were retrieved from NCBI BioProject ID PRJNA1311704 [14].

Pre-processed data were mapped to the haplotype-resolved Queensland's avocado cv. Hass reference genome assembly (publicly available in the AvoBase repository: https://www.avocado.uma.es/easy_gdb/downloads.php, accessed on 23 March 2026) [30] using BWA *v.*0.7.12-r1039. Resultant SAM files were transformed to BAM files using Picard *v.*1.107 software (http://www.psc.edu/index.php/user-resources/software/picard, accessed on 23 March 2026) keeping paired-end reads' consistency with command *FixMateInformation*, and removing duplicates with the *MarkDuplicates* function.

GATK software *v.*4.5.0.0 was used for SNP calling [96]. Variant sites were called using the *Haplotype Caller* function. The resultant GVCF files per sample were merged by a chromosome with the *Combine GVCFs* function, and later combined across chromosomes with the *GatherVcfs* function. SNP markers were extracted from the integrated GVCF file with all samples and chromosomes using the *Select Variants* function, and quality filtering was performed with the *Variant Filtration* function. Specifically, SNPs were retained if Fisher's test of strand bias (FS) ≤ 60, mapping quality (MQ) ≥ 40, quality depth (QD) ≥ 2, ReadPosRankSum ≥ -8.0, and MQRankSum > -12.5. Retained SNP markers were annotated using ANNOVAR software *v.*4.0 [97]. Finally, a more depurated SNP dataset with higher coverage was obtained using VCFtools software *v.*2015 [98], only keeping markers with a minimum of 10× of sequencing depth (SD) per sample per position, missing data per variant not exceeding 20% (i.e., variants with genotypes present in at least 80% of the samples), and a minimum allele frequency (*maf*) of 5%.

4.4. Population Genomics Analysis

To infer the racial ancestry of the 95 Tanzanian samples from Morogoro and the Southern Highlands, phylogenies were built using the full and high-coverage SNP dataset with the 205 published avocado racial controls. Trees were reconstructed using Maximum Likelihood with Shimodaira–Hasegawa node support in Geneious *v.*2026.0.2 [99] FastTree plugin (http://www.microbesonline.org/fasttree/, accessed on 23 March 2026).

Two complementary unsupervised genetic clustering strategies were implemented to confirm the racial ancestry of the 95 Tanzanian samples by relying on the NCBI-available lcWGS data from 205 trees with reported race assignation. First, principal component analysis (PCA) was carried out, retaining three orthogonal dimensions, using GCTA software *v.*1.93.3 (https://github.com/jianyangqt/gcta, accessed on 23 March 2026). This analysis was carried out in the full panel as well as in the high-coverage SNP dataset, inspecting for each case two of the three principal components (PCs) at a time up to the third dimension for the full dataset (i.e., PC1 vs. PC2, PC1 vs. PC3, and PC2 vs. PC3). Second, admixture genetic structure was analyzed using ADMIXTURE software *v.*1.3.1 (http://dalexander.github.io/admixture, accessed on 23 March 2026) by assuming $K = 2 - 10$ *a priori* populations. The analysis provided, given a specific *K* value of assumed populations, a stacked bar plot detailing for each bar (individual) the population ancestry proportions. *K* value was optimized as per the profile of cross-validation (CV) error scores.

5. Conclusions

The lcWGS-screened Tanzanian avocado germplasm from the Morogoro region and the Southern Highlands constitutes a predominantly West Indian-derived secondary diversity genepool with introgression signatures from the Guatemalan and Mexican races, and emerging local differentiation. Specifically, the Morogoro low- to mid-altitude region had purer West Indian-type trees, whereas the Southern Highlands (i.e., Iringa, Mbeya and Ruvuma regions) harbored admixed genotypes with putative West Indian × Guatemalan and West Indian × Mexican ancestry, potentially adapted to cooler and high-altitude climates. These findings offer the first insights into the evolutionary history of avocado diversification in East Africa, providing an innovative genomic framework to further harness germplasm conservation and breeding efforts in secondary diversity centers.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ijms27073083/s1.

Author Contributions: A.J.C. conceived this study. I.J. and A.J.C. designed the sampling. I.J., J.M.H. and A.J.C. carried out field expeditions in the Morogoro low- to mid-altitude region and the Southern Highlands (i.e., Iringa, Mbeya and Ruvuma regions) of Tanzania during early 2024. A.J.C. processed data and prepared artwork. I.J. and A.J.C. interpreted the results. J.M.H. additionally assisted with the contextualization of outcomes. A.J.C. wrote the first draft of the manuscript. I.J. and J.M.H. substantially revised and edited subsequent versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the Swedish Research Council, Vetenskapsrådet—VR, through grant 2022-04411 awarded to A.J.C.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Processed data is already contained within the article and its Supplementary Materials. On the other hand, raw 95 FASTQ files are available through NCBI BioProject ID PRJNA1442766.

Acknowledgments: The authors are grateful to R. Ortiz for enabling the scientific networking that made this work possible. Fieldwork assistance from interns, district agricultural officers and local communities in early 2024 is equally acknowledged. Genotyping assistance from the staff of CD Genomics (NJ, USA), including N. Zhao, is very much appreciated. I.J. and A.J.C. also recognize B. Valencia's invitation to provocative discussions that took place during the African Plant Nutrition Institute (APNI) annual retreat in Sea Cliff, Zanzibar, during September 2024. S. Cong is acknowledged for accompanying and supporting the submission. A.J.C recognizes the Department of Food Science and Technology in the College of Agricultural Sciences and Food Technology (CoAF) at University of Dar es Salaam (UDSM) for hosting him through permit AD.203/319/01 from 16 October 2023. Lastly, I.J. and A.J.C are very appreciative that this research was supported by the University of Dar es Salaam (UDSM) through Memorandum of Understating (MOU) SLU.ltv.2023.2.6- with the Swedish University of Agricultural Sciences (Sveriges Lantbruksuniversitet—SLU), and approved by the Tanzania Agricultural Research Institute (TARI) through research clearance on 25 September 2023 and the Tanzania Commission for Science and Technology (COSTECH) through research permit 2023-914-NA-2023-1016 on 20 October 2023.

Conflicts of Interest: The authors declare that this research was conducted without any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Juma, I. A review on avocado (*Persea americana* Mill.) leaf: Morphology, nutritional and phytochemical profiles, functional properties, health benefits, and toxicity. *Genet. Resour. Crop Evol.* **2025**, *72*, 9087–9103. [CrossRef]
2. Calderón-Vázquez, C.; Durbin, M.L.; Ashworth, V.E.T.M.; Tommasini, L.; Meyer, K.K.T.; Clegg, M.T. Quantitative Genetic Analysis of Three Important Nutritive Traits in the Fruit of Avocado. *J. Am. Soc. Hortic. Sci.* **2013**, *138*, 283–289. [CrossRef]
3. Bhuyan, D.J.; Alsherbiny, M.A.; Perera, S.; Low, M.; Basu, A.; Devi, O.A.; Barooah, M.S.; Li, C.G.; Papoutsis, K. The odyssey of bioactive compounds in avocado (*Persea americana*) and their health benefits. *Antioxidants* **2019**, *8*, 426. [CrossRef]
4. Juma, I. Nature's Pharmacy: Avocado (*Persea americana* Mill.) Leaves as a Novel Therapeutic Agent for Diabetes, Hypertension, High Cholesterol, Gastric Ulcers, Kidney Stones, and Gout. *Int. Res. J. Biol. Sci.* **2024**, *13*, 27–32.
5. Rodríguez, P.; Cortes-Herrera, C.; Artavia, G.; Granados-Chinchilla, F.; Vaillant, F. Comparative Analysis of Bioactive Compounds in Colombian Caribbean Avocados and Hass: Exploring Metabolomic Diversity. *ACS Food Sci. Technol.* **2025**, *5*, 3191–3200. [CrossRef]
6. Hausch, B.J.; Arpaia, M.L.; Kawagoe, Z.; Walse, S.; Obenland, D. Chemical Characterization of Two California-Grown Avocado Varieties (*Persea americana* Mill.) over the Harvest Season with an Emphasis on Sensory-Directed Flavor Analysis. *J. Agric. Food Chem.* **2020**, *68*, 15301–15310. Correction in *J. Agric. Food. Chem.* **2025**, *73*, 20526. https://doi.org/10.1021/acs.jafc.5c09310. [CrossRef]
7. Bergh, B.; Ellstrand, N. Taxonomy of the Avocado. *Calif. Avocado Soc.* **1986**, *70*, 135–146.
8. Chen, H.; Morrell, P.L.; Ashworth, V.E.; de la Cruz, M.; Clegg, M.T. Tracing the geographic origins of major avocado cultivars. *J. Hered.* **2009**, *100*, 56–65. [CrossRef]
9. Ge, Y.; Dong, X.; Wu, B.; Wang, N.; Chen, D.; Chen, H.; Zou, M.; Xu, Z.; Tan, L.; Zhan, R. Evolutionary analysis of six chloroplast genomes from three *Persea americana* ecological races: Insights into sequence divergences and phylogenetic relationships. *PLoS ONE* **2019**, *14*, e0221827. [CrossRef]
10. Boza, E.; Tondo, C.; Ledesma, N.; Campbell, R.; Bost, J.; Schnell, R.; Gutiérrez, O. 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]
11. Ashworth, V.E.; Clegg, M.T. Microsatellite markers in avocado (*Persea americana* Mill.): Genealogical relationships among cultivated avocado genotypes. *J. Hered.* **2003**, *94*, 407–415. [CrossRef]
12. Galindo-Tovar, M.E.; Ogata-Aguilar, N.; Arzate-Fernández, A.M. Some aspects of avocado (*Persea americana* Mill.) diversity and domestication in Mesoamerica. *Genet. Resour. Crop Evol.* **2007**, *55*, 441–450. [CrossRef]
13. Galindo-Tovar, M.E.; Arzate-Fernández, A.M.; Ogata-Aguilar, N.; Landero-Torres, I. The Avocado (*Persea americana*, *Lauraceae*) Crop in Mesoamerica: 10,000 Years of History. *Harv. Pap. Bot.* **2007**, *12*, 325–334. [CrossRef]
14. Cañas-Gutierrez, G.P.; López-Hernandez, F.; Cortés, A.J. Whole Genome Resequencing of 205 Avocado Trees Unveils the Genomic Patterns of Racial Divergence in the Americas. *Int. J. Mol. Sci.* **2025**, *26*, 10353. [CrossRef]
15. Berdugo-Cely, J.A.; Cortés, A.J.; López-Hernández, F.; Delgadillo-Durán, P.; Cerón-Souza, I.; Reyes-Herrera, P.H.; Navas-Arboleda, A.A.; Yockteng, R. Pleistocene-dated genomic divergence of avocado trees supports cryptic diversity in the Colombian germplasm. *Tree Genet. Genomes* **2023**, *19*, 42. [CrossRef]
16. 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]
17. 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]
18. Rendón-Anaya, M.; Ibarra-Laclette, E.; Méndez-Bravo, A.; Lan, T.; Zheng, C.; Carretero-Paulet, L.; Perez-Torres, C.; Chacón-López, A.; Hernandez-Guzmán, G.; Chang, T.; et al. The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 17081–17089. [PubMed]
19. 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] [PubMed]
20. Ruiz-Chután, J.A.; Kalousová, M.; Maňourová, A.; Degu, H.D.; Berdúo-Sandoval, J.E.; Villanueva-González, C.E.; Lojka, B. Core Collection Formation in Guatemalan Wild Avocado Germplasm with Phenotypic and SSR Data. *Agronomy* **2023**, *13*, 2385. [CrossRef]
21. 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*, 1–14. [CrossRef]

22. 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]
23. Reyes-Herrera, P.H.; Muñoz-Baena, L.; Velásquez-Zapata, V.; Patiño, L.; Delgado-Paz, O.A.; Díaz-Diez, C.A.; Navas-Arboleda, A.A.; Cortés, A.J. Inheritance of Rootstock Effects in Avocado (*Persea americana* Mill.) cv. Hass. *Front. Plant Sci.* **2020**, *11*, 555071. [CrossRef]
24. Cañas-Gutierrez, G.P.; Arango-Isaza, R.E.; Saldamando-Benjumea, C.I. Microsatellites revealed genetic diversity and population structure in Colombian avocado (*Persea americana* Mill.) germplasm collection and its natural populations. *J. Plant Breed. Crop Sci.* **2019**, *11*, 106–119. [CrossRef]
25. Ferrer-Pereira, H.; Pérez-Almeida, I.; Raymúndez-Urrutia, M.; Suárez, L. Genetic relationship analysis for avocado cultivars from Venezuelan germplasm bank (INIA-CENIAP) using molecular markers. *Tree Genet. Genomes* **2017**, *13*, 82. [CrossRef]
26. 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, Erratum in *Sci. Rep.* **2023**, *13*, 3016.. [CrossRef]
27. 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. *Turk. J. Agric. For.* **2019**, *43*, 437–449. [CrossRef]
28. Wienk, R.; Mostert-O'Neill, M.; Abeysekara, N.; Manosalva, P.; Freeman, B.; Van den Berg, N. Genetic diversity, population structure, and clonal veri cation in South African avocado cultivars using single nucleotide polymorphism (SNP) markers. *Tree Genet. Genomes* **2022**, *18*, 41. [CrossRef]
29. Sharon, D.; Cregan, P.B.; Mhameed, S.; Kusharska, M.; Hillel, J.; Lahav, E.; Lavi, U. An integrated genetic linkage map of avocado. *Theor. Appl. Genet.* **1997**, *95*, 911–921. [CrossRef]
30. Nath, O.; Fletcher, S.J.; Hayward, A.; Shaw, L.M.; Masouleh, A.K.; Furtado, A.; Henry, R.J.; Mitter, N. A haplotype resolved chromosomal level avocado genome allows analysis of novel avocado genes. *Hortic. Res.* **2022**, *9*, uhac157. [CrossRef]
31. Solares, E.; Morales-Cruz, A.; Balderas, R.F.; Focht, E.; Ashworth, V.E.T.M.; Wyant, S.; Minio, A.; Cantu, D.; Arpaia, M.L.; Gaut, B.S.; et al. Insights into the domestication of avocado and potential genetic contributors to heterodichogamy. *G3 Genes Genomes Genet.* **2023**, *13*, jkac323. [CrossRef]
32. Cortés, A.J.; Restrepo-Montoya, M.; Bedoya-Canas, L.E. Modern Strategies to Assess and Breed Forest Tree Adaptation to Changing Climate. *Front. Plant Sci.* **2020**, *11*, 583323. [CrossRef]
33. Guevara-Escudero, M.; Osorio, A.N.; Cortés, A.J. Integrative Pre-Breeding for Biotic Resistance in Forest Trees. *Plants* **2021**, *10*, 2022. [CrossRef] [PubMed]
34. Badenes, M.L.; Fernandez, I.M.A.; Rios, G.; Rubio-Cabetas, M.J. Application of Genomic Technologies to the Breeding of Trees. *Front. Genet.* **2016**, *7*, 198. [CrossRef] [PubMed]
35. 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]
36. Juma, I.; Fors, H.; Hovmalm, H.P.; Nyomora, A.; Fatih, M.; Geleta, M.; Carlsson, A.S.; Ortiz, R.O. Avocado Production and Local Trade in the Southern Highlands of Tanzania: A Case of an Emerging Trade Commodity from Horticulture. *Agronomy* **2019**, *9*, 749. [CrossRef]
37. Fuentes-Pardo, A.P.; Ruzzante, D.E. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Mol. Ecol.* **2017**, *26*, 5369–5406. [CrossRef]
38. Wakeley, J. *Coalescent Theory: An Introduction*; Harvard University: Cambridge, MA, USA, 2008.
39. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585–595. [CrossRef]
40. Juma, I.; Valencia, J.B.; Cortés, A.J. Predicting Suitable Regions for Avocado (*Persea americana* Mill.) Tree Cultivation in Tanzania. *Horticulturae* **2025**, *12*, 24. [CrossRef]
41. Ramírez-Gil, J.G.; Henao-Rojas, J.C.; Diaz-Diez, C.A.; Peña-Quiñones, A.J.; León, N.; Parra-Coronado, A.; Bernal-Estrada, J.A. Phenological variations of avocado cv. Hass and their relationship with thermal time under tropical conditions. *Heliyon* **2023**, *9*, e19642. [CrossRef]
42. Sandoval-Castro, E.; Peraza-Magallanes, A.Y.; Dodd, R.S.; Ashworth, V.E.T.M.; Cruz-Mendívil, A.; Calderón-Vásquez, C.L. Exploring genetic diversity of lowland avocado (*Persea americana* Mill.) as a genetic reservoir for breeding. *Genet. Resour. Crop Evol.* **2021**, *68*, 2757–2766. [CrossRef]
43. Sánchez-González, E.I.; Gutiérrez-Soto, J.G.; Olivares-Sáenz, E.; Gutiérrez-Díez, A.; Barrientos-Priego, A.F.; Ochoa-Ascencio, S. Screening Progenies of Mexican Race Avocado Genotypes for Resistance to *Phytophthora cinnamomi* Rands. *HortScience* **2019**, *54*, 809–813. [CrossRef]
44. 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]

45. Salazar-García, S.; Garner, L.C.; Lovatt, C.J. Reproductive Biology. In *The Avocado: Botany, Production and Uses*, 2nd ed.; Schaffer, B., Wolstenholme, N., Whiley, A.W., Eds.; CABI: Boston, MA, USA, 2013; pp. 118–167.
46. Herrera-González, J.A.; Salazar-García, S.; Gutiérrez-Martínez, P.; González-Durán, I.J.L. Postharvest performance of 'Hass' avocado fruit is influenced by rootstock. *Rev. Mex. Cienc. Agrícolas* **2013**, *4*, 19–32.
47. Salazar-García, S.; Velasco-Cárdenas, J.J.; Medina-Torres, R.; Gómez-Aguilar, J.R. Avocado selections with potential use as rootstocks. I. Grafting and scion growth. *Rev. Fitotec. Mex.* **2004**, *27*, 23–30.
48. Mickelbart, M.V.; Bender, G.S.; Witney, G.W.; Adams, C.; Arpaia, M.L. Effects of clonal rootstocks on 'Hass' avocado yield components, alternate bearing, and nutrition. *J. Hortic. Sci. Biotechnol.* **2007**, *82*, 460–466. [CrossRef]
49. López-Galé, Y.D.; Martínez, M.F.; Rios-Rojas, L.; García-Dávila, M.A. Avocado genotypes resistant to *Phytophthora cinnamomi* in Colombia. *Sci. Agric.* **2025**, *82*, e20240117. [CrossRef]
50. Fick, A.; Swart, V.; Bombarely, A.; van den Berg, N. Comparative transcriptional analysis of *Persea americana MYB*, *WRKY* and *AP2/ERF* transcription factors following *Phytophthora cinnamomi* infection. *Mol. Plant Pathol.* **2024**, *25*, e13453. [CrossRef]
51. Fick, A.; Swart, V.; Backer, R.; Bombarely, A.; Engelbrecht, J.; van den Berg, N. Partially Resistant Avocado Rootstock Dusa® Shows Prolonged Upregulation of Nucleotide Binding-Leucine Rich Repeat Genes in Response to *Phytophthora cinnamomi* Infection. *Front. Plant Sci.* **2022**, *13*, 793644. [CrossRef]
52. Toapanta-Gallegos, D.E.; Morillo-Velastegui, L.E.; Viera-Arroyo, W.F. Molecular diagnosis of *Phytophthora cinnamomi* associated with root rot in avocado producing areas of Ecuador. *Cienc. Tecnol. Agropecu.* **2017**, *18*, 285. [CrossRef]
53. Reeksting, B.J.; Olivier, N.A.; van den Berg, N. Transcriptome responses of an ungrafted *Phytophthora* root rot tolerant avocado (*Persea americana*) rootstock to flooding and *Phytophthora cinnamomi*. *BMC Plant Biol.* **2016**, *16*, 205. [CrossRef] [PubMed]
54. Smith, L.A.; Dann, E.K.; Pegg, K.G.; Whiley, A.W.; Giblin, F.R.; Doogan, V.; Kopittke, R. Field assessment of avocado rootstock selections for resistance to *Phytophthora* root rot. *Australas. Plant Pathol.* **2011**, *40*, 39–47. [CrossRef]
55. Coffey, M.D. *Phytophthora* root rot of avocado. *Plant Dis.* **1987**, *71*, 1046–1052.
56. Willingham, S.L.; Pegg, K.G.; Cooke, A.W.; Coates, L.M.; Langdon, P.W.B.; Dean, J.R. Rootstock influences postharvest anthracnose development in 'Hass' avocado. *Aust. J. Agric. Res.* **2001**, *52*, 1017–1022. [CrossRef]
57. Lazare, S.; Yasuor, H.; Yermiyahu, U.; Kuhalskaya, A.; Brotman, Y.; Ben-Gal, A.; Dag, A. It takes two: Reciprocal scion-rootstock relationships enable salt tolerance in 'Hass' avocado. *Plant Sci.* **2021**, *312*, 111048. [CrossRef]
58. Bernstein, N.; Loffe, M.; Zilberstaine, M. Salt-stress effects on avocado rootstock growth. I. Establishing criteria for determination of shoot growth sensitivity to the stress. *Plant Soil* **2001**, *233*, 1–11. [CrossRef]
59. McCouch, S. Diversifying Selection in Plant Breeding. *PLoS Biol.* **2004**, *2*, 1507–1512. [CrossRef]
60. Iwata, H.; Minamikawa, M.F.; Kajiya-Kanegae, H.; Ishimori, M.; Hayashi, T. Genomics-assisted breeding in fruit trees. *Breed. Sci.* **2016**, *66*, 100–115. [CrossRef]
61. Blair, M.W.; Izquierdo, P. Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. *Theor. Appl. Genet.* **2012**, *125*, 1015–1031. [CrossRef]
62. Vetharaniam, I.; Stanley, C.J.; Cummins, M.; van den Dijssel, C.; Müller, K. Modelling Climate Change Impacts on Location Suitability for Cultivating Avocado and Blueberry in New Zealand. *Land* **2024**, *13*, 1753. [CrossRef]
63. Lazare, S.; Haberman, A.; Yermiyahu, U.; Erel, R.; Simenski, E.; Dag, A. Avocado rootstock influences scion leaf mineral content. *Arch. Agron. Soil Sci.* **2020**, *66*, 1399–1409. [CrossRef]
64. Ceballos, R.; Rioja, T. Rootstock affects the blend of biogenic volatile organic compounds emitted by 'Hass' avocado. *Chil. J. Agric. Res.* **2019**, *79*, 330–334. [CrossRef]
65. Mickelbart, M.V.; Arpaia, M.L. Rootstock Influences Changes in Ion Concentrations, Growth, and Photosynthesis of 'Hass' Avocado Trees in Response to Salinity. *J. Am. Soc. Hortic. Sci.* **2002**, *127*, 649–655. [CrossRef]
66. Tamayo-Vélez, Á.d.J.; Bernal-Estrada, J.A.; Díaz-Díez, C.A.; Cano-Gallego, L.E.; Córdoba-Gaona, O. Effect of rootstock/scion compatibility on fruit and foliar nutrient composition in avocado (*Persea americana* Mill.) cv. Hass in Colombia. *Rev. Colomb. Cienc. Hortícolas* **2022**, *16*, e14833. [CrossRef]
67. Shu, B.; Liu, L.; Jue, D.; Wang, Y.; Wei, Y.; Shi, S. Effects of avocado (*Persea americana* Mill.) scion on arbuscular mycorrhizal and root hair development in rootstock. *Arch. Agron. Soil Sci.* **2017**, *63*, 1951–1962. [CrossRef]
68. Peláez, D.; Aguilar, P.A.; Mercado, M.; López-Hernández, F.; Guzmán, M.; Burbano-Erazo, E.; Denning-James, K.; Medina, C.I.; Blair, M.W.; De Vega, J.J.; et al. Genotype selection, and seed uniformity and multiplication to ensure common bean (*Phaseolus vulgaris* L.) var. Liborino. *Agronomy* **2022**, *12*, 2285. [CrossRef]
69. Alcaraz, M.L.; Hormaza, J.I. Fruit Set in Avocado: Pollen Limitation, Pollen Load Size, and Selective Fruit Abortion. *Agronomy* **2021**, *11*, 1603. [CrossRef]
70. Gepts, P.; Papa, R. Possible effects of (trans)gene flow from crops on the genetic diversity from landraces and wild relatives. *Environ. Biosaf. Res.* **2003**, *2*, 89–103. [CrossRef]
71. Smale, M.; Jamora, N. Valuing genebanks. *Food Secur.* **2020**, *12*, 905–918. [CrossRef]

72. Sommaruga, R.; Eldridge, H.M. Avocado Production: Water Footprint and Socio-economic Implications. *EuroChoices* **2020**, *20*, 48–53. [CrossRef]
73. Fulgoni, V.L.; Dreher, M.; Davenport, A.J. Avocado consumption is associated with better diet quality and nutrient intake, and lower metabolic syndrome risk in US adults: Results from the National Health and Nutrition Examination Survey (NHANES) 2001–2008. *Nutr. J.* **2013**, *12*, 1. [CrossRef]
74. Kuhn, D.N.; Livingstone, D.S.; Richards, J.H.; Manosalva, P.; Van den Berg, N.; Chambers, A.H. Application of genomic tools to avocado (*Persea americana*) breeding: SNP discovery for genotyping and germplasm characterization. *Sci. Hortic.* **2019**, *246*, 1–11. [CrossRef]
75. Arenas, S.; Cortés, A.J. Unveiling the Genomic Architecture of Phenotypic Plasticity Using Multiple GWAS Approaches Under Contrasting Conditions of Water Availability: A Model for Barley. *Int. J. Mol. Sci.* **2026**, *27*, 652. [CrossRef]
76. Denning-James, K.E.; Chater, C.; Cortés, A.J.; Blair, M.W.; Peláez, D.; Hall, A.; De Vega, J.J. Genome-wide association mapping dissects the selective breeding of determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris* L.). *G3 Genes Genomes Genet.* **2025**, *15*, jkaf090. [CrossRef]
77. Bedoya-Londoño, S.; Cañas-Gutiérrez, G.P.; Cortés, A.J. Breeding Without Breeding: Enabling Indirect Selection Schemes for Tropical Tree Improvement. In *Genomics Based Approaches for Tropical Tree Improvement and Conservation*; Springer Nature: Singapore, 2025; pp. 19–42. [CrossRef]
78. Arenas, S.; Cortés, A.J.; Jaramillo-Correa, J.P. Toward a Genomic-Enabled Selection in Natural Tree Populations for Long-Term Management and Conservation. In *Genomics Based Approaches for Tropical Tree Improvement and Conservation*; Springer Nature: Singapore, 2025; pp. 243–278. [CrossRef]
79. Parra-Londono, S.; López-Hernández, F.; Montoya, G.; Henao-Rojas, J.C.; Ossa-Ossa, G.A.; Cortés, A.J. Merging Phenotypic Stability Analysis and Genomic Prediction for Multi-Environment Breeding in *Capsicum* spp. *Agronomy* **2025**, *15*, 2690. [CrossRef]
80. Kuhn, D.N.; Groh, A.; Rahaman, J.; Freeman, B.; Arpaia, M.L.; Van den Berg, N.; Abeysekara, N.; Manosalva, P.; Chambers, A.H. Creation of an avocado unambiguous genotype SNP database for germplasm curation and as an aid to breeders. *Tree Genet. Genomes* **2019**, *15*, 71. [CrossRef]
81. Ramírez-Gil, J.G.; López-Hernandez, F.; Conejo-Rodriguez, D.F.; Henao-Rojas, J.C.; Quiroga-Benavides, K.E.; Cortés, A.J.; Chacón-Sanchez, M.I. GermVersity: A free and user-friendly interface to enhance the visualization and analysis of genebank data. *PLoS ONE* **2026**, *21*, e0340826. [CrossRef]
82. López-Hernández, F.; Villanueva-Mejía, D.F.; Tofiño-Rivera, A.P.; Cortés, A.J. Genomic prediction of adaptation in common bean (*Phaseolus vulgaris* L.) × Tepary bean (*P. acutifolius* A. Gray) hybrids. *Int. J. Mol. Sci.* **2025**, *26*, 7370. [CrossRef] [PubMed]
83. Arenas, S.; Cortés, A.J.; Mastretta-Yanes, A.; Jaramillo-Correa, J.P. Evaluating the accuracy of genomic prediction for the management and conservation of relictual natural tree populations. *Tree Genet. Genomes* **2021**, *17*, 12. [CrossRef]
84. Ernst, A.A. Micro cloning: A multiple cloning technique for avocados using micro containers. *Rev. Chapingo Ser. Hortic.* **1999**, *5*, 217–220.
85. Cortés, A.J. Unlocking genebanks for climate adaptation. *Nat. Clim. Change* **2025**, *15*, 590–592. [CrossRef]
86. Lasky, J.R.; Josephs, E.B.; Morris, G.P. Genotype-environment associations to reveal the molecular basis of environmental adaptation. *Plant Cell* **2023**, *35*, 125–138. [CrossRef] [PubMed]
87. Barton, N.H.; Etheridge, A.M. The effect of selection on genealogies. *Genetics* **2004**, *166*, 1115–1131.
88. Fu, R.; Gelfand, A.E.; Holsinger, K.E. Exact moment calculations for genetic models with migration, models with migration, models with migration, mutation, and drift. *Theor. Popul. Biol.* **2003**, *63*, 231–243.
89. Maddison, W.P.; Knowles, L.L. Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* **2006**, *55*, 21–30. [CrossRef]
90. Wilkie, J.D.; Conway, J.; Griffin, J.; Toegel, H. Relationships between canopy size, light interception and productivity in conventional avocado planting systems. *J. Hortic. Sci. Biotechnol.* **2018**, *94*, 481–487. [CrossRef]
91. Rodriguez, P.; Henao, J.C.; Correa, G.; Aristizabal, A. Identification of Harvest Maturity Indicators for 'Hass' Avocado Adaptable to Field Conditions. *HortTechnology* **2018**, *28*, 815–821. [CrossRef]
92. Rodriguez, P.; Astudillo, C. Physicochemical parameters of avocado *Persea americana* Mill. cv. Hass (*Lauraceae*) grown in Antioquia (Colombia) for export. *Cienc. Tecnol. Agropecu.* **2018**, *19*, 383–392. [CrossRef]
93. Vega-Munoz, M.A.; López-Hernandez, F.; Cortés, A.J.; Roda, F.; Castano, E.; Montoya, G.; Henao-Rojas, J.C. Pangenomic and Phenotypic Characterization of Colombian *Capsicum* Germplasm Reveals the Genetic Basis of Fruit Quality Traits. *Int. J. Mol. Sci.* **2025**, *26*, 8205. [CrossRef]
94. López-Hernández, F.; Roda, F.; Cortés, A.J. Machine Learning and Pangenomics: Revolutionizing Plant Breeding for a Sustainable Future. In *Plant Breeding 2050: Next-Gen Crops*; Priyadarshan, P.M., Ortiz, R., Eds.; Springer: Singapore, 2025; pp. 509–523. [CrossRef]
95. Hu, H.; Zhao, J.; Thomas, W.J.W.; Batley, J.; Edwards, D. The role of pangenomics in orphan crop improvement. *Nat. Commun.* **2025**, *16*, 118. [CrossRef] [PubMed]

96. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]
97. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef]
98. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
99. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.