

Development and Application of Microsatellite Markers for Genetic Diversity Assessment and Construction of a Core Collection of *Myrciaria dubia* Germplasm from the Peruvian Amazon

[Juan C. Castro](#)^{*}, Stalin J. Vasquez-Guizado, Bianca E. Vigil, Francisco Ascue, Naysha Rojas-Villa, Jae D. Paredes, Marianela Cobos, Carlos G. Castro, Daniel E. Motta, [Pedro M. Adrianzen](#), [Sixto A. Imán](#), [J. Dylan Maddox](#)^{*}

Posted Date: 31 July 2024

doi: 10.20944/preprints202407.2528.v1

Keywords: Amazonian crop; ex situ conservation; core collection; molecular markers; plant genetic resources; simple sequence repeats



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Development and Application of Microsatellite Markers for Genetic Diversity Assessment and Construction of a Core Collection of *Myrciaria dubia* Germplasm from the Peruvian Amazon

Juan C. Castro ^{1,2,*}, Stalin J. Vasquez-Guizado ³, Bianca E. Vigil ⁴, Francisco Ascue ⁵,
Naysha Rojas-Villa ⁶, Jae D. Paredes ¹, Marianela Cobos ^{1,2}, Carlos G. Castro ¹, Daniel E. Motta ¹,
Pedro M. Adrianzen ^{1,2}, Sixto A. Imán ^{7,8} and J. Dylan Maddox ^{9,10,*}

¹ Unidad Especializada del Laboratorio de Investigación en Biotecnología (UELIB), Centro de Investigaciones de Recursos Naturales de la UNAP (CIRNA), Universidad Nacional de la Amazonía Peruana (UNAP), Iquitos, Perú; jaedianap@gmail.com (J.D.P.); marianela.cobos@unapiquitos.edu.pe (M.C.); castrocoboscarlosgilbertox13@gmail.com (C.G.C.); dmottas17@gmail.com (D.E.M.); pedro.adrianzen@unapiquitos.edu.pe (P.M.A.)

² Departamento Académico de Ciencias Biomédicas y Biotecnología; Facultad de Ciencias Biológicas, Universidad Nacional de la Amazonía Peruana (UNAP)

³ Universidad Nacional de San Agustín de Arequipa; svasquezg@unsa.edu.pe

⁴ FORESLAB; bianca.vigil@foreslab.com

⁵ Universidad de Ingeniería y Tecnología; francisco.ascue@utec.edu.pe

⁶ Centro Internacional de Investigación para la Sustentabilidad, Universidad Nacional de Cañete; n_rojas@undc.edu.pe

⁷ Estación Experimental Agraria San Roque, Dirección de Recursos Genéticos y Biotecnología, Instituto Nacional de Innovación Agraria (INIA); siman@inia.gob.pe or sixto.iman@unapiquitos.edu.pe

⁸ Departamento Académico de Ecología y Conservación, Facultad de Ciencias Forestales, Universidad Nacional de la Amazonía Peruana (UNAP), Iquitos, Perú

⁹ Pritzker Laboratory for Molecular Systematics and Evolution, Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, IL 60605, USA

¹⁰ Environmental Sciences, American Public University System, Charles Town, WV 25414, USA

* Correspondence: juan.castro@unapiquitos.edu.pe (J.C.C.); dmaddox@fieldmuseum.org (J.D.M.); Tel.: +51-965-017215 (J.C.C.); +1-(217)-390-5081 (J.D.M.)

Abstract: The Amazonian shrub *Myrciaria dubia* (Kunth) McVaugh, known as “camu-camu”, produces vitamin C-rich fruits of growing commercial interest. However, sustainable utilization requires assessment and protection of the genetic diversity of the available germplasm. We hypothesized that the ex situ *M. dubia* germplasm bank assembled from eight river basins across the Peruvian Amazon would harbor substantial genetic diversity and have a genetic population structure. This study aimed to (1) develop new polymorphic microsatellite markers, (2) characterize genetic diversity and validate the hypothesis in the ex situ germplasm, and (3) construct a core subset representing maximum allelic variability. Sixteen polymorphic microsatellite loci were developed using an enrichment approach. The evaluation of 336 genotypes from 43 accessions originating from eight river basins of the germplasm bank corroborated this hypothesis, revealing high gene diversity, with observed heterozygosity ranging from 0.468 to 0.644 and expected heterozygosity from 0.684 to 0.817 at the river basin level. Analysis of molecular variance showed a higher genetic variation within accessions and river basins, at 73% and 86%, respectively, than among accessions and river basins, at 27% and 14%, respectively. Bayesian clustering detected the presence of ten genetic clusters, with several degrees of admixture among river basins. The Putumayo River basin showed a clear genetic homogeneity. A core collection of 84 genotypes was constructed, thus covering 86.7% of the global allelic diversity. These results have important implications for *M. dubia* conservation strategies and breeding programs, in demonstrating a need for genetic connectivity between populations but preserving unique genetic resources in isolated

basins. These results validate the expected levels of diversity and population subdivision in a crop and stress the need to secure genetically diverse germplasms, underscoring the importance of thorough genetic characterization for ex situ germplasm management.

Keywords: amazonian crop; ex situ conservation; core collection; molecular markers; plant genetic resources; simple sequence repeats.

1. Introduction

Myrciaria dubia (H.B.K) McVaugh, camu-camu, is a shrub native to the Amazon rainforest [1]. It belongs to the Myrtaceae family and is an underutilized crop that produces one of the richest natural sources of vitamin C, as its fruits have up to 7355 mg L-ascorbic acid per 100 g of their pulp [2,3]. In addition to its exceptionally high vitamin C content, camu-camu fruit is a good source of other bioactive compounds with antioxidant and anti-inflammatory properties, such as anthocyanins, flavonoids, and phenolic acids [4–7]. The growing demand for natural sources of antioxidants and functional foods has increased interest in domesticating and commercializing camu-camu [8,9].

While the nutritional and economic potential of camu-camu has driven interest in domestication, long-term conservation and sustainable utilization of this genetic resource requires securing representative germplasm ex situ. Establishing ex situ field gene banks and seed banks is an important strategy, but maximizing captured genetic diversity remains a major challenge [10,11]. Many ex situ collections, however, have been founded based on comprehensive sampling in the native distribution area of plant species. This can give a better representation of the gene pool and reduce the potential for the processes of genetic erosion in successive generations [12,13]. Questions relating to these concerns are discussed in several publications about ex situ conservation and genetic diversity [14–19]. Thus, comprehensive assessments of genetic diversity in available ex situ germplasms become very critical in optimizing management practices, identifying unique accessions, and delineating subsets that maximize diversity for breeding and long-term conservation [10,20]. For *M. dubia*, studies evaluating genetic diversity within and among ex situ collections have been sparse.

Microsatellites or simple sequence repeats (SSRs) are highly polymorphic codominant molecular markers that have been widely used to assess genetic diversity, population structure, gene flow, and mating systems in many plant species [21]. Several recent studies have successfully developed and applied microsatellite markers for assessing the genetic diversity of various Myrtaceae species [22–26]. For *M. dubia*, only a few microsatellite markers have been published to date [27,28], which may be insufficient for a comprehensive germplasm characterization.

Given the wide geographic distribution of camu-camu across the Amazon basin, we hypothesize that the ex situ germplasm collection assembled from multiple regions in the Peruvian Amazon harbors substantial genetic diversity. However, owing to potential sampling biases, the collection may not be representative of the entire population, with specific genotypes underrepresented or missing. Developing a set of informative microsatellite markers will enable robust quantification of genetic diversity parameters and delineation of a core subset of individuals that maximizes the allelic variability present in the collection.

The main objectives of this study were to 1) develop a set of novel polymorphic microsatellite markers for *M. dubia* suitable for characterizing the germplasm bank, 2) evaluate the genetic diversity and population structure of an ex situ *M. dubia* germplasm collection assembled from multiple locations across the Peruvian Amazon, and 3) construct a core collection that maximizes genetic diversity within this collection to facilitate conservation and breeding efforts. In most cases, limited accessibility to wild populations makes comprehensive field sampling rather difficult; therefore, ex situ collection becomes very important in any genetic diversity study. Accordingly, these recently developed microsatellite markers and the genetically characterized core collection are highly valuable genomic resources that will facilitate the effective management of germplasm and genetic

improvement initiatives for this crop species from the Amazon region that is still underutilized but holds high nutritional and economic potential.

2. Materials and Methods

2.1. Plant Materials

A total of 336 genotypes of *Myrciaria dubia* were obtained from the ex situ germplasm bank maintained by the National Institute of Agrarian Innovation (INIA) in Peru. These genotypes represent 43 accessions originating from eight major river basins (Nanay, Itaya, Napo, Ucayali, Putumayo, Curaray, Tigre, and Amazonas) across the Loreto region of the Peruvian Amazon (Figure 1, Table S1). Young leaf samples were collected from each plant genotype and immediately transported to the laboratory facilities at the Natural Resources Research Center of UNAP (CIRNA) in containers with dry ice to preserve the tissue integrity. Leaf samples were stored at -20 °C until further processing.

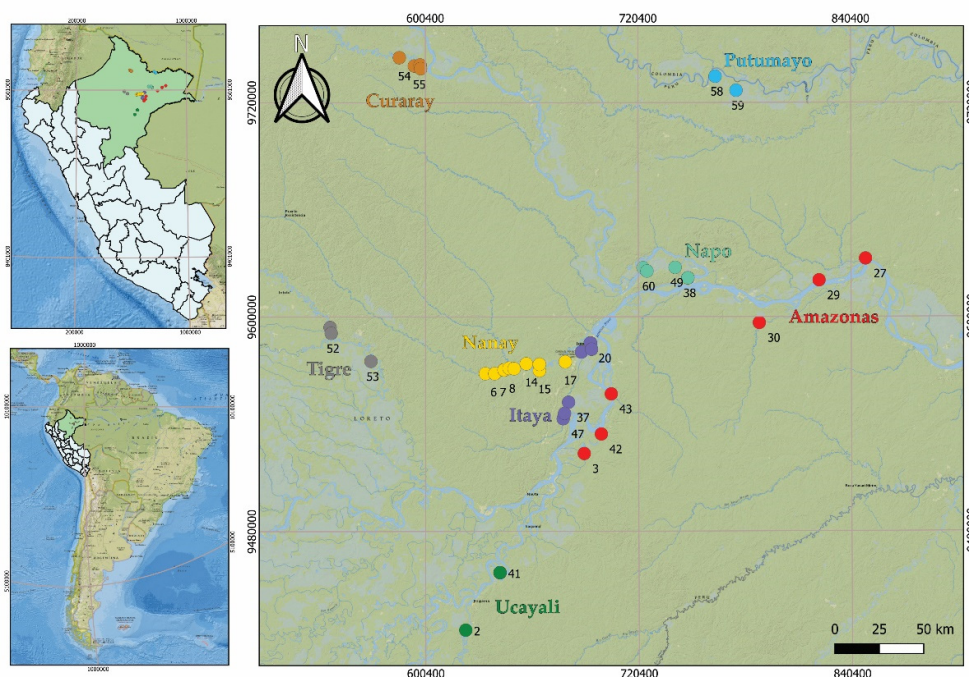


Figure 1. Geographical locations of the eight river basins in the Peruvian Amazon where INIA germplasm bank accessions of *M. dubia* were collected from wild populations.

2.2. DNA Isolation and Analysis

Genomic DNA was isolated from young leaf tissues following a cetyltrimethylammonium bromide (CTAB) extraction protocol with minor modifications [29]. Approximately 100 mg of fresh leaf material was ground to a fine powder in liquid nitrogen using a pre-chilled mortar and pestle. The powdered tissue was then incubated in pre-warmed CTAB extraction buffer supplemented with 2% β -mercaptoethanol at 65 °C for 60 min with intermittent mixing. Cell debris were removed by centrifugation, and the supernatant was extracted with a chloroform:isoamyl alcohol solution (24:1). Nucleic acids were precipitated from the aqueous phase by adding ice-cold isopropanol and centrifugation. The resulting DNA pellet was washed twice with 70% ethanol, air-dried briefly, and resuspended in 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) for long-term storage. The integrity and purity of the extracted DNA were evaluated using a combination of agarose gel electrophoresis and UV-Vis spectrophotometry. Agarose gel electrophoresis was performed by loading an aliquot of each DNA sample onto a 0.8% agarose gel containing 0.5 μ g/mL ethidium bromide [30]. After electrophoresis in 1X TBE buffer, the gels were visualized under UV illumination.

to assess DNA integrity based on the presence of a prominent high-molecular-weight band without significant smearing or degradation. DNA purity was determined by measuring the absorbance ratios at 260/280 and 260/230 nm using a NanoDrop 2000c UV-Vis spectrophotometer (Thermo Scientific). Only DNA samples exhibiting an A260/280 ratio between 1.8–2.0 and an A260/230 ratio greater than 1.8, indicative of high purity with minimal protein/polysaccharide contamination, were considered suitable for downstream applications and selected for microsatellite genotyping.

2.2. Isolation of Microsatellite DNA Loci and Analysis

Microsatellite markers were isolated from the genomic DNA of a single *M. dubia* individual following the enrichment protocol described by Glenn and Schable [31]. This process resulted in the identification of 300 candidate microsatellite loci. After sequencing and PCR assay optimization, a set of 16 polymorphic microsatellite loci was successfully developed and used in this genetic diversity assessment study. Detailed information on these 16 microsatellite markers, including their GenBank accession numbers, primer sequences, annealing temperatures (T_m), repeat motif sequences, size ranges, number of alleles, and polymorphic information content (PIC) values, is provided in Supplementary Table S2.

Plants were genotyped following Schuelke [32] with all forward primers tagged with M13-tails (5'-TGTAACGACGGCCAGT-3') to incorporate fluorescently labeled M13 primers. Reverse primers were developed with a "pigtail" (5'-GTGTCTT-3') on the 5' end to facilitate the addition of an adenine to all PCR products. Microsatellite loci were amplified individually in 10 μ L reactions with the following final concentrations: $\approx$ 25 ng of genomic DNA, 1 U Taq DNA polymerase, 1x PCR buffer (10 mM Tris-HCL, 50 mM KCL), 0.5 mM dNTPs, 2.0–2.5 mM $MgCl_2$, 1x BSA, 0.16 μ M fluorescently labeled M13 primer (VIC, PET, NED, or 6-FAM), 0.04 μ M forward primer, and 0.16 μ M reverse primer. The thermal profile for all PCRs consisted of an initial denaturation step at 94 °C for 4 min, followed by 30 cycles at 94 °C for 15 s, 62 °C for 15 s, 72 °C for 45 s, followed by 8 cycles at 94 °C for 15 s, 53 °C for 15 s, 72 °C for 45 s, and a final extension step for 10 min at 72 °C. PCR reactions (1 μ L) were combined with 8 μ L formamide and 1 μ L of a custom standard [33] and run on an ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, California, USA) at the Pritzker DNA Laboratory for Molecular Systematics and Evolution (Field Museum, USA). Alleles were visualized and called using the microsatellite plugin v1.4.7 in Geneious Prime v2024.0.5 [34].

2.3. Data Analysis

2.3.1. Analysis of the Genetic Diversity and Population Structure

Standard genetic diversity parameters, including the number of alleles (N_a), the effective number of alleles (N_e), Shannon's information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), and Wright's fixation index (F), were calculated using GenAlEx v6.5 [35]. Analysis of molecular variance (AMOVA) was implemented in GenAlEx to partition genetic variability within and among populations [36]. Pairwise genetic differentiation (F_{ST}) and gene flow (N_m) among river basins were estimated using GenAlEx. N_m values among river basins were determined based on the equation [37,38]:

$$N_m = 0.25 \frac{(1 - F_{ST})}{F_{ST}}$$

Clustering based on Bayesian analysis was performed using the number of clusters (K), in the "find.cluster" function of the R package adegenet v1.3.1 [39,40], which does not take Hardy-Weinberg equilibrium into account. The basic algorithm was described by Pritchard, Stephens, and Donnelly [41]. Extensions to the method were published by Falush, Stephens, and Pritchard [42,43], and Hubisz, Falush, Stephens, and Pritchard [44]. Twenty independent runs were performed for each K value ranging from 1 to 34, with a burn-in of 100,000 iterations and 500,000 Markov Chain Monte Carlo (MCMC) repetitions [41–44]. The best value of K is the first break in the curve that corresponds to the lowest value of the Bayesian Information Criterion (BIC) [45].

The number of genetic clusters was further corroborated using discriminant analysis of principal components (DAPC), an effective method for visualizing population structure that enables us to ascertain the most influential factors contributing to variation among the populations under investigation. To accomplish this, it was used the R package adegenet v1.3.1 [39,40], which comprised 200 principal components and a priori grouping of the number of river basins.

To visualize the genetic relationships among populations of river basins, a neighbor-joining tree based on Nei's genetic distances [46,47] was constructed. A distance matrix was first generated using the R package StAMPP v1.6.3 [48,49] and a tree was built with 1000 bootstrap replicates using the R package Poppr v2.9.6 [50,51].

2.3.3. Construction of the Core Collection

To capture the maximum allelic diversity in a representative core collection, CoreHunter3 v3.2.2 [52] was used to select a core subset of 84 genotypes (25% of the full collection) based on a maximization strategy using Modified Rogers distance [53]. Allelic richness was calculated for the full collection and the core subset using the R package hierfstat v0.5-11 [54,55] to evaluate allelic representation in the core collection.

All statistical analyses were performed using R v4.4.0 [56]. Figures were generated (see scripts in <https://github.com/FranciscoAscue/diversity-core-collection-myrciaria-dubia>) using R packages ggplot2 v3.3.6 [57,58], pheatmap v1.0.12 [59], and SRplot [60].

3. Results

3.1. Genetic Diversity Parameters

Genetic diversity analysis of the *M. dubia* germplasm bank, consisting of 43 accessions obtained from eight river basins (Amazonas, Curaray, Itaya, Nanay, Napo, Putumayo, Tigre, and Ucayali) in the Peruvian Amazon, revealed substantial variation in genetic parameters, indicating a rich and complex genetic structure within this species. In total, 336 genotypes were examined using 16 polymorphic microsatellite loci. All microsatellite loci were polymorphic, with the number of alleles per locus ranging from 10 to 28, resulting in 313 alleles. The polymorphic information content ranged from 0.66 to 0.94 (Figure 2, Table S2).

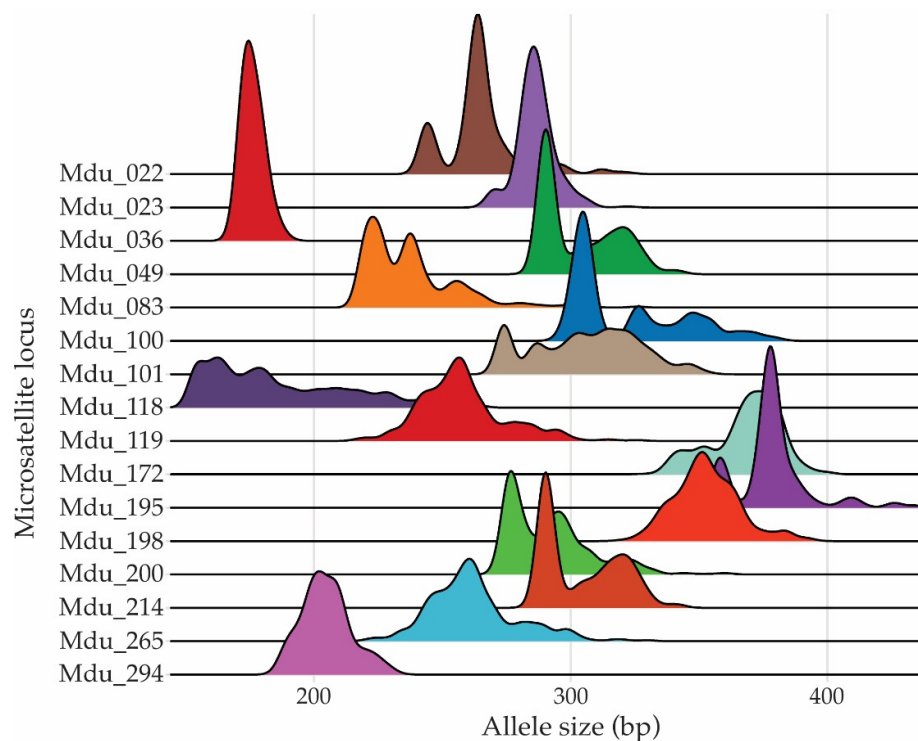


Figure 2. Allele size distribution of sixteen polymorphic microsatellite loci used for assessing genetic diversity and constructing the core collection of the *M. dubia* germplasm bank from the Peruvian Amazon.

At the accessions level (Table S3), the number of alleles per locus (*Na*) varied from 1.875 to 6.688, with an average of 5.438 ± 1.140 , while the effective number of alleles (*Ne*) ranged from 1.395 to 4.721, with an average of 3.899 ± 0.813 . The Shannon information index (*I*), which measures allelic richness and evenness, ranged from 0.397 to 1.623, with an average of 1.381 ± 0.287 . The observed heterozygosity (*Ho*) ranged from 0.104 to 0.854, with an average of 0.604 ± 0.152 , whereas the expected heterozygosity (*He*) varied from 0.236 to 0.760, with an average of 0.675 ± 0.119 . The degree of inbreeding, as measured by the fixation index (*F_{IS}*), ranged from -0.259 to 0.511, with an average of 0.103 ± 0.177 .

At the river basin level (Table 1), the average number of alleles (*Na*) per locus was 10.180 ± 1.849 , with Nanay containing the highest (12.438) and Putumayo the lowest (6.813) number of alleles. The effective number of alleles (*Ne*) ranged from 3.834 (Tigre) to 6.303 (Napo), with an average of 5.395 ± 0.947 . Shannon’s diversity index averaged 1.800 ± 0.203 . The highest Shannon’s diversity value was observed in Napo (2.006), whereas the lowest was noted in Putumayo (1.489). Observed heterozygosity (*Ho*) showed an average of 0.551 ± 0.06 ; Amazonas presented the highest value, with 0.644, while Nanay had the lowest, with 0.468. The expected heterozygosity (*He*) was in general higher, averaging 0.757 ± 0.059 , with Napo recording the maximum value of 0.817 and Tigre the minimum of 0.672. The fixation index (*F_{IS}*) averaged 0.276 ± 0.077 , with values ranging from 0.160 in Putumayo to 0.371 in Itaya, showing a trend toward deficiency of heterozygotes in all populations. These results taken together show considerable genetic diversity within the *M. dubia* germplasm bank in the Peruvian Amazon, evidencing remarkable variation among river basins.

Table 1. Assessment of genetic diversity parameters at the river basin level for *M. dubia* germplasm bank from the Peruvian Amazon using sixteen microsatellite markers.

River Basin	N	Genetic Diversity Parameters					
		<i>Na</i>	<i>Ne</i>	<i>I</i>	<i>Ho</i>	<i>He</i>	<i>F_{IS}</i>
Amazonas	50	11.375	6.154	1.939	0.644	0.797	0.191
Curaray	32	11.063	6.294	1.992	0.556	0.809	0.313
Itaya	53	10.625	5.847	1.910	0.507	0.803	0.371
Nanay	91	12.438	5.247	1.779	0.468	0.717	0.356
Napo	37	11.313	6.303	2.006	0.638	0.817	0.225
Putumayo	19	6.813	4.219	1.489	0.580	0.684	0.160
Tigre	24	8.375	3.834	1.520	0.473	0.672	0.306
Ucayali	30	9.438	5.262	1.765	0.546	0.754	0.284
Average/Total	336	10.180	5.395	1.800	0.551	0.757	0.276
Standard deviation	-	1.849	0.947	0.203	0.068	0.059	0.077

Note: *N*, number of genotyped individuals; *Na*, number of alleles; *Ne*, number of effective alleles; *I*, Shannon’s diversity index; *Ho*, observed heterozygosity; *He*, expected heterozygosity; *F_{IS}*, fixation index.

An evaluation of the exclusive genetic characteristics across the eight river basins revealed substantial differences in the prevalence of unique genetic variations within the germplasm bank of *M. dubia*. A total of 237 private alleles were identified across the river basins (Figure S2). Although it had the fewest number of individuals (19), the content of Putumayo River basin was very high, comprising 91 private alleles. Not much behind was the Tigre River basin, with 70, which indicated a significant pool of diversity. The third position was occupied by Nanay River basin with 31 private alleles, followed by Curaray and Amazonas River basins with 17 and 15 unique alleles, respectively.

In sharp contrast, only 13 private alleles were found in the Itaya, Napo, and Ucayali River basins, indicating a general lack of regional genetic uniqueness among those populations.

3.2. Analysis of Molecular Variance (AMOVA)

Analysis of Molecular Variance (AMOVA) showed very useful information with respect to the distribution of genetic variation within the *M. dubia* germplasm bank (Table 2). At the accessions level, it was observed that there was a greater degree of genetic variation within accessions than among accessions. In particular, 73% of the total genetic variance was attributed to within-accession variation, while 27% was attributed to among-accession variation. This indicates that individual accessions of *M. dubia* remain high in levels of internal genetic variation and this can be useful for conservation and breeding purposes. This was further justified by the fact that the estimated variance of components within-accession was high (13.821) as opposed to the estimated variance of components among-accession variation (5.200).

Table 2. Genetic variation among and within accessions and river basins of *M. dubia* germplasm bank from the Peruvian Amazon based on sixteen microsatellite markers and determined with analysis of molecular variance (AMOVA).

Source	df	SS	MS	Est. Var.	%
At accessions level					
Among accessions	42	2284.679	54.397	5.200	27
Within accessions	293	4049.461	13.821	13.821	73
Total	335	6334.140		19.020	100
At river basins level					
Among river basins	7	904.018	129.145	2.785	14
Within river basins	328	5430.121	16.555	16.555	86
Total	335	6334.140		19.341	100

Abbreviations: *df*, degrees of freedom; *SS*, sum of squares; *MS*, mean of the squares; *Est. Var.*, estimated variance of components; %, percentage of total variance contributed by each component.

Upon examining the genetic variance within the river basin, a more pronounced pattern was detected. Out of the total genetic variation, 86% was found within river basins, while just 14% occurred among different river basins. These findings suggest that *M. dubia* populations from various river basins in the Peruvian Amazon have significant genetic similarities. The estimated variation of components within river basins was much greater at 16.555 compared to the variation among river basins, which was 2.785. This observation provides evidence for the hypothesis that local populations maintain a high level of genetic variation.

3.3. Analysis of Genetic Differentiation and Gene Flow among River Basins

Figure 3 shows the results for genetic differentiation (*F_{ST}*) among the plants of the *M. dubia* germplasm bank derived from eight different river basins: Amazonas, Curaray, Itaya, Nanay, Napo, Putumayo, Tigre, and Ucayali. The pairwise *F_{ST}* values ranged from 0.018 to 0.166, thus showing different degrees of genetic differentiation among the river basins. The highest *F_{ST}* value was observed between Putumayo and Tigre, with an *F_{ST}* = 0.166. This would indicate very strong genetic differentiation between the two populations. Other high *F_{ST}* values were observed between Putumayo and Nanay, with an *F_{ST}* = 0.154; Putumayo and Ucayali, with an *F_{ST}* = 0.141; and Putumayo and Amazonas, with an *F_{ST}* = 0.130. In contrast, relatively low *F_{ST}* values were observed between Amazonas and the other river basins, ranging from 0.023 to 0.046. Likewise, for the Ucayali River basin, there were low *F_{ST}* values with Nanay, of 0.018; Amazonas, of 0.023; Itaya, of 0.037; and Napo, of 0.042; hence, indicating closer genetic similarity among these populations. The overall *F_{ST}* values showed high genetic differentiation among the river basins; Putumayo and Tigre represent the largest differentiation in relation to other river basins.

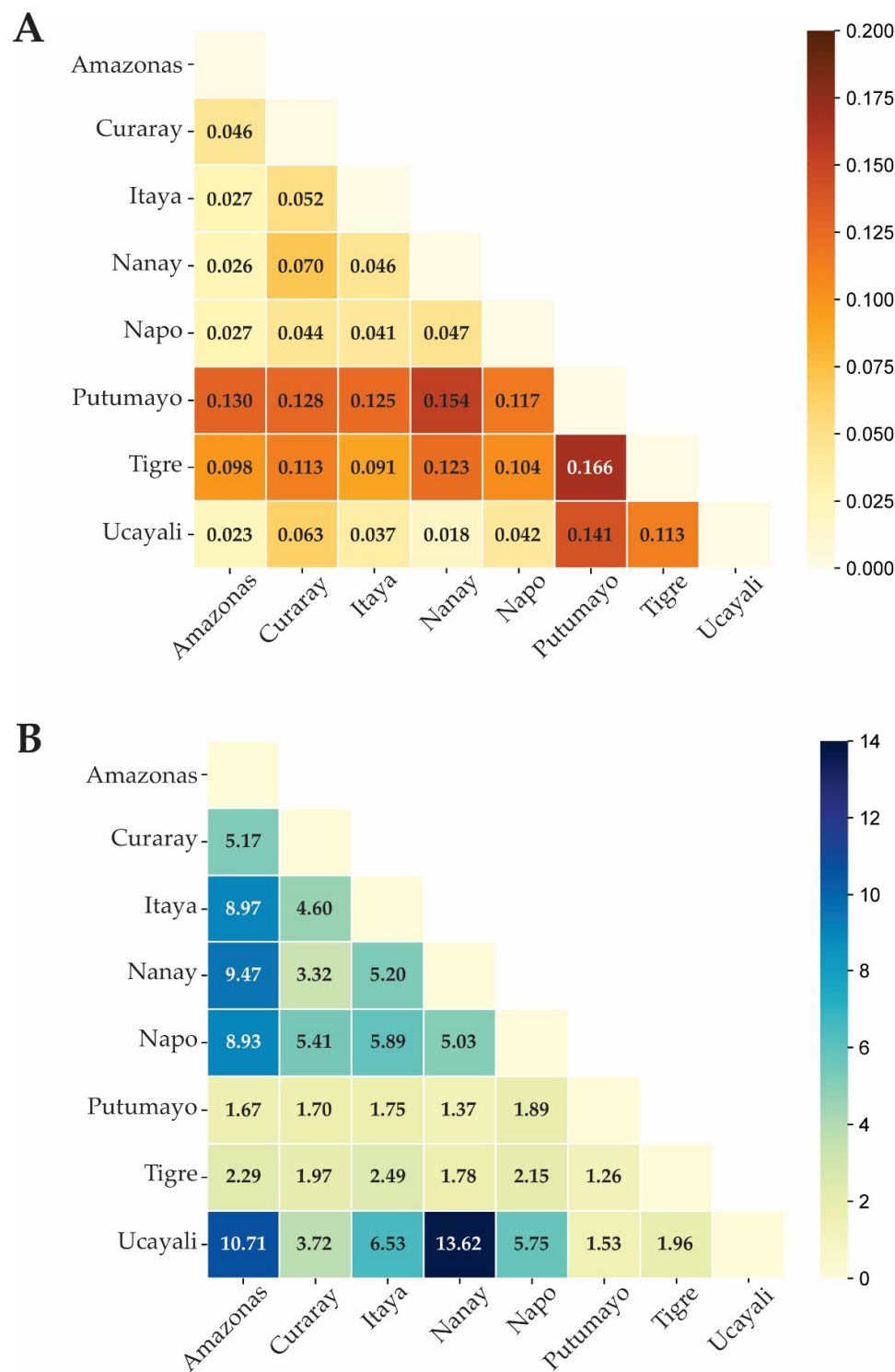


Figure 3. Pairwise comparisons of genetic differentiation and gene flow among *M. dubia* plants from the germplasm bank derived from eight river basins. A) Heatmap of pairwise F_{ST} values representing genetic differentiation among river basins. Higher F_{ST} values indicate greater genetic differentiation among populations. B) Heatmap of pairwise N_m values representing gene flow (number of migrants per generation) among *M. dubia* plants from the germplasm bank derived from eight river basins, with higher N_m values indicating greater gene flow among populations.

The gene flow, measured as the number of migrants per generation (Nm), between individuals of the same river basins was estimated (Figure 3). The higher the value of Nm , the more intense the gene flow between populations, opposing genetic differentiation. The values of Nm ranged from 1.26 to 13.62 in this study, indicating a moderate gene flow among river basins. Itaya and Ucayali had the highest gene flow, with $Nm = 13.62$, indicating that there had been a high degree of genetic exchange between these two populations. Other high values of Nm were between Amazonas and Ucayali, with an Nm of 10.71; between Nanay and Amazonas, with an Nm of 9.47; and between Itaya and Amazonas, with an Nm of 8.97. In contrast, the river basins of Putumayo against all others presented low values of Nm . The lowest value was noticed between Putumayo and Tigre, where $Nm = 1.26$, pointing to a very low rate of gene flow between these populations. In the same way, low Nm values were recorded between Putumayo and Nanay, Putumayo and Ucayali, and Putumayo and Amazonas, with values of 1.37, 1.53, and 1.67, respectively. In general, the Nm values indicated that, in fact, there is fairly high gene flow between several river basins, mostly between Amazonas, Itaya, Nanay, and Ucayali, responsible for keeping genetic connectivity. However, low gene flow detected among Putumayo and with other river basins, particularly Tigre, indicated restricted genetic exchange that adds up to the genetic differentiation observed.

When the relationship between genetic differentiation (F_{ST}) and gene flow (Nm) among the river basins was analyzed, an inverse relationship was observed, indicating that as gene flow increased, genetic differentiation decreased (Figure S1). The equation describing this relationship had a very high coefficient of determination ($R^2 = 0.9995$), suggesting an excellent fit of the model to the data. This suggests that the transfer of genes between various plant genotypes is crucial in minimizing the genetic variation among the river basins of *M. dubia* accessions from the Peruvian Amazon.

3.4. Principal Component Discriminant Analysis (DAPC)

Discriminant Analysis of Principal Components (DAPC) was conducted based on eight river basins, with 200 principal components representing 98% of the total variation. The scatterplots of DAPC are shown in Figure 4 and delimit genetic relationships among plant genotypes from different river basins.

The results indicated that the plant genotypes from the Putumayo, Curaray, and Tigre River basins could be separated from the rest to fall into two distinct main groups. Figure 3A clearly allows the separation to be seen in this scatter plot, underlining the different genetic profiles of these river basins. Even if the other remaining plant genotypes from the other river basins do not have complete differentiation, this already hinted at substantial genetic connectivity among the other river basins, particularly the Itaya, Nanay, Amazonas, Napo, and Ucayali.

To attain a greater level of resolution among plant genotypes belonging to the remaining river basins, we conducted an additional DAPC with the same parameters, excluding individuals from the Putumayo River basin (Figure 4B). This analysis revealed more separation of the Tigre, Napo, and Curaray River basins. Although the Itaya, Nanay, Amazonas, and Ucayali River basins do not fully differentiate from one another, they do exhibit some distinctiveness, suggesting subtle genetic distinctions despite the overall genetic connectivity.

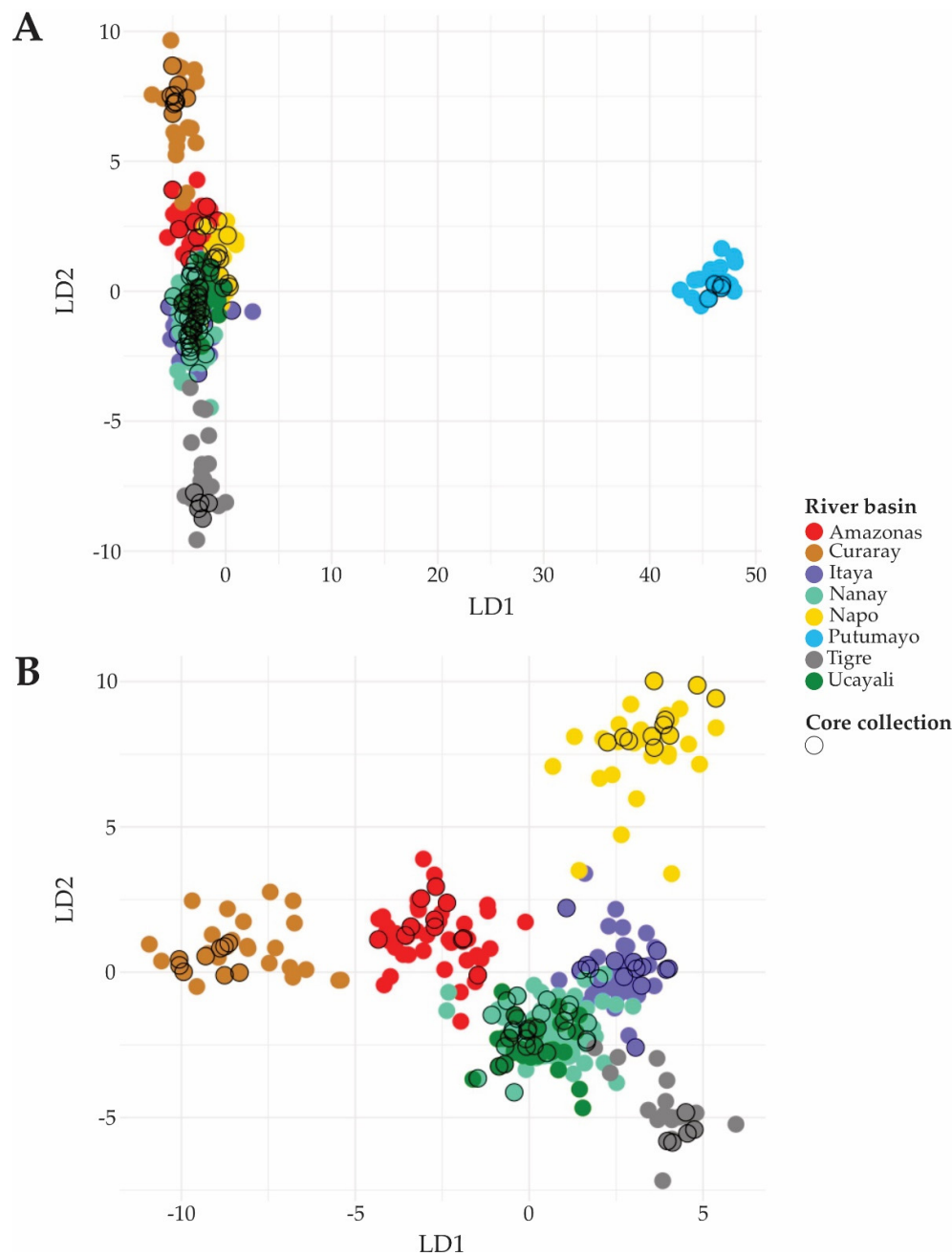


Figure 4. Discriminant analysis of principal components (DAPC) for plant genotypes of *M. dubia* from eight river basins of the germplasm bank, based on sixteen microsatellite loci. A) DAPC scatterplot showing genetic differentiation among all eight river basins, highlighting the distinct separation of Putumayo, Curaray, and Tigre River basins. B) DAPC scatterplot excluding the Putumayo River basin genotypes, revealing finer genetic structure among the remaining seven river basins.

3.5. Bayesian Cluster Analysis

Bayesian cluster analysis revealed unique genetic structures and admixture patterns of the *M. dubia* accessions across the river basins (Figure 5). There was evidence that $K = 10$ was the optimal number of genetic clusters since at this point, the Bayesian Information Criterion (BIC) had the lowest value; this is indicated by the red dot in Figure 5A. From $K = 1$ to $K = 10$, the BIC values decreased sharply before increasing again for higher K values. The pattern suggests that, up to ten clusters,

model fit improves appreciably, after which additional clusters provide diminishing returns or even reduce model explanatory power.

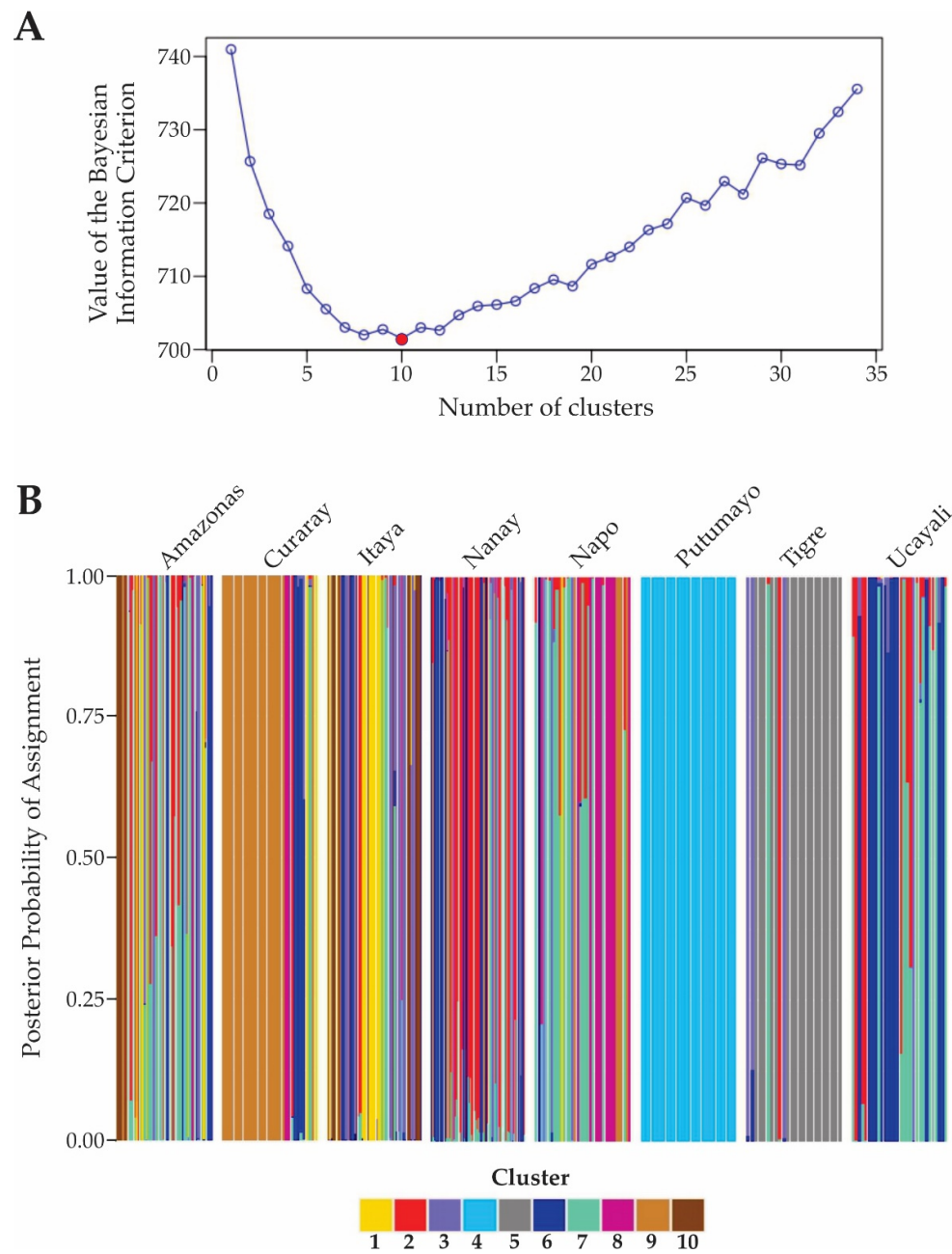


Figure 5. Genetic structure analysis of the *M. dubia* germplasm bank at the river basin level. A) A graph showing the optimal number of genetic clusters based on the Bayesian Information Criterion, B) A bar plot representing the probability of assignment of individuals to different genetic clusters across the eight river basins.

Figure 5B depicts the results in bar plots of the posterior probability of the assignment for each river basin of the individual plant genotypes to the ten genetic clusters identified. Each vertical bar thus represents an individual plant genotype from the respective river basin, while colors indicate the proportion of their genome from the different genetic clusters.

The results exhibited a complex genetic structure, with different degrees of admixture of the analyzed *M. dubia* germplasm accessions from eight river basins. Germplasm accessions from the

Amazonas, Itaya, Nanay, Napo, and Ucayali River basins showed some degree of admixture; that is, plant genotypes were assigned to multiple genetic clusters. Those germplasm accessions derived from the Tigre and Curaray River basins displayed moderate grades of admixture, mostly being assigned to clusters 5 (grey) and 9 (light brown), respectively. In sharp contrast to this, germplasm accessions from the Putumayo River basin did not show any admixture, in which all of the plant genotypes are classified in the genetic cluster 5 (light blue). These findings give evidence regarding the complex genetic structure and variable levels of admixture present in the eight river basins and their corresponding germplasm accessions of *M. dubia*.

3.6. Phylogenetic Relationship among River Basins

Based on Nei's genetic distances, the neighbor-joining tree provided insights into the genetic relationships and divergence among germplasm accessions from the eight river basins in the study area, as depicted in Figure 6. Tree topology revealed two main genetic clusters that would suggest the existence of distinct evolutionary lineages. One large cluster comprised the Curaray, Napo, Amazonas, Ucayali, Nanay, and Itaya River basins, thus indicating their proximity genetically and therefore higher historical gene flow rates among these river basins. Their close genetic relationships among these six river basins agree with the high admixture levels observed in these populations. Larger genetic distances of the second major cluster, including the Putumayo and Tigre River basins from all others, suggest their independent genetic background and reduced admixture. The results underline the complex genetic structure and heterogeneous genetic connectivity among plant genotypes of *M. dubia* in the different river basins.

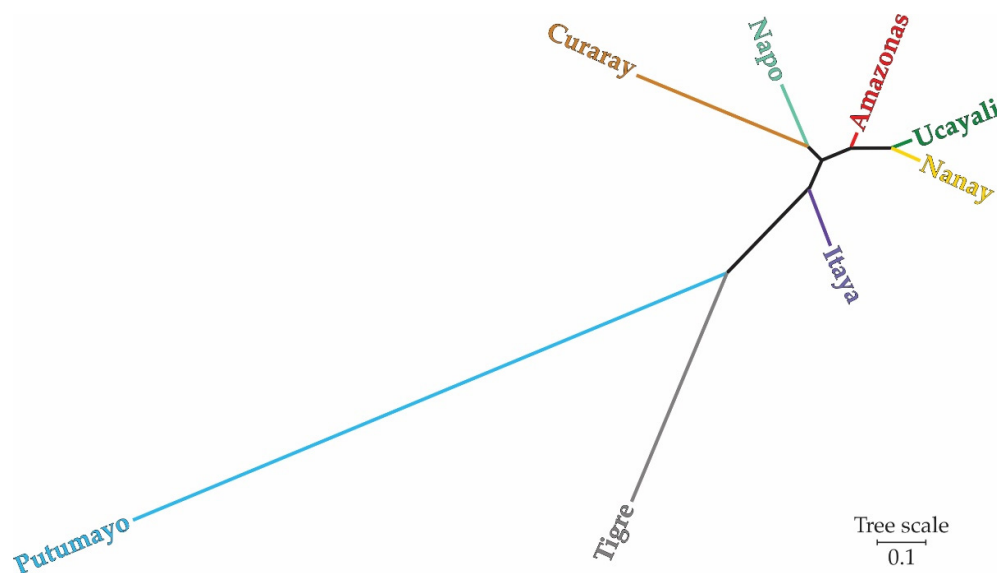


Figure 6. Neighbor-Joining phylogenetic tree depicting the genetic relationships among *M. dubia* germplasm accessions from eight river basins. The tree was constructed based on Nei's genetic distances and was generated with 1000 bootstrap replicates.

3.7. Construction of the Core Collection

The core collection was constructed using a thorough and precise approach through the use of multipurpose core subset selection tools. These tools utilized local search algorithms when generating subsets based on one or more distance and allelic richness metrics derived from the information given by the sixteen polymorphic microsatellite loci. This strategy ensured that the selected core collection would represent the entire germplasm bank and would be optimum in genetic diversity and allelic richness for the conservation of general genetic structure of *M. dubia*. In the end, a total of 84 representative plant genotypes were selected to fully cover genetic variation. The core collection was made up of genotypes from the eight river basins: Amazonas (13.10%),

Curaray (10.71%), Itaya (19.05%), Nanay (25.00%), Napo (13.10%), Putumayo (4.76%), Tigre (5.95%), and Ucayali (8.33%). This distribution mirrors the proportional representation of each river basin in the core collection (Figure 4, Table S4).

The allele coverage in the core collection was high at 0.867, which in itself is indicative of substantial allelic richness. This allelic richness for the whole sample set and just the core collection is depicted graphically by the heatmap in Figure 7. A high value for allele coverage showed that this core collection had managed to pick up most of the genetic diversity in the larger germplasm bank.

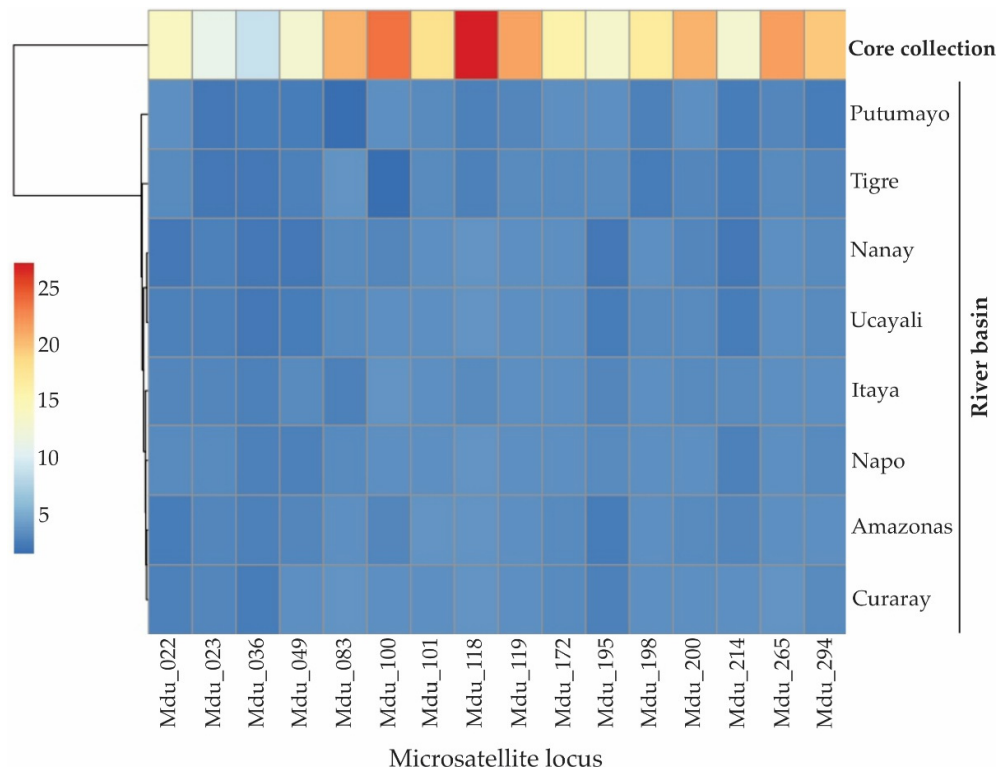


Figure 7. Heatmap depicting the allelic richness of a core collection from the *M. dubia* germplasm bank in the Peruvian Amazon based on sixteen microsatellite loci.

4. Discussion

4.1. Genetic Diversity Parameters

The estimates of genetic diversity parameters among river basins were highly variable, thus underlining the necessity for germplasm sampling and conservation efforts that cover the entire Amazonian region. Such moderate to high levels of genetic diversity agreed with the allelic richness and heterozygosity measures seen for *M. dubia*; this is consistent with its wide geographic distribution and outcrossing nature.

Despite the high genetic diversity, positive fixation indices and heterozygote deficiencies detected in several river basins of the *M. dubia* germplasm bank may indicate the presence of inbreeding within the populations. Such a trend was observed in a study investigating the genetic diversity of an ex situ germplasm collection of 139 *M. dubia* accessions from 17 distinct populations, originating from at least six tributary rivers to the Amazon River in the Brazilian Amazon region, and preserved at the INPA (Instituto Nacional de Pesquisas da Amazônia) Active Germplasm Bank in Manaus. For this purpose, the researchers used eight expressed sequence tag-derived microsatellite markers, which revealed a deficit of heterozygotes and high values of fixation index (F_{IS}), with H_o ranging from 0.266 to 0.539 versus H_e ranging from 0.691 to 0.903, and the F_{IS} ranging from 0.107 to 0.454 [28].

Another study evaluated 126 samples from 13 populations representing 10 wild and three cultivated population types of *M. dubia* from five river basins, namely Curaray, Itaya, Napo, Putumayo, and Tigre, within the Peruvian Amazon region, using seven polymorphic EST-SSR markers. The results mirrored those from our results and the Brazilian Amazon, showing a deficiency in heterozygotes (H_o ranging from 0.137 to 0.527 vs. H_e ranging from 0.218 to 0.680) and high F_{is} values ranging from 0.137 to 0.527 [27]. In a nutshell, these results, when combined, mean that heterozygote deficiencies and positive fixation indices within *M. dubia* germplasm collections and within wild and cultivated population types reflect inbreeding. It could be attributed to genetic drift, restricted gene flow, or population fragmentation—all factors underlying local adaptation, similarly to what has been demonstrated for *Eugenia dysenterica* [61]. Such trends call for focused strategies of conservation that would help retain genetic diversity and reduce inbreeding in *M. dubia* populations within the Amazonian region.

A high number of private alleles, notably in accessions from the Putumayo, Tigre, and Nanay River basins, reveals unique genetic variations within those populations. These basin-specific alleles add to the total genetic diversity of the species, thus underscoring the importance of conservatism in the germplasm bank from diverse geographic origins, which captures maximum allelic richness. The result also underlines the role of these particular river basins as potential reservoirs of distinct genetic diversity for *M. dubia*. Nevertheless, such peculiar genetic resources do need targeted conservation efforts for the maintenance of the adaptive potential of the species, as has been shown by studies on other Amazonian fruit trees like *Theobroma grandiflorum* [62]

4.2. Analysis of Molecular Variance (AMOVA)

The results from the analysis of molecular variance conducted using the *M. dubia* germplasm bank revealed a significant pattern of genetic diversity that had implications for conservation and breeding programs. The fact that there was higher genetic variation within accessions (73%) than among accessions (27%) means that there was reasonable maintenance of genetic diversity within the local populations. Additionally, the highest genetic variation was within river basins compared to among river basins, going for as high as 86% versus 14%. This means that *M. dubia* populations from the several river systems of the Peruvian Amazon have much of their genetic diversity in common (Table 2). The result coincides with studies of genetic grouping in the same species from the Brazilian Amazon region conducted by Rojas et al. [28], particularly within the active germplasm bank of INPA. Also, this pattern is in conformity with other findings on tropical tree species, such as *Bertholletia excelsa* [63], *Theobroma cacao* [64], and *Euterpe edulis* [65], for which high within-population diversity has been attributed to outcrossing mating systems and efficient gene flow mechanisms. Long-distance gene flow, possibly mediated by seed dispersal via water or animal vectors, may be involved in the maintenance of high within-population diversity. This has long been reported for other Myrtaceae species [66].

Genetic variation of *M. dubia* therefore holds some important implications for the design of appropriate conservation strategies. The high within-accession and within-river basin diversity suggests that collecting strategies focused on sampling multiple individuals within fewer locations may efficiently capture a large proportion of the species' genetic diversity. In fact, this strategy is recommended for other plant species of comparable genetic structure [11,15,67]. However, still a non-negligible among-accession (27%) and among-river basin variation (14%) exists, indicating that sampling from several accessions and river basins is necessary to capture all the existing genetic diversity. The results emphasize the need for broad geographic representation in germplasm collections, as suggested by studies on other Amazonian fruit species [68].

This genetic structure could be explained by a combination of historical and contemporary processes of population dynamics, gene flow, and local adaptation. Further studies on these factors, probably using landscape genetics methods [69], may provide insight into the evolutionary history and contemporary dynamics of *M. dubia* populations in the Peruvian Amazon.

Finally, several implications for breeding programs are derived from these results. High genetic variation in germplasm banks has been found to be essential for breeding strategies that maximize

genetic diversity in many studies across different plant species, including *Arabidopsis thaliana* [70], *Melia dubia* [71], and other plant species [72]. This probably induces heterosis effects [73]; therefore, all these studies underline the importance of genetic diversity within and among populations. Increased genetic variability through the use of diverse germplasm offers better traits for hybridization with maximum heterosis effects. It is through the exploitation of the existing genetic diversity by incorporating genetic material from different populations that breeders can come up with superior plant varieties having the desired traits, which is evidenced from various promising tropical fruit tree breeding programs in Latin America [74–76].

4.3. Analysis of Genetic Differentiation and Gene Flow among River Basins

The pairwise F_{ST} values ranging from 0.018 to 0.166 indicate varying degrees of genetic differentiation among the river basins. The highest differentiation observed between Putumayo and other basins suggests that this population may have been relatively isolated or subjected to different selective pressures. Similar patterns of population differentiation have been observed in other Amazonian tree species, such as *Swietenia macrophylla*, where genetic differentiation values were positively and significantly correlated with geographical distance under the isolation-by-distance model [77]. According to Gamba and Muchhala higher F_{ST} values are reported for tropical plants that are non-woody species, have mixed-mating systems, and are pollinated by small insects. Also, among the ecological factors tested, the latitudinal region explained the largest portion of variance, followed by pollination mode, mating system, and growth form, while seed dispersal mode did not significantly relate to genetic differentiation [78].

The patterns of gene flow among the river basins revealed a complex landscape of genetic connectivity and isolation. The high levels of gene flow observed between the Ucayali with Nanay and with Amazonas River basins and between Amazonas with Napo, Nanay and Itaya River basins (Figure 3) suggest extensive historical gene exchange, potentially facilitated by interconnected river networks, animal or human-mediated dispersal and other ecological factors [78]. Conversely, the low to moderate gene flow levels between certain basin pairs, such as Putumayo-Tigre and Putumayo-Nanay, may be attributable to geographic barriers or insufficient dispersal mechanisms, leading to greater genetic isolation.

The inverse relationship between genetic differentiation (F_{ST}) and gene flow (Nm) is a well-established pattern in plant population genetics, indicating that as gene flow increases, genetic differentiation between populations decreases [79–81]. In the case of *M. dubia*, the high coefficient of determination ($R^2 = 0.9995$) in this relationship underscores the critical role of gene flow in shaping the genetic structure of these populations. This strong correlation suggests that gene flow acts as a homogenizing force, reducing genetic differentiation by enabling the exchange of alleles across populations. Consequently, maintaining landscape connectivity becomes paramount for the long-term genetic health and viability of *M. dubia*. Connectivity facilitates gene flow, thereby preserving genetic diversity, enhancing adaptive potential, and reducing the risks associated with inbreeding and genetic drift. This finding aligns with broader conservation principles, emphasizing that fragmented habitats and disrupted dispersal pathways can lead to isolated populations with diminished genetic variability and resilience. Thus, maintaining stable, genetically varied populations of *M. dubia* and other related species requires maintaining habitats and ecological pathways.

4.4. Principal Component Discriminant Analysis (DAPC)

Genetic structure complexity was also reflected by the results of DAPC for the *M. dubia* germplasm bank. Actually, the main clusters from Putumayo, Curaray, and Tigre River basins appear clearly separated from the remaining populations, which shows that they have particular genetic profiles probably due to isolation or adaptation to special environmental conditions. Similar genetic distinctiveness was detected in geographically isolated populations of other Amazonian tree species, including *Bertholletia excelsa* [82].

The weak structuring among the other basins (Itaya, Nanay, Amazonas, Napo, and Ucayali) is a signal of relatively good genetic connectivity among these populations. The observed pattern of admixture for this species suggests that the existence of these mechanisms of gene flow (such as wind, water dispersion, or animal-mediated pollination and seed dispersal) would be effective in the exchange of genetic material over long distances. This forms a critical aspect for maintaining genetic diversity, enabling maximum adaptive potential in populations facing environmental changes. Patterns of genetic admixture similar to these have been described in other broadly distributed Amazonian trees, such as *Ceiba pentandra*, where high levels of gene flow have been attributed to their broad distribution and effective pollination mechanisms [83]. Understanding these dynamics is critical to conservation efforts because they underline the importance of conserving natural habitats and processes that permit gene flow, hence ensuring long-term viability of genetic diversity in such populations.

4.5. Bayesian Cluster Analysis

The Bayesian cluster analysis that returned ten genetic clusters ($K = 10$) underlines the complex genetic structure of *M. dubia* within the studied river basins. High levels of admixture, as observed in most populations, particularly in Amazonas, Itaya, Nanay, Napo, and Ucayali, suggest historic and contemporary gene flow among these river basins. This pattern of admixture is similar to that found in other broadly distributed Amazonian tree species, such as *Jacaranda copaia*, in which gene flow maintains genetic connectivity over large areas [84]. Also, Zhang et al. [64], using the Bayesian clustering method, revealed appreciable genetic structure in a collection of *Theobroma cacao* due to the river systems in the Peruvian Amazon.

That genetic homogeneity, observed in this particular case in the Putumayo River basin, is striking and special in its potential underlying mechanisms. This homogeneity would represent a founder effect where a new population was established by a few individuals who had limited genetic variation. It could also be due to a recent population bottleneck—some environmental event or human activity that reduced population size and lost genetic diversity. Additionally, strong adaptation to local conditions in this region of Putumayo is what can explain genetic uniformity. Such distinct genetic clusters have been described for other Amazonian species, and these have been noted as being relevant as conservation units. For example, studies on *Theobroma cacao* have proved that these genetically distinct populations can possess very useful traits useful for conservation and breeding purposes [85]. Preservation of such genetic clusters relies on their understanding and preservation to maintain the integrity of the general genetic diversity and resilience of species in the Amazon basin.

4.6. Phylogenetic Relationship among River Basins

The neighbor-joining tree, based on Nei's genetic distances, displays two principal genetic clusters that give clues about the evolutionary history of *M. dubia* from the Peruvian Amazon. Closely related genes among the Curaray, Napo, Amazonas, Ucayali, Nanay, and Itaya River basins suggest a common evolutionary history that may have been driven by historical and ongoing gene flow. This result may indicate similar selective pressures and continuing gene flow between these populations, which contributes to the retention of overall genetic diversity and, therefore, the adaptability of these river basins.

In sharp contrast, the clustering of the Putumayo and Tigre River basins suggests an independent evolutionary history. This could have been the consequence of geographic barriers, environmental gradients, or historic isolation, reducing gene flow, and hence enhancing genetic differentiation. These results suggest that the Putumayo and Tigre River basins are rich in privately held alleles and adaptations; hence, they become very important targets for conservation. The results of such phylogenetic analyses will be important in understanding the evolutionary history of and the genetic structure of *M. dubia* for the formulation of effective conservation strategies. This ensures that both the genetic diversity within the interconnected populations and the unique genetic resources found in isolated populations are conserved.

The pattern of genetic relationships among populations from different river basins that was obtained for the eight river basins corresponds very much to other phylogeographic studies dealing with the genetic structure of other Amazonian tree species, such as *Cedrela fissilis*, in which the river systems clearly show a strong imprint on their current genetic structure [86]. An interplay of historical processes—including past climatic changes and geological events—and more recent anthropogenic influences is likely responsible for the genetic structure observed in *M. dubia*.

4.7. Construction of the Core Collection

This is an important dimension for the conservation and utilization of *M. dubia* genetic resources: a core collection can be built, representing 86.7% of the allelic diversity present in the entire germplasm bank. At such a high level of allele capturing, the core collection represents the genetic diversity of the species, whereas the number of accessions to be maintained is greatly reduced. It offers an effective method of management and conservation and helps to ensure that important genetic variations are conserved with minimal input.

The methodologies that had been adapted for the construction of core collection using multipurpose core subset selection tools and local search algorithms had prior antecedents already applied with success in other crop species such as *Cunninghamia lanceolata* [87], *Ginkgo biloba* [88], *Larix decidua* [89], *Phoebe zhennan* [90], and *Pinus koraiensis* [91]. Such strategies maximize the capture of genetic diversity as part of the core collection, enhancing its representativeness and utility. Albeit the genotypes were distributed proportional to the source river basin origin, the *M. dubia* core collection was, therefore, suitable for further breeding programs and genetic studies—allowing targeted research and development efforts to work on the discovery of beneficial traits and improvement of cultivars.

The establishment of such a core collection is therefore of great significance for an underutilized species like *M. dubia*, as it offers an avenue for efficient characterization, evaluation, and utilization of genetic resources. Core collections have contributed to the genetic enhancement of tropical crops including cassava, strawberry, and cacao. Optimization procedures based on genetic distances and phenotypic and genotypic data have kept significant genetic diversity in cassava core collections and as such, useful for breeding programs and genomic investigations [92]. In strawberries, the integration of genomic and pedigree information in a core collection captured the maximum genetic variation in a small subset of genotypes for its future-proofing and facilitating haplotype reference panel development to be used for genotyping [93]. Finally, in cacao, the construction of the core collection was performed using 15 microsatellite loci by researchers who performed it with five different sampling algorithms [94]. As these data showed, the role of core collections is actually important in the genetic improvement of tropical plants. The well-curated and diversified genetic foundation of the *M. dubia* core collection will help for the selection of superior genotypes useful for cultivation and study, in species conservation and development.

5. Conclusions

This constitutes the first comprehensive assessment of genetic diversity and core collection development with regard to *M. dubia* germplasm from the Peruvian Amazon. The new microsatellite markers developed herein enabled the quantification of the genetic parameters, showing high diversity among the accessions and river basins. This study demonstrated the need for conserving diversity at several scales due to the significant genetic differentiation among accessions and river basins.

The complicated patterns of gene flow and genetic structure among river basins reflect the geographic, dispersal, and evolutionary processes interplay in the formation of genetic diversity of *M. dubia* in the Amazonian landscape. Basins with a higher level of admixture and connectivity, like Amazonas, Curaray, Nanay, and Ucayali, most likely underwent historical gene exchange events that should be driven by the interconnection of river systems, animal or human-mediated dispersals. Other basins, such as Putumayo and Tigre, would appear to have been relatively more genetically isolated, perhaps because of geographical barriers or reduced effective levels of dispersal.

This will be useful in developing a core collection that is rationally selected, with maximum allelic richness and, therefore, representation of genetic diversity. Such a core collection is a grained resource in germplasm management, conservation, and breeding. In such a constructed core collection, the robust coverage of alleles and proportional basin representation will ensure that the total genetic diversity present in this larger germplasm bank is preserved.

These findings validate expected levels of diversity and structure of populations, therefore securing genetically diverse germplasm accessions and the role of effective genetic characterization for ex situ germplasm management. These newly developed microsatellite markers and the genetically characterized core collection comprise the available genomic resources that will empower future research and conservation in this underutilized Amazonian crop species.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Relationship between genetic differentiation (F_{ST}) and gene flow (Nm) among plant genotypes belonging to the eight river basins in the *M. dubia* germplasm bank from the Peruvian Amazon; Figure S2: Distribution of private alleles across plant genotypes belonging to the eight river basins in the *M. dubia* germplasm bank from the Peruvian Amazon. Table S1: Comprehensive dataset of *M. dubia* germplasm bank from the Peruvian Amazon, including codes, river basin data, collection site, geographical information, and profiles of the sixteen microsatellite loci; Table S2: Information of the sixteen polymorphic microsatellite loci used for assessment the genetic diversity parameters and construction of the core collection of *M. dubia* germplasm bank from the Peruvian Amazon; Table S3: Assessment of genetic diversity parameters at the accessions level for *M. dubia* germplasm bank from the Peruvian Amazon using sixteen microsatellite loci; Table S4: A core collection of *M. dubia* germplasm bank from the Peruvian Amazon obtained based on the sixteen polymorphic microsatellite loci.

Author Contributions: Conceptualization, J.C.C., and J.D.M.; methodology, J.D.P., C.G.C., D.E.M., S.A.I., and P.M.A.; validation, S.A.I., C.G.C., and P.M.A.; formal analysis, S.J.V.G., B.E.V., F.A., and N.R.V.; resources, J.C.C. and J.D.M.; writing—original draft preparation, S.J.V.G., and J.C.C.; writing—review and editing, J.C.C. and J.D.M.; supervision, J.C.C.; project administration, M.C.; funding acquisition, J.C.C. and J.D.M. All authors have read and agreed to the published version of the manuscript.

Funding: Funding was provided by the Universidad Nacional de la Amazonía Peruana (grant numbers: R.R. 0686-2015-UNAP and R.R. 0449-2024-UNAP), the American Public University System, the Grainger Bioinformatics Center, and the Pritzker Laboratory for Molecular Systematics and Evolution operated with support from the Pritzker Foundation.

Data Availability Statement: The data are included in the article or Supplemental Materials.

Acknowledgments: In this section, you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Villachica, H. *El Cultivo Del Camu Camu (Myrciaria Dubia H.B.K. McVaugh) En La Amazoni'a Peruana*; Pro-Tempore, S., Ed.; 1st ed.; Tratado de Cooperación Amazónica: Lima, 1996;
2. Bradfield, R.B.; Roca, A. Camu-Camu-a Fruit High in Ascorbic Acid. *J Am Diet Assoc* **1964**, *44*, 28–30.
3. Imán Correa, S.; Luz Bravo Zamudio; Solís, V.S.; Cruz, C.O. Contenido de vitamina C en frutos de camu camu *Myrciaria dubia* (H.B.K) Mc Vaugh, en cuatro estados de maduración, procedentes de la Colección de Germoplasma del INIA Loreto, Perú. *Sci. Agropecu* **2011**, *2*, 123–130.
4. Akter, M.S.; Oh, S.; Eun, J.B.; Ahmed, M. Nutritional compositions and health promoting phytochemicals of Camu-Camu (*Myrciaria dubia*) Fruit: A Review. *Food Res. Int* **2011**, *44*, 1728–1732, doi:10.1016/j.foodres.2011.03.045.
5. Akachi, T.; Shiina, Y.; Kawaguchi, T.; Kawagishi, H.; Morita, T.; Sugiyama, K. 1-Methylmalate from Camu-Camu (*Myrciaria dubia*) suppressed D-galactosamine-induced liver injury in rats. *Biosci. Biotechnol. Biochem.* **2010**, *74*, 573–578, doi:10.1271/bbb.90775.
6. Castro, J.C.; Maddox, J.D.; Cobos, M.; Imán, S.A. *Myrciaria dubia* “Camu Camu” fruit: Health-promoting phytochemicals and functional genomic characteristics. *Breeding and Health Benefits of Fruit and Nut Crops* **2018**, 85–116, doi:10.5772/intechopen.73213.

7. Azevedo, L.; de Araujo Ribeiro, P.F.; de Carvalho Oliveira, J.A.; Correia, M.G.; Ramos, F.M.; de Oliveira, E.B.; Barros, F.; Stringheta, P.C. Camu-Camu (*Myrciaria dubia*) from commercial cultivation has higher levels of bioactive compounds than native cultivation (Amazon forest) and presents antimutagenic effects in vivo. *J. Sci. Food Agric* **2019**, *99*, 624–631, doi:10.1002/jsfa.9224.
8. Akinnifesi, F.K. Indigenous Fruit Trees in the Tropics: Domestication, Utilization and Commercialization; 1st Edition.; CABI: Cambridge, MA, USA, 2008; ISBN 978-1-84593-110-0.
9. Penn, J. The cultivation of Camu Camu (*Myrciaria dubia*): A tree planting programme in the Peruvian Amazon. *For Trees livelihood* **2006**, *16*, 85–101.
10. Khoury, C.; Laliberté, B.; Guarino, L. Trends in ex situ conservation of plant genetic resources: A review of global crop and regional conservation strategies. *Genet Resour Crop Evol* **2010**, *57*, 625–639, doi:10.1007/s10722-010-9534-z.
11. Priyanka, V.; Kumar, R.; Dhaliwal, I.; Kaushik, P. Germplasm conservation: Instrumental in agricultural biodiversity—A review. *Sustainability* **2021**, *13*, 6743, doi:10.3390/su13126743.
12. Cohen, J.I.; Williams, J.T.; Plucknett, D.L.; Shands, H. Ex situ conservation of plant genetic resources: Global development and environmental concerns. *Science* **1991**, *253*, 866–872, doi:10.1126/science.253.5022.866.
13. Rajpurohit, D.; Jhang, T. In situ and ex situ conservation of plant genetic resources and traditional knowledge. In *Plant genetic resources and traditional knowledge for food security*; Salgotra, R.K., Gupta, B.B., Eds.; Springer: Singapore, 2015; pp. 137–162 ISBN 978-981-10-0060-7.
14. Li, D.-Z.; Pritchard, H.W. The science and economics of ex situ plant conservation. *Trends Plant Sci* **2009**, *14*, 614–621, doi:10.1016/j.tplants.2009.09.005.
15. Plucknett, D.L.; Smith, N.J.H.; Williams, J.T.; Anishetty, N.M. Crop germplasm conservation and developing countries. *Science* **1983**, *220*, 163–169, doi:10.1126/science.220.4593.163.
16. Griffith, M.P.; Cartwright, F.; Dosmann, M.; Fant, J.; Freid, E.; Havens, K.; Jestrow, B.; Kramer, A.T.; Magellan, T.M.; Meerow, A.W.; et al. Ex situ conservation of large and small plant populations illustrates limitations of common conservation metrics. *Int. J. Plant Sci* **2021**, *182*, 263–276, doi:10.1086/713446.
17. Volis, S.; Blecher, M. Quasi in situ: A bridge between ex situ and in situ conservation of plants. *Biodivers Conserv* **2010**, *19*, 2441–2454, doi:10.1007/s10531-010-9849-2.
18. Wei, X.; Jiang, M. Meta-Analysis of Genetic Representativeness of Plant Populations under Ex Situ Conservation in Contrast to Wild Source Populations. *Conserv. Biol* **2021**, *35*, 12–23, doi:10.1111/cobi.13617.
19. Engels, J.M.M.; Ebert, A.W. A Critical Review of the Current Global Ex Situ Conservation System for Plant Agrobiodiversity. I. History of the Development of the Global System in the Context of the Political/Legal Framework and Its Major Conservation Components. *Plants* **2021**, *10*, 1557, doi:10.3390/plants10081557.
20. Ebert, A.W.; Engels, J.M.M. Plant Biodiversity and Genetic Resources Matter! *Plants* **2020**, *9*, 1706, doi:10.3390/plants9121706.
21. Ml, V.; L, S.; Al, D.; F, M.C. Microsatellite Markers: What They Mean and Why They Are so Useful. *Genet Mol Biol* **2016**, *39*, 312–328, doi:10.1590/1678-4685-GMB-2016-0027.
22. Pérez, F.; Irrarázabal, C.C.; Cossio, M.; Peralta, G.; Segovia, R.; Bosshard, M.; Hinojosa, L.F. Microsatellite Markers for the Endangered Shrub *Myrcogenia Rufa* (Myrtaceae) and Three Closely Related Species. *Conserv. Genet Resour* **2014**, *6*, 773–775, doi:10.1007/s12686-014-0212-x.
23. Miwa, M.; Tanaka, R.; Yamanoshita, T.; Norisada, M.; Kojima, K.; Hogetsu, T. Analysis of Clonal Structure of *Melaleuca cajuputi* (Myrtaceae) at a Barren Sandy Site in Thailand Using Microsatellite Polymorphism. *Trees* **2001**, *15*, 242–248, doi:10.1007/s004680100098.
24. Albaladejo, R.G.; Sebastiani, F.; González-Martínez, S.C.; González-Varo, J.P.; Vendramin, G.G.; Aparicio, A. Isolation of Microsatellite Markers for the Common Mediterranean Shrub *Myrtus Communis* (Myrtaceae). *Am. J. Bot.* **2010**, *97*, e23–e25, doi:10.3732/ajb.1000060.
25. Steane, D.A.; Conod, N.; Jones, R.C.; Vaillancourt, R.E.; Potts, B.M. A Comparative Analysis of Population Structure of a Forest Tree, *Eucalyptus Globulus* (Myrtaceae), Using Microsatellite Markers and Quantitative Traits. *Tree Genet. Genomes* **2006**, *2*, 30–38, doi:10.1007/s11295-005-0028-7.
26. Rossetto, M.; Slade, R.W.; Baverstock, P.R.; Henry, R.J.; Lee, L.S. Microsatellite Variation and Assessment of Genetic Structure in Tea Tree (*Melaleuca Alternifolia*—Myrtaceae). *Mol Ecol* **1999**, *8*, 633–643, doi:10.1046/j.1365-294x.1999.00622.x.
27. Šmíd, J.; Kalousová, M.; Mandák, B.; Houška, J.; Chládková, A.; Pinedo, M.; Lojka, B. Morphological and Genetic Diversity of Camu-Camu [*Myrciaria Dubia* (Kunth) McVaugh] in the Peruvian Amazon. *Plos One* **2017**, *12*, e0179886, doi:10.1371/journal.pone.0179886.
28. Rojas, S.; Clement, Y.; Nagao, E.O. Diversidade genética em acessos do banco de germoplasma de camu-camu (*Myrciaria dubia* [H.B.K.] McVaugh) do INPA usando marcadores microssatélites (EST-SSR). *Cienc. Tecnol. Agropecuaria* **2011**, *12*, 51–64, doi:10.21930/rcta.vol12_num1_art:215.
29. Doyle, J.J.; Doyle, J.L. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phyt. Bull.* **1987**, *19*, 11–15.
30. Sambrook, J.; Frisch, E.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: New York, USA, 1989; Vol. 2;.

31. Glenn, T.C.; Schable, N.A. Isolating Microsatellite DNA Loci. In *Methods in Enzymology*; Molecular Evolution: Producing the Biochemical Data; Academic Press, 2005; Vol. 395, pp. 202–222.
32. Schuelke, M. An Economic Method for the Fluorescent Labeling of PCR Fragments. *Nat Biotechnol* **2000**, *18*, 233–234, doi:10.1038/72708.
33. Maddox, J.D.; Feldheim, K.A. A Cost-Effective Size Standard for Fragment Analysis That Maximizes Throughput on Five Dye Set Platforms. *Conserv Genet Resour* **2014**, *6*, 5–7, doi:10.1007/s12686-013-0019-1.
34. 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, doi:10.1093/bioinformatics/bts199.
35. Peakall, R.; Smouse, P. GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—an Update. *Bioinformatics* **2012**, *28*, 2537–2539, doi:10.1093/bioinformatics/bts460.
36. Excoffier, L.; Smouse, P.; Quattro, J. Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* **1992**, *131*, 479–491, doi:10.1093/genetics/131.2.479.
37. Slatkin, M. Gene Flow in Natural Populations. *Annu Rev Ecol Evol Syst* **1985**, *16*, 393–430, doi:10.1146/annurev.es.16.110185.002141.
38. Reynolds, J.; Weir, B.S.; Cockerham, C.C. Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. *Genetics* **1983**, *105*, 767–779, doi:10.1093/genetics/105.3.767.
39. Jombart, T. Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* **2008**, *24*, 1403–1405, doi:10.1093/bioinformatics/btn129.
40. Jombart, T.; Kamvar, Z.N.; Collins, C.; Lustrik, R.; Beugin, M.-P.; Knaus, B.J.; Solymos, P.; Mikryukov, V.; Schliep, K.; Maié, T.; et al. Adegenet: Exploratory Analysis of Genetic and Genomic Data 2023.
41. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959, doi:10.1093/genetics/155.2.945.
42. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* **2003**, *164*, 1567–1587, doi:10.1093/genetics/164.4.1567.
43. Falush, D.; Stephens, M.; Pritchard, J.K. Inference of Population Structure Using Multilocus Genotype Data: Dominant Markers and Null Alleles. *Mol Ecol Notes* **2007**, *7*, 574–578, doi:10.1111/j.1471-8286.2007.01758.x.
44. Hubisz, M.J.; Falush, D.; Stephens, M.; Pritchard, J.K. Inferring Weak Population Structure with the Assistance of Sample Group Information. *Mol Ecol Resour* **2009**, *9*, 1322–1332, doi:10.1111/j.1755-0998.2009.02591.x.
45. Miller, J.M.; Cullingham, C.I.; Peery, R.M. The Influence of a Priori Grouping on Inference of Genetic Clusters: Simulation Study and Literature Review of the DAPC Method. *Heredity* **2020**, *125*, 269–280, doi:10.1038/s41437-020-0348-2.
46. Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425, doi:10.1093/oxfordjournals.molbev.a040454.
47. Nei, M. Genetic Distance between Populations. *Am. Nat* **1972**, *106*, 283–292.
48. Pembleton, L.W.; Cogan, N.O.I.; Forster, J.W. StAMPP: An R Package for Calculation of Genetic Differentiation and Structure of Mixed-Ploidy Level Populations. *Mol Ecol Resour* **2013**, *13*, 946–952, doi:10.1111/1755-0998.12129.
49. Pembleton, L.W. StAMPP: Statistical Analysis of Mixed Ploidy Populations 2021.
50. 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, doi:10.7717/peerj.281.
51. Kamvar, Z.N.; Tabima, J.F.; Everhart, S.E.; Brooks, J.C.; Krueger-Hadfield, S.A.; Sotka, E.; Knaus, B.J.; Meirmans, P.G.; Chevalier, F.D.; Folarin, D.; et al. Poppr: Genetic Analysis of Populations with Mixed Reproduction 2024.
52. De Beukelaer, H.; Davenport, G.F.; Fack, V. Core Hunter 3: Flexible Core Subset Selection. *BMC Bioinformatics* **2018**, *19*, 203, doi:10.1186/s12859-018-2209-z.
53. Wright, S. *Evolution and the Genetics of Populations: Variability Within and Among Natural Populations*; University of Chicago Press: Chicago, IL, 1984; Vol. 4; ISBN 978-0-226-91041-3.
54. Goudet, J. Hierfstat, a Package for r to Compute and Test Hierarchical F-Statistics. *Mol Ecol Notes* **2005**, *5*, 184–186, doi:10.1111/j.1471-8286.2004.00828.x.
55. Goudet, J.; Jombart, T.; Kamvar, Z.N.; Archer, E.; Hardy, O. Hierfstat: Estimation and Tests of Hierarchical F-Statistics 2022.
56. R: The R Project for Statistical Computing Available online: <https://www.r-project.org/> (accessed on 13 May 2024).
57. Wickham, H.; Navarro, D.; Lin, T. *Ggplot2: Elegant Graphics for Data Analysis*; 1st Edition.; Springer, 2009; ISBN 0387981403.

58. Wickham, H.; Chang, W.; Henry, L.; Pedersen, T.L.; Takahashi, K.; Wilke, C.; Woo, K.; Yutani, H.; Dunnington, D.; Brand, T. van den; et al. Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics 2024.
59. Kolde, R. Pheatmap: Pretty Heatmaps 2019.
60. Tang, D.; Chen, M.; Huang, X.; Zhang, G.; Zeng, L.; Zhang, G.; Wu, S.; Wang, Y. SRplot: A Free Online Platform for Data Visualization and Graphing. *PLoS One* **2023**, *18*, e0294236.
61. de Campos Telles, M.P.; Coelho, A.S.G.; Chaves, L.J.; Diniz-Filho, J.A.F.; Valva, F.D. Genetic Diversity and Population Structure of *Eugenia Dysenterica* DC. ("cagaiteira"—Myrtaceae) in Central Brazil: Spatial Analysis and Implications for Conservation and Management. *Conserv. Genet* **2003**, *4*, 685–695, doi:10.1023/B:COGE.0000006124.55109.5e.
62. Alves, R.M.; Silva, C.R. de S.; Silva, M.S. da C.; Silva, D.C. de S.; Sebbenn, A.M. Diversidade Genética Em Coleções Amazônicas de Germoplasma de Cupuaçuzeiro [*Theobroma Grandiflorum* (Willd. Ex Spreng.) Schum.]. *Rev. Bras. Frutic.* **2013**, *35*, 818–828, doi:https://doi.org/10.1590/S0100-29452013000300019.
63. Baldoni, A.B.; Ribeiro Teodoro, L.P.; Eduardo Teodoro, P.; Tonini, H.; Dessaune Tardin, F.; Alves Botin, A.; Hoogerheide, E.S.S.; de Carvalho Campos Botelho, S.; Lulu, J.; de Farias Neto, A.L.; et al. Genetic Diversity of Brazil Nut Tree (*Bertholletia excelsa* Bonpl.) in Southern Brazilian Amazon. *For. Ecol. Mang.* **2020**, *458*, 117795, doi:10.1016/j.foreco.2019.117795.
64. Zhang, D.; Boccara, M.; Motilal, L.; Mischke, S.; Johnson, E.S.; Butler, D.R.; Bailey, B.; Meinhardt, L. Molecular Characterization of an Earliest Cacao (*Theobroma cacao* L.) Collection from Upper Amazon Using Microsatellite DNA Markers. *Tree Genet Genomes* **2009**, *5*, 595–607, doi:10.1007/s11295-009-0212-2.
65. Gaiotto, F.A.; Grattapaglia, D.; Vencovsky, R. Genetic Structure, Mating System, and Long-Distance Gene Flow in Heart of Palm (*Euterpe edulis* Mart.). *J. Hered.* **2003**, *94*, 399–406, doi:10.1093/jhered/esg087.
66. Sytsma, K.J.; Litt, A.; Zjhra, M.L.; Chris Pires, J.; Nepokroeff, M.; Conti, E.; Walker, J.; Wilson, P.G. Clades, Clocks, and Continents: Historical and Biogeographical Analysis of Myrtaceae, Vochysiaceae, and Relatives in the Southern Hemisphere. *Int. J. Plant Sci.* **2004**, *165*, S85–S105, doi:10.1086/421066.
67. Migicovsky, Z.; Warschefsky, E.; Klein, L.L.; Miller, A.J. Using Living Germplasm Collections to Characterize, Improve, and Conserve Woody Perennials. *Crop Sci.* **2019**, *59*, 2365–2380, doi:10.2135/cropsci2019.05.0353.
68. Cardoso, R.; Ruas, C.F.; Giacomini, R.M.; Ruas, P.M.; Ruas, E.A.; Barbieri, R.L.; Rodrigues, R.; Gonçalves, L.S.A. Genetic Variability in Brazilian *Capsicum baccatum* Germplasm Collection Assessed by Morphological Fruit Traits and AFLP Markers. *PLoS One* **2018**, *13*, e0196468, doi:10.1371/journal.pone.0196468.
69. Manel, S.; Schwartz, M.K.; Luikart, G.; Taberlet, P. Landscape Genetics: Combining Landscape Ecology and Population Genetics. *Trends Ecol Evol.* **2003**, *18*, 189–197, doi:https://doi.org/10.1016/S0169-5347(03)00008-9.
70. Oakley, C.G.; Lundemo, S.; Ågren, J.; Schemske, D.W. Heterosis Is Common and Inbreeding Depression Absent in Natural Populations of *Arabidopsis thaliana*. *J. Evol. Biol.* **2019**, *32*, 592–603, doi:10.1111/jeb.13441.
71. Rawat, S.; Annapurna, D.; Arunkumar, A.N.; Geeta, J. Genetic Diversity and Population Structure in Fragmented Natural Populations of *Melia dubia* Cav. Revealed by SSR Markers—Its Implications on Conservation. *Plant Mol Biol Rep.* **2021**, *40*, 247–255, doi:10.1007/s11105-021-01309-3.
72. Yu, K.; Wang, H.; Liu, X.; Xu, C.; Li, Z.; Xu, X.; Liu, J.; Wang, Z.; Xu, Y. Large-Scale Analysis of Combining Ability and Heterosis for Development of Hybrid Maize Breeding Strategies Using Diverse Germplasm Resources. *Front Plant Sci.* **2020**, *11*, 660, doi:10.3389/fpls.2020.00660.
73. Mr, L.; Aj, S.; Je, R. Heterosis and Hybrid Crop Breeding: A Multidisciplinary Review. *Front Genet.* **2021**, *12*, 643761, doi:10.3389/fgene.2021.643761.
74. Ramírez, F. Breeding Programs. In *Latin American Blackberries Biology*; Springer: Cham, 2023; pp. 157–162.
75. León, N.; Murillo, O.; Badilla, Y.; Ávila, C.; Murillo, R. Expected Genetic Gain and Genotype by Environment Interaction in Almond (*Dipteryx Panamensis* (Pittier) Rec. and Mell) in Costa Rica. *Silvae Genet* **2017**, *66*, 9–17, doi:10.1515/sg-2017-0002.
76. Bost, J. *Persea Schiedeana*: A High Oil "Cinderella Species" Fruit with Potential for Tropical Agroforestry Systems. *Sustainability* **2013**, *6*, 99–111, doi:10.3390/su6010099.
77. Lemes, M.R.; Gribel, R.; Proctor, J.; Grattapaglia, D. Population Genetic Structure of Mahogany (*Swietenia Macrophylla* King, Meliaceae) across the Brazilian Amazon, Based on Variation at Microsatellite Loci: Implications for Conservation. *Mol Ecol.* **2003**, *12*, 2875–2883, doi:10.1046/j.1365-294X.2003.01950.x.
78. Gamba, D.; Muchhala, N. Global Patterns of Population Genetic Differentiation in Seed Plants. *Mol Ecol* **2020**, *29*, 3413–3428, doi:10.1111/mec.15575.
79. Slatkin, M. Gene Flow and the Geographic Structure of Natural Populations. *Science* **1987**, *236*, 787–792, doi:10.1126/science.3576198.
80. Dick, C.W.; Hardy, O.J.; Jones, F.A.; Petit, R.J. Spatial Scales of Pollen and Seed-Mediated Gene Flow in Tropical Rain Forest Trees. *Trop. Plant Biol* **2008**, *1*, 20–33, doi:10.1007/s12042-007-9006-6.
81. Wright, S. Isolation by Distance. *Genetics* **1943**, *28*, 114–138, doi:10.1093/genetics/28.2.114.

82. Sujii, P.S.; Martins, K.; Wadt, L.H. de O.; Azevedo, V.C.R.; Solferini, V.N. Genetic Structure of *Bertholletia excelsa* Populations from the Amazon at Different Spatial Scales. *Conserv. Genet.* **2015**, *16*, 955–964, doi:10.1007/s10592-015-0714-4.
83. Dick, C.W.; Bermingham, E.; Lemes, M.R.; Gribel, R. Extreme Long-Distance Dispersal of the Lowland Tropical Rainforest Tree *Ceiba pentandra* L. (Malvaceae) in Africa and the Neotropics. *Mol Ecol* **2007**, *16*, 3039–3049, doi:10.1111/j.1365-294X.2007.03341.x.
84. Scotti-Saintagne, C.; Dick, C.W.; Caron, H.; Vendramin, G.G.; Troispoux, V.; Sire, P.; Casalis, M.; Buonamici, A.; Valencia, R.; Lemes, M.R.; et al. Amazon Diversification and Cross-Andean Dispersal of the Widespread Neotropical Tree Species *Jacaranda copaia* (Bignoniaceae). *J. Biogeogr.* **2013**, *40*, 707–719, doi:10.1111/j.1365-2699.2012.02797.x.
85. Motamayor, J.C.; Lachenaud, P.; Mota, J.W. da S. e; Loor, R.; Kuhn, D.N.; Brown, J.S.; Schnell, R.J. Geographic and Genetic Population Differentiation of the Amazonian Chocolate Tree (*Theobroma cacao* L.). *PLoS One* **2008**, *3*, e3311, doi:10.1371/journal.pone.0003311.
86. Gandara, F.B.; Da-Silva, P.R.; de Moura, T.M.; Pereira, F.B.; Gobatto, C.R.; Ferraz, E.M.; Kageyama, P.Y.; Tambarussi, E.V. The Effects of Habitat Loss on Genetic Diversity and Population Structure of *Cedrela fissilis* Vell. *Trop. Plant Biol* **2019**, *12*, 282–292, doi:10.1007/s12042-019-09234-2.
87. Wu, H.; Duan, A.; Wang, X.; Chen, Z.; Zhang, X.; He, G.; Zhang, J. Construction of a Core Collection of Germplasms from Chinese Fir Seed Orchards. *Forests* **2023**, *14*, 305, doi:10.3390/f14020305.
88. Yao, Z.; Feng, Z.; Wu, C.; Tang, L.; Wu, X.; Chen, D.; Wang, Q.; Fan, K.; Wang, Y.; Li, M. Analysis of Genetic Diversity and Construction of a Core Collection of *Ginkgo biloba* Germplasm Using EST-SSR Markers. *Forests* **2023**, *14*, 2155, doi:10.3390/f14112155.
89. Teodosiu, M.; Mihai, G.; Ciocîrlan, E.; Curtu, A.L. Genetic Characterisation and Core Collection Construction of European Larch (*Larix decidua* Mill.) from Seed Orchards in Romania. *Forests* **2023**, *14*, 1575, doi:10.3390/f14081575.
90. Zhu, Y.; An, W.; Peng, J.; Li, J.; Gu, Y.; Jiang, B.; Chen, L.; Zhu, P.; Yang, H. Genetic Diversity of Nanmu (Phoebe Zhennan S. Lee. et F. N. Wei) Breeding Population and Extraction of Core Collection Using nSSR, cpSSR and Phenotypic Markers. *Forests* **2022**, *13*, 1320, doi:10.3390/f13081320.
91. Yan, P.; Zhang, L.; Hao, J.; Sun, G.; Hu, Z.; Wang, J.; Wang, R.; Li, Z.; Zhang, H. Construction of a Core Collection of Korean Pine (*Pinus koraiensis*) Clones Based on Morphological and Physiological Traits and Genetic Analysis. *Forests* **2024**, *15*, 534, doi:10.3390/f15030534.
92. Dos Santos, C.; de Andrade, L.; do Carmo, C.; de Oliveira, E. Development of Cassava Core Collections Based on Morphological and Agronomic Traits and SNPS Markers. *Front. Plant Sci.* **2023**, *14*, 1250205, doi:10.3389/fpls.2023.1250205.
93. Koorevaar, T.; Willemsen, J.H.; Visser, R.G.F.; Arens, P.; Maliepaard, C. Construction of a Strawberry Breeding Core Collection to Capture and Exploit Genetic Variation. *BMC Genom.* **2023**, *24*, 1–13, doi:10.1186/s12864-023-09824-1.
94. Bidot Martínez, I.; Valdés de la Cruz, M.; Riera Nelson, M.; Bertin, P. Establishment of a Core Collection of Traditional Cuban *Theobroma cacao* Plants for Conservation and Utilization Purposes. *Plant Mol Biol Rep.* **2016**, *35*, 47–60, doi:10.1007/s11105-016-0999-6.

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.