horticulturae

MDPI

Article

Genetic Diversity and Structure in *Coffea canephora* Genotypes from the Amazon Region

Shayenne Hevelyn Farias Fernandes [1], Caroline de Souza Bezerra [1], Santiago Linorio Ferreyra Ramos [1], Ricardo Lopes [2], Marcelo Curitiba Espíndula [3], Thaynara Silva Ramos [4], Rodrigo Rodrigues Matiello [5], Maria José Marques [1], Carlos Henrique Salvino Gadelha Meneses [4,*] and Maria Teresa Gomes Lopes [1,*]

[1] Faculdade de Ciências Agrárias, Universidade Federal do Amazonas, Avenida Rodrigo Otávio Ramos, 3.000, Bairro Coroado, Manaus 69077-000, AM, Brazil; shayennefarias@gmail.com (S.H.F.F.); caroline.bezerra@ufam.edu.br (C.d.S.B.); slfr@ufam.edu.br (S.L.F.R.); marques.f.j.m@gmail.com (M.J.M.)
[2] Embrapa Amazônia Ocidental, Empresa Brasileira de Pesquisa Agropecuária, Rodovia AM-010, Km 29, Manaus 69010-970, AM, Brazil; ricardo.lopes@embrapa.br
[3] Centro de Pesquisa Agroflorestal de Rondônia, Empresa Brasileira de Pesquisa Agropecuária, Rodovia BR-364, Km 5.5, Porto Velho 76815-800, RO, Brazil; marcelo.espindula@embrapa.br
[4] Programa de Pós-Graduação em Ciências Agrárias, Departamento de Biologia, Centro de Ciências Biológicas e da Saúde, Universidade Estadual da Paraíba, Campina Grande 58429-500, PB, Brazil; thaynara.ramos@aluno.uepb.edu.br
[5] Setor de Ciências Agrárias e de Tecnologia, Departamento de Fitotecnia e Fitossanidade, Universidade Estadual de Ponta Grossa, Av. Carlos Cavalcanti 4748, Ponta Grossa 84030-900, PR, Brazil; rrmatiel@uepg.br
* Correspondence: carlos.meneses@servidor.uepb.edu.br (C.H.S.G.M.); mtglopes@ufam.edu.br (M.T.G.L.); Tel.: +55-(83)-99991-7717 (C.H.S.G.M.); +55-(92)-98121-0021 (M.T.G.L.)

Abstract

Coffea canephora is economically and socially important for small-scale agriculture in Northern Brazil. To identify genotypes adapted to Amazonian edaphoclimatic conditions, clones of the species have been evaluated across multiple locations in Amazonas. Introducing genetically selected materials into comparable environments may promote consistent productivity gains in the short and medium term. In this context, the aim of this study was to assess the genetic diversity of different *C. canephora* genotypes using microsatellite markers, which will support the development of superior genotypes adapted to Amazon conditions. A total of 43 *C. canephora* genotypes were analyzed. Leaves were collected for genomic DNA extraction and were standardized and amplified by PCR using microsatellite primers. Genotyping was performed via capillary electrophoresis, allowing for the determination of allele sizes. Genetic structure was inferred, and genetic diversity parameters were estimated. The average observed heterozygosity (H_O = 0.64) exceeded the expected heterozygosity (H_E = 0.53), and the average inbreeding coefficient (f = −0.19) indicated an excess of heterozygotes. The results revealed high genetic variability among the evaluated genotypes. These findings highlight the broad genetic diversity of *C. canephora*, reinforcing its potential as a genetic basis for selection and the development of cultivars adapted to the environmental conditions of the Amazon.

Keywords: molecular markers; heterozygosity; germplasm bank; genetic structure

Academic Editors: Marco Cristancho and Jorge Berny

Received: 15 January 2026
Revised: 10 February 2026
Accepted: 17 February 2026
Published: 20 February 2026

1. Introduction

Coffee plays an essential role in the global scenario and influences cultural, economic, and social aspects. It is estimated that roughly 3 billion cups are consumed daily, generating approximately US$200 billion per year and providing employment for 125 million

people [1,2]. The genus *Coffea* is diverse and comprises 130 different species; however, only two are considered commercially important [3,4], which together account for 99% of world production [5]. *Coffea arabica* L. (Arabica coffee) dominates the market and represents 56% of production, while *Coffea canephora* Pierre ex Froehner A. (Robusta/Conilon) corresponds to the remaining 44% [6,7].

Brazil is the world's largest coffee producer and, in 2024, exported 54.2 million 60 kg bags. Of this total, *C. canephora* represented 26.94%, while *C. arabica* accounted for 73.06%. The country occupies the second position in global *C. canephora* production and stands out as the largest producer and the second largest consumer of coffee worldwide. The country has consolidated its leadership in the sector, with investment in the cultivation of the two most relevant species, *C. arabica* and *C. canephora*. Thus, the Brazilian market share was boosted, reinforcing the status of *C. canephora* as the second most traded coffee species in the world [8,9].

In relation to *C. arabica*, which has higher commercial value, *C. canephora* is a species with economic and social relevance in the northern region of Brazil, and it stands out for having rusticity, greater resistance to adverse climatic conditions and pests and diseases, lower sensitivity to the more stable productivity, and greater genetic variability. Thus, robusta/conilon is gaining prominence as a sustainable alternative to coffee [2,8,10].

Widely adopted by small farmers in family and labor-intensive production systems, the cultivation of *C. canephora*, in addition to generating income for these producers, stands out as a viable alternative to recover degraded areas in the Amazon, contributing to reducing pressure on the forest and boosting regional agribusiness [11]. The increase in productivity improvement results from the adoption of clonal varieties adapted to the region, together with management practices, including dense planting, pruning, fertilization, and irrigation. These innovations increased production by 37.11%, from 758,797 tons in 2012 to 1,040,401 tons in 2023 [12]. In 2023, the main producing states were Espírito Santo (61.66%), Rondônia (21.59%), and Bahia (13.24%), followed by Minas Gerais, Mato Grosso, Acre, Amazonas, Pará, and São Paulo. The average annual net profit per hectare was R$27,664.90, with an average productivity of 2.56 tons per hectare/year in established crops [12].

The species *C. canephora* was introduced in Brazil in 1912, in the state of Espírito Santo; from the 1980s, the state consolidated itself as a reference, driven by improvement programs of Incaper and Embrapa [13]. Subsequently, Embrapa Rondônia, in collaboration with farmers of the state, worked on the selection of genotypes adapted to the conditions of the Western Amazon and, in 2016, in partnership with the Federal University of Amazonas (UFAM), evaluated clones of *C. canephora* in Manaus, Silves, and Itacoatiara to identify superior genetic materials for adaptation in these regions of the Amazon [14].

The clonal genotypes resulting from the genetic improvement program of Embrapa originate from controlled hybridization between plants of the groups 'Conilon' and 'Robusta', in a process that began in 2004 in the experimental field of Embrapa Rondônia. These genotypes represent an advance over the variety Conilon—BRS Ouro Preto for being an intraspecific hybrid that presents high vegetative vigor, high productivity, and tolerance to the main coffee diseases, such as orange rust and cercosporiosis [15].

Understanding and utilizing genetic diversity in existing breeding program genotypes is essential for sustaining hybridization-based breeding programs involving the use of genotypes and for supporting the recommendation of clones for planting [16]. The breeder needs to know the available germplasm in terms of genetic divergence because when combined with the knowledge of the behavior of the parents alone, it can result in a greater heterotic effect. In addition, genetic distance measurements have been useful in germplasm

conservation, in establishing relationships between genetic and geographic diversity, and in avoiding the genetic vulnerability of crops [17].

Co-dominant molecular markers, such as microsatellites or Sequence Simple Repeat (SSR), are a fast and effective tool in the analysis of genetic variations in the study of several plant species [18]. These allow us to obtain the degree of polymorphisms in one or more populations, answering important questions in population genetics studies [19]. In genetic conservation, it is possible to obtain estimates of genetic variation, population structure, gene flow, demographic history, and hybridization events [20,21]. These markers also assist in the identification of accessions in germplasm conservation units or to investigate genetic processes, drift occurrence, and variety protection [21,22]. This is because the genomes of eukaryotes contain complex and simple repetitive sequences that can be used as DNA markers, which occur at high frequencies and are arbitrarily better distributed, leading to polymorphic genetic loci [23]. The SSRs have been widely used in studies of the structure and genetic diversity of different species native to the Amazon, such as *Astrocaryum aculeatum* [24], *Bixa orellana* var. *urucurana* [25], and *Euterpe precatoria* [26].

After noting the importance of *C. canephora*, it becomes essential to understand the structure and genetic diversity of the new genotypes of this species in comparison to hybrid cultivars, aiming to advance their improvement. Given this need, the present study aimed to characterize the new genotypes of *C. canephora* Pierre ex Froehner, as well as hybrid cultivars belonging to the germplasm Bank of Embrapa Rondônia, based on the analysis of genetic structure and diversity. For this, specific and transferred SSR markers were used, which made it possible to estimate genetic diversity parameters and identify groups of divergence among the 43 evaluated genotypes.

2. Materials and Methods

2.1. Plant Material and Collection

Forty-three *C. canephora* genotypes from Embrapa's Rondônia genetic improvement program were evaluated and made available for this study. The genotypes analyzed in the following order were: BRS1216, BRS2299, BRS2314, BRS2336, BRS2357, BRS3137, BRS3193, BRS3210, BRS3213, BRS3220, BRS0012, BRS056, BRS061, BRS073, BRS088, BRS089, BRS120, BRS125, BRS130, BRS155, BRS160, BRS184, BRS189, BRS199, BRS203, RL1PL01, RL1PL12, RL2PL01, RL2PL12, RL3PL01, RL3PL12, RL4PL01, RL4PL12, BI101, BI102, BI103, BI104, BI105, BI106, BI107, BI108, BI109, and BI201. The first 11 genotypes are the new hybrid cultivars called Amazonian Robustas, with BRS0012 not yet released as a cultivar [14]. The following 14 are genotypes called Conilon—BRS Ouro Preto, with their different clones [14]. The next eight genotypes are materials that have been phenotypically cataloged as Robustas. The final 10 genotypes are Conilon plants, originating from the state of Espírito Santo, RN, which were accessed by Embrapa's Rondônia genetic improvement program.

From each genotype, at least two new healthy leaves were collected. The foliar samples were individually packed in zip-lock plastic bags, previously identified, containing silica gel for dehydration. Subsequently, the samples were stored at −20 °C in the Coffee Breeding Program Laboratory of Embrapa Rondônia, where they were lyophilized. Then, the samples were sent to the Plant Genetic Improvement Laboratory of the Faculty of Agricultural Sciences of the Federal University of Amazonas (UFAM) for molecular analysis.

2.2. DNA Extraction and Genotyping

The genomic DNA extraction was performed using the cationic detergent protocol CTAB 2X (Cationic Hexadecyl Trimethyl Ammonium Bromide) described by [27] and quantified with GelRed dye. The genomic DNA was standardized to a concentration of 10 ng·μL^{-1} for use in amplification. The 43 genotypes were amplified by polymerase chain

reaction (PCR) using 10 initiators developed for *C. canephora* (Me_13; Mg_257; Mg_358; Mg_445; Mg_753; Mg_461; Mg_501; Mg_779; Mg_368; and Mg_447) and amplified according to the protocol proposed by [28]. PCRs presented a total volume of 10 µL, containing 10 ng of genomic DNA, 1X buffer (10× standard Taq reaction buffer), 210 µM of each dNTP, 1.5 mM of $MgCl_2$, 0.16 µM of *forward primer* and M13 (FAM or NED dyes) [29], 0.32 µM of *reverse primer*, 1.05 U of Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA), and 3.49 µL of ultrapure water. PCR amplifications were developed in two stages, according to the procedure described in [26] Ramos et al. (2022). PCR products were previously evaluated as 1.5% agarose gel stained with GelRed dye (Biotium) in 1× TBE buffer (pH 8.0).

A genotyping of the progenies was performed by capillary electrophoresis in the Automatic DNA Analyzer ABI 3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). GeneScan^{TM}-500 ROX^{R} standard size (Life Technologies, Carlsbad, CA, USA) was used to determine the size of alleles. The amplified fragments were observed and analyzed with GENEMAPPER software version 4.0 (Applied Biosystems, Foster City, CA, USA).

2.3. Statistical Analysis

To identify genetic differences between progenies data (Table S1), allowing the formation of clusters and the estimation of genetic diversity parameters, two analyses were conducted. The first consisted of a Bayesian approach to determine the number of genetic clusters present in the set of evaluated genotypes, using the STRUCTURE software version 2.3.4 [30], with the *Admixture* model, widely applied in real or natural populations. The number of clusters (K) was defined from 1 to 10, and for each K, 10 interactions were performed, with a *burn-in* of 100,000 followed by 500,000 interactions through the *Markov Chain Monte Carlo* algorithm (MCMC). The determination of the most probable value of K was performed based on the method of *Evanno* ΔK [31]. After defining the optimal number of clusters, the repetitions were consolidated using CLUMPP—Cluster Matching and Permutation Program version 1.1.2 [32]—and the structure of the clusters was visualized graphically with the aid of Distruct v.1.1 [33].

In the second analysis of the genetic structure, the Discriminant Analysis of Principal Components (DAPC) [34] was used, considering the groups as unknown [35]. First, the probability of adhesion of each genotype sampled to diverse groups was determined using *k-means* [34]. Next, the ideal number of clusters was identified using *k-means*, comparing and discriminating against possible clusters using the Bayesian Information Criterion (BIC) [34]. It used 1,000,000 permutations, within the functions *find.structure* and DAPC of the package adegenet version 2.1.0 [36], on the platform R version 4.5.2 [37].

In order to confirm the existence of differentiation or genetic structure among the 43 genotypes of *C. canephora* sampled, the F statistics of [38] were calculated through the algorithms F_{IT}, F_{IS}, and F_{ST} of Weir and [39], in which the genetic structure analyses characterized by the analysis of variance of gene frequencies for the progenies collected consider the levels of total endogamy in individuals of all progenies ($F = F_{IT}$), the index of endogamy within subpopulations due to the reproductive system ($f = F_{IS}$), and genetic divergence or differentiation between genotypes ($\theta_P = F_{ST}$) [39].

A matrix was also calculated with the values of the fixing index (F_{ST}) pair to pair [39], between the four clusters obtained in the genetic structure analysis from the 43 progenies. These statistics were calculated and evaluated with significance ($p \leq 0.05$) through 20,000 *bootstrapping* interations, using the functions *wc* and *pairwise*, and *WCfst* of the *hierfstat* package [40] and the GDA software version 1.7-23 [41].

To verify the degree of genetic variation according to the hierarchical levels between and within the individual samples of the 43 genotypes, molecular variance analysis

(AMOVA) was performed based on the estimate of the extent of the genetic differentiation statistics, using the *poppr.amova* function within the poppr package version 2 [41]. The significance level of the variance components was performed using the Monte Carlo test, and *ΦCP*, population differentiation, *ΦCAP*, sample differentiation within populations, and *ΦCA* were evaluated using a permutation test implemented in the randtest function of *ade4* [42], with 9999 permutations.

After identifying the clusters by Bayesian methods (software STRUCTURE) and DAPC, the progenies were grouped by the largest number of clusters detected in genetic structure analyses. The clusters were numbered, and the genetic diversity parameters, such as mean number of alleles/loci (A), allelic richness (A_R), observed heterozygosity (H_O), expected heterozygosity (H_E), and fixation index (f), were estimated for each population using the functions *basic.stats* and *allelic.richness* of the *hierfstat* Package [40], within the R package [37]. The number of total alleles (A_T), with the function *nAll* in the package Adegenet version 2.1.1 [36], and private alleles (A_P) was estimated with the function *private_alleles* in the poppr package [41]. The linkage imbalance (*LI*) and Hardy–Weinberg equilibrium (*HWE*) were calculated using the GDA software [41]. *HWE* and *LI* were evaluated by Fisher's exact test with 20,000 permutations. The significance level ($p \leq 0.05$) of *HWE* and *LI* was adjusted with Bonferroni's correction [43].

3. Results

3.1. Clusters of Evaluated Progenies

The population structure analysis of the progenies, conducted in the STRUCTURE software, indicated that the model with $K = 3$ was the one that best represented the partition of *C. canephora* genotypes (Figure 1), showing the presence of three main genetic groupings in the evaluated set. Based on the individual allocation probabilities (q-values), the forty-three genotypes could be distributed into three genetically more homogeneous groups, formed by 26, 7, and 10 genotypes, respectively (Figure 1). Table 1 presents the detailed composition of these clusters, allowing us to identify both the genotypes allocated unequivocally in a single group and those that exhibited signs of admixture, that is, those that retain genome proportions associated with more than one cluster, which is consistent with the history of improvement of the material, in which there is use of Conilon and Robusta sources.

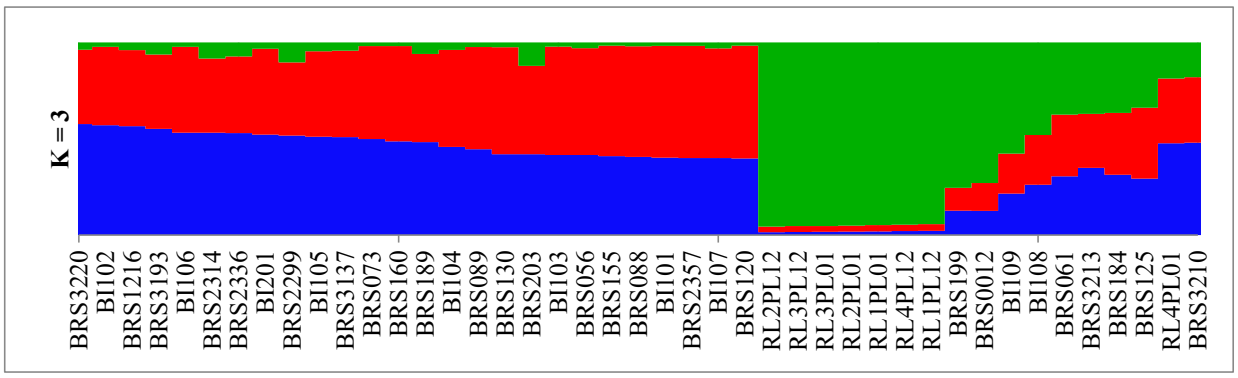

Figure 1. Structuring of the 43 progenies of *C. canephora*, indicating the presence of three groups ($K = 3$). Each color represents one of the inferred genetic clusters ($K = 3$) identified by the Bayesian analysis implemented in STRUCTURE. The proportional assignment of colors within each bar corresponds to the estimated membership coefficient (*q-value*) of each genotype.

Table 1. Distribution of *C. canephora* individuals among the clusters formed by Discriminant Analysis of Principal Components (DAPC) and STRUCTURE (*K*), based on four microsatellite loci.

Number	Progenies	DAPC Groupings				Clusters *K*		
		1	2	3	4	1	2	3
1	BRS0012	x						x
2	BRS155	x				x		
3	BRS088	x				x		
4	BI107	x				x		
5	BRS061	x						x
6	BRS184	x						x
7	BI104		x			x		
8	BRS073		x			x		
9	BRS130		x			x		
10	BI103		x			x		
11	BRS2357		x			x		
12	BRS120		x			x		
13	BRS199		x					x
14	BRS3213		x					x
15	BRS125		x					x
16	RL4PL01		x					x
17	BRS3210		x					x
18	BRS3220			x		x		
19	BI102			x		x		
20	BRS1216			x		x		
21	BRS3193			x		x		
22	BI106			x		x		
23	BRS2314			x		x		
24	BRS2336			x		x		
25	BI201			x		x		
26	BRS2299			x		x		
27	BI105			x		x		
28	BRS3137			x		x		
29	BRS160			x		x		
30	BRS189			x		x		
31	BRS089			x		x		
32	BRS203			x		x		
33	BRS056			x		x		
34	BI101			x		x		
35	BI109			x				x
36	BI108			x				x
37	RL2PL12				x		x	
38	RL3PL12				x		x	
39	RL3PL01				x		x	
40	RL2PL01				x		x	
41	RL1PL01				x		x	
42	RL4PL12				x		x	
43	RL1PL12				x		x	

By DAPC, the genotypes of *C. canephora* presented a more refined structure, with the formation of four genetically distinct groups (Figures 2 and 3; Table 1). The first three DAPC eigenvalues, corresponding to 34.94%, 26.28%, and 22.68% of the variance, together explained 83.90% of the genetic variation between the groups, a value considered high for this type of multivariate approach. This result indicates that most of the genetic divergence between genotypes was already captured in the first discriminant axes. In practice, this means that the groups formed by DAPC are not classification artifacts but reflect real

genetic differences between subsets of genotypes, possibly associated with the origin of the materials, the use of different parents in crosses, and/or different selection cycles.

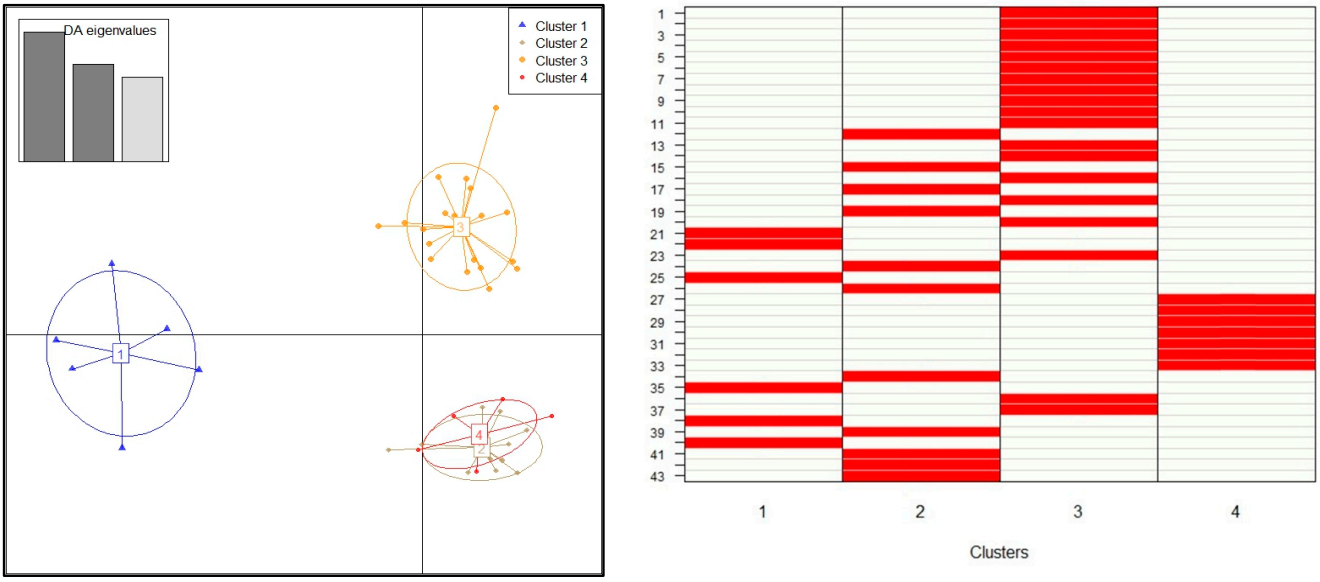

Figure 2. Distribution graph of the clusters formed by *k-means* method, each color represents one of the four genetic clusters identified by DAPC, assigns each genotype to a single group based on discriminant functions. The spatial separation among colored groups reflects the genetic differentiation captured by the first discriminant axes (**a**). Distribution graph of progenies in different clusters (clusters), where color and symbol combinations represent discrete group assignments (**b**).

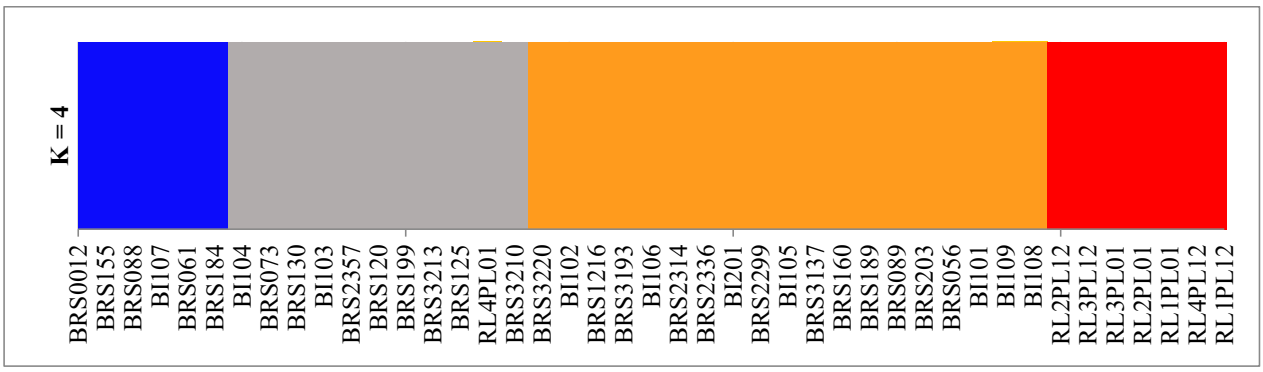

Figure 3. Graph of the structuring according to the groupings identified by the progenies according to the discriminant function obtained in the DAPC for the 43 progenies of *C. canephora*. Each color represents one of the four genetic clusters identified by DAPC, based on discriminant functions.

The results differ in the number of clusters inferred by the methods: the Bayesian analysis in STRUCTURE indicated three groups, while the DAPC revealed four. Despite this discrepancy, there is partial agreement between the approaches: the fourth grouping identified by DAPC corresponds mostly to Group 2 of structure (Table 1; Figures 1–3). This shared cluster is composed of the progenies RL2PL12, RL3PL12, RL3PL01, RL2PL01, RL1PL1, RL4PL12, and RL1PL12 (Table 1).

Table 1 presents the groupings obtained by the two genetic structure analyses (STRUCTURE and DAPC). It is noteworthy that the genotypes resulting from directed hybridizations between Robusta and Conilon parents, known as Amazonian Robustas and recommended by Embrapa (BRS1216, BRS2299, BRS2314, BRS2336, BRS2357, BRS3137, BRS3193, BRS3210, BRS3213, and BRS3220), show a consistent structuring pattern between the two

approaches, being allocated similarly in Group 1 of the Bayesian analysis (STRUCTURE) and in Group 3 of the DAPC.

Estimates of Wright's *F*-coefficients calculated for the four clusters of *C. canephora*, based on the 43 genotypes defined by the DAPC, indicate that the largest portion of genetic variability is concentrated within the clusters and not among them, which is typical of cultivated populations submitted to recombination and selection. The results indicate that the identified clusters represent genetically recognizable subsets within the panel (Table 2). The values of F_{iT} = 0.0278 and F_{IS} = −0.1636 were lower than the F_{ST}, indicating that the global level of inbreeding in the panel is low and that, within the clusters, there is even an excess of heterozygosity (negative F_{IS}). This pattern is consistent with the absence or low frequency of inbred matings and/or with the use of genetically contrasting parents in breeding programs. The confidence intervals obtained by bootstrapping indicated that F_{ST} was significantly greater than zero, indicating moderate genetic structuring among the clusters. F_{IS} presented significantly negative values, suggesting excess heterozygosity within the groups. In contrast, F_{IT} did not differ significantly from zero, indicating no significant global heterozygosity imbalance in the total set evaluated. Taken together, these results demonstrate that the panel simultaneously shows detectable genetic structuring (useful for the formation of crossing groups) and high intra-group diversity (favorable for selection within each cluster).

Table 2. Results of the estimates of Wright's *F* statistics (1951) obtained for the four clusters, composed of 43 *Coffea canephora* progenies, using 10 microsatellite loci.

	F_{IS}	F_{ST}	F_{IT}
Under all the loci	−0.1636	0.1645	0.0278
Superior ($CI_{95\%}$) [1]	−0.0019	0.2596	0.2264
Inferior ($CI_{95\%}$) [1]	−0.3389	0.0836	−0.1907

[1] $CI_{95\%}$ = 95% confidence interval through 20,000 *bootstrapping* interations. *Fis*: Inbreeding coefficient; F_{st}: Fixation index; *Fit*: Total inbreeding coefficient.

The paired estimates of F_{ST} revealed that all combinations between the clusters presented significant genetic differentiation (Table 3), which confirms that the four identified groups represent genetically distinguishable units within the set of *C. canephora* evaluated. Among the comparisons, it was highlighted that it was performed between Groups 1 and 4, which presented the highest value of F_{ST} (0.2102; 21.02%), indicating moderate-to-high differentiation and suggesting that these two groups combine genotypes with more contrasting genetic histories, possibly derived from different germplasm sources, different selection cycles, or less material exchange throughout the program. In the opposite direction, the smallest divergence was observed between Groups 2 and 3 (F_{ST} = 0.1206; 12.06%), a value that falls within the range of low-to-moderate differentiation, compatible with groups that, although recognizable as separate units, still maintain some degree of genetic proximity, either by partial sharing of parents or by more recent gene flow between them. This F_{ST} gradient (from the most divergent pair, G1 × G4, to the closest pair, G2 × G3) shows that the structure is not uniform among the groups and offers useful information for crossing planning: more distant pairs (such as 1 × 4) tend to maximize the recombination of alleles and may be preferred when the objective is to broaden variability or explore heterosis, while closer pairs (such as 2 × 3) can be used in consolidation schemes of features already present in the germplasm.

Table 3. Pairwise comparisons of the F_{ST} among the four clusters of *C. canephora* collected in the germplasm Bank of Embrapa in the state of Rondônia.

Groups	1	2	3
1			
2	0.140649 *		
3	0.141663 *	0.120582 *	
4	0.210178 *	0.201876 *	0.205776 *

* = F_{ST} pairwise significant comparisons, indicating difference among populations ($p \leq 0.05$).

The analysis of molecular variance (AMOVA) performed on the *C. canephora* genotype panel revealed significant genetic differentiation among the identified clusters (15.52% of the total variation; Φ = 0.1552; p = 0.0001), suggesting that these groups capture a real fraction of the genetic structure present (Table 4). Most of the genetic variation was associated with the residual level among genotypes (97.68%), a pattern commonly observed in studies based on SSR markers and consistent with high intra-group variability. However, permutation tests in AMOVA assess the statistical significance of the Φ statistics rather than directly testing the magnitude of the variance components. At this level, the Φ value was low (Φ = 0.0232) and not statistically significant (p = 0.7446), indicating the absence of detectable additional genetic structure in this contrast.

Table 4. Analysis of molecular variance (AMOVA) was performed for four populations of *Coffea canephora* collected in the germplasm Bank of Embrapa in the state of Rondônia.

AMOVA—No Individual Levels (Populations)					
Variation Source	**Medium Square**	**Variance Component**	**Percentage of Variation**	***p*-Value**	**Φ**
Among clusters	27,668	1131	15.52	0.0001	0.1552
Among genotypes within clusters	5193	−0.962	−13.20	1000	−0.1563
Among genotypes	7117	7117	97.68	0.7446	0.0232
Total	6.960	7.286	100.00		

Variation source, medium square, variance component, percentage of variation, and p-value = 0.000 (estimated based on 9999 permutations).

The component corresponding to genotypes within clusters showed a negative estimate (−13.20%; Φ = −0.1563), which is interpreted as differentiation close to zero and/or estimation instability at this hierarchical level and therefore should not be over-interpreted biologically. Taken together, the AMOVA results support the conclusion that detectable genetic structure exists among groups, although most genetic diversity remains distributed within clusters, a scenario consistent with the use of multiple germplasm sources and favorable for the exploitation of genetic variability in breeding programs.

3.2. Genetic Diversity

Of the 10 microsatellite loci analyzed, three showed significant deviation from the Hardy–Weinberg equilibrium (*HWE*) in clusters 2 and 3 ($p < 0.005$), indicating that, in these groups, the observed genotypic distribution does not follow the expected pattern for an ideal panmictic population. This deviation may be associated with the group formation history itself (use of a reduced number of parents, directed selection, recent introduction of materials, or even sampling effect), and not necessarily inbreeding. In addition, the linkage imbalance analysis showed that only 4.44% of locus × locus combinations were unbalanced for these same clusters (Table 5), which means that most markers acted approximately

independently and, therefore, SSR data are adequate to describe the genetic structure of the panel.

Table 5. Hardy–Weinberg equilibrium (*HWE*) and linkage disequilibrium (*LD*) for the four clusters of progenies of the de café *Coffea canephora* of active bank localized in Rondônia in Brazil.

Agru	CLD (*LD*%)		Loci—*HWE*									
			Me_13	**Mg_257**	**Mg_358**	**Mg_445**	**Mg_753**	**Mg_461**	**Mg_501**	**Mg_779**	**Mg_368**	**Mg_447**
1	0	(0.00)	0.7773	0.6242	0.0568	0.1116 [1]	0.0203	0.4232	0.3266 [1]	0.0802	0.5213 [1]	0.0810 [1]
2	2	(4.44)	0.9893	0.0016 *	0.0107 [1]	0.0003 * [1]	0.431	0.0661 [1]	0.1167	0.0047 * [1]	0.0430	0.8895
3	2	(4.44)	0.9962	0.0005 *	0.0074 [1]	0.0003 *	0.7352 [3]	0.0014 *	0.1745	0.0015 * [1]	0.9610 [1]	0.9004
4	0	(0.00)	0.4936 [2]	0.1172 [2]	0.0082	0.0719	0.1199 [1]	0.3208	0.0719 [1]	0.0211	0.8293 [2]	0.0364 [1]

* Deviations from the Hardy–Weinberg equilibrium with Fisher's exact test ($p < 0.005$, after adjusting the Bonferroni correction [44]); CLD = number of combinations with binding imbalance, selected with $p < 0.0011$, after adjusting the Bonferroni correction (RICE, 1989); [ ] = number of private alleles. Agru = Clusters of coffee progenies according to the result of the Main Discriminant Analysis of Principal Components (DAPC).

The DAPC analysis, performed with 43 genotypes of *C. canephora*, identified four distinct genetic groupings, all exhibiting high levels of diversity, which reinforces what had been observed in the structure analyses. The 10 loci used in genotyping revealed a total of 76 alleles, with an average of 4.4 alleles per locus, a value considered good for SSRs in perennial cultures. When the diversity was analyzed by group, it was observed to have varied from 3.8 alleles/loci (grouping 1) to 5.1 alleles/loci (grouping 3), showing that there are slightly richer alleles, which may indicate a broader origin of germplasm or lower selection pressure in this subset (Table 6).

Table 6. Genetic diversity indices of the four clusters obtained by the Discriminant Analysis of Principal Components (DAPC) method in 43 *Coffea canephora* progenies, using 10 microsatellite loci.

DAPC Groupings	n	A_T	A	A_R	H_O	H_E	f
01	6	38	3.8	3.29	0.585	0.592	−0.0251
02	11	46	4.6	3.29	0.680	0.594	−0.1892
03	19	51	5.1	3.11	0.683	0.559	−0.2269
04	7	39	3.9	3.25	0.786	0.664	−0.2328
Average	10.8	43.5	4.4	3.24	0.683	0.602	−0.1685

n = mean number of individuals analyzed by loci; A_T = total number of alleles identified in the population; A = mean number of alleles per locus; allele richness (A_R); H_O = observed heterozygosity; H_E = expected heterozygosity; f = endogamy coefficient.

The heterozygosity parameters also point to high variability within groups. The observed heterozygosity (H_O) ranged from 0.585 to 0.786, with an average of 0.683, being the lowest value recorded in group 1 and the highest in group 4. Already, the expected heterozygosity (H_E) presented an average of 0.602, oscillating from 0.592 (group 1) to 0.664 (group 4). In practically all cases, $H_O > H_E$, which is a strong indication of excess heterozygotes in the evaluated material. This pattern is confirmed by the mean endogamy coefficient per locus (f), which was negative ($f = -0.1685$), indicating that the frequency of homozygotes was lower than expected under HWE and suggesting the absence of systematic endogamous crosses, a result consistent with breeding programs that use directed crosses between divergent parents to increase variability. The average allele richness was estimated at 3.24 alleles per locus, a value that, even corrected for the sample size, confirms that the panel retains a good genetic base and can be exploited both for intra-group selection and for assembly of crosses between more distant groups.

The diversity analysis also revealed the presence of 23 private alleles distributed between the four clusters of *C. canephora* for the 10 microsatellite loci evaluated. The occurrence of unique alleles in more than one group is an important indication of real genetic differentiation between groups, since it shows that each subset of genotypes maintains variants that were not detected in the others. These private alleles are particularly valuable for the diagnosis of belonging (assignment) of new materials, monitoring of access conservation in germplasm banks, and expansion of the genetic basis in crosses, since they represent variability that can be introgressed in other groups.

As for the link imbalance (*LD*), most of the loci presented a pattern of independence in locus × locus combinations. In groups 1 and 4, no significant *LD* was detected (0%), suggesting that, in these groups, the markers are segregated approximately randomly and that there is no strong sampling or joint selection effect on the evaluated loci. In clusters 2 and 3, 4.44% combinations were unbalanced (Table 5), which is still low for studies with SSR and compatible with panels formed by materials of partially common or recently assembled origin. In practical terms, this low percentage of *LD* indicates that the set of markers is informative and a little redundant, and can be used safely for population structure analysis, estimates of diversity, and selection of parents without a significant risk of overestimating genetic relationships due to correlations between loci.

4. Discussion

The economic relevance of coffee is well established in national and international markets, particularly for the species *C. arabica*. However, *C. canephora* has shown notable expansion since its introduction in the 1980s, marked by significant increases in cultivated area and productivity. It also stands out for its adaptability to the edaphoclimatic conditions of the Legal Amazon, which has generated opportunities and socioeconomic development for farmers in the region who dedicate themselves to its cultivation [14]. In this context, the conservation and genetic resource management carried out by both Embrapa and on-farm farmers have been instrumental in accelerating the advancement of culture [13]. The introduction of new materials in Embrapa's *C. canephora* germplasm bank is therefore an essential strategy to expand the available genetic base and subsidize breeding programs. The characterization of the genetic diversity present in this collection is a priority to guide the genetic management and selection of parents. The identification of divergent groups, superior clones, heterotic groups, and potential parents is crucial for the design of crosses aimed at obtaining agronomic characteristics of interest [45–47]. Studies of this nature are fundamental for the strategic planning and progress of genetic improvement programs [48].

The genetic characterization of 43 genotypes of *C. canephora* from the program of Embrapa Rondônia revealed a panel with high variability and detectable structuring. At the intra-group diversity level, genotyping with 10 SSR loci recovered 76 alleles (an average of 4.4 alleles/locus) and an average allelic richness of 3.24, with 23 private alleles distributed among the groups, a robust indicator of exclusive variation that can be exploited in direct crosses and germplasm bank management.

As for genetic structure, complementary approaches were employed. STRUCTURE indicated $K = 3$ as the most parsimonious partition, revealing three main groups and admixture signs compatible with the history of improvement integrating Conilon and Robusta sources. In parallel, DAPC identified four genetically distinguishable clusters, all with high internal diversity. This duality (K = three by the Bayesian method vs. four clusters by DAPC) is consistent with recent methodological literature, which recognizes the greater accuracy of DAPC/PCA to separate sets with mixing gradients, while STRUCTURE describes well the thickness of the partition, especially in the presence of admixture. In operational breeding terms, the coexistence of three "macrogroups" and four DAPC sets

provides a practical map for organizing crossovers, heterotic consortia, and variability management [49]. The genetic structure analyses applied to *C. canephora* genotypes can be significantly influenced by factors such as geographic distance between populations [14] and gene flow, the latter occurring inversely proportional to geographical distance [50]. The formation of the four clusters identified by means of the DAPC analysis seems to reflect this dynamic, corroborated by pair-to-pair comparisons of F_{ST} values, which indicated significant genetic differentiation between all clusters (Table 3). The observed heterozygosity exceeded what was expected for most of the groups (mean H_O = 0.68; H_E = 0.60), and the inbreeding coefficient was consistently negative (f mean = −0.17), reinforcing the absence of systematic inbreeding and suggesting cross-mating and/or hybrid composition of the evaluated materials. These patterns were accompanied by a low proportion of linkage imbalances and *HWE* point deviations confined to some loci in two groups, without compromising population inference. This suggests that the gene flow between the evaluated genotypes is strongly influenced by their geographical origin, pointing to a pattern of distance isolation. This interpretation is supported by the positive and significant correlation between geographic distance and genetic distance, reflecting the observed genetic structuring [50,51].

The congruence between the groups obtained in different analyses (DAPC and STRUCTURE) reinforces the consistency of the detected genetic structure. Genotypes such as RL2PL12, RL3PL12, RL3PL01, RL2PL01, RL1PL01, RL4PL12, and RL1PL12 were grouped in a similar way in both approaches, as well as the hybrids known as Amazonian Robustas (BRS1216, BRS2299, BRS2314, BRS2336, BRS2357, BRS3137, BRS3193, SBR3210, BRS3213, and BRS3220), showing stable patterns of genetic connectivity between genotypes. The magnitude of the differentiation between groups was quantified by the *F*-statistics of Wright and by AMOVA. The global F_{ST} around 0.16 (95%CI = 0.08–0.26) characterizes moderate-to-high differentiation between groups; paired comparisons reached ~0.21 among the most distant groups, while the closest ones were close to ~0.12. In AMOVA, ~15.5% of the variance was attributed to differences between groups (Φ_{CP} = 0.155; p = 0.0001), with the largest fraction of the variation retained in individual genotypes, a typical scenario of elite panels where directed selection preserves intra-group heterogeneity. This reading is in line with updated references that classify F_{ST} from 0.05 to 0.15 as moderate differentiation and 0.15–0.25 as high, corroborating the usefulness of the detected structure to form crossing groups and explore heterosis without losing the internal genetic base [52].

The genetic differentiation by Wright's F statistic indicates the existence of a genetic structure among the genotypes evaluated by the F_{ST} [53]. Thus, much of the genetic variability observed for the set of sampled genotypes is mostly within the clusters, influenced by the reproductive system (F_{IS}), and presents significant differences when analyzed in pairs (F_{ST}), confirming the existence of genetic structure among the 43 genotypes sampled [38]. Similar information is observed in the AMOVA, which confirms the existence of genetic structure in the set of evaluated genotypes, showing significant differences between populations (Table 4); however, most of the genetic variation occurred among the evaluated genotypes (Table 4), as observed in the probability obtained. The analysis shows that the greatest concentration of diversity occurs within populations. This result can be strongly influenced by the reproductive characteristics of the species. *C. canephora* is an allogamous species, and in addition to having the ability to disperse its genetic material by the self-incompatibility system [54] and by the allele diversity that it presents [55], these germplasm sources and/or the use of materials of different origin are used in the breeding program. In biological terms, this means that there is still wide variation to be explored within each cluster, which is favorable for intra-group selection and maintenance of the genetic base.

Complementary evidence for the genetic organization of the evaluated panel was provided by the AMOVA results. The predominance of genetic variation at the level of individual genotypes is consistent with the existence of multiple germplasm sources and the use of materials of different origins in the breeding program. From a biological perspective, this pattern indicates that substantial genetic variability remains within clusters, which is favorable for intra-group selection and for maintaining a broad genetic base. The lack of statistical significance of this component in the permutation test may be associated with sample size, number of markers, or the distribution of variance among hierarchical levels.

From the applied point of view, these results have three direct implications for Amazonian coffee cultivation. First and foremost, the combination "high intra-group diversity associated with moderate/high F_{ST} between groups" allows drawing to cross between more distant pairs to maximize recombination and heterosis, while closer pairs favor consolidation of specific attributes. The presence of private alleles suggests that each cluster hosts rare variants that are potentially associated with agronomic performance, quality, or resilience; their planned introgression may broaden the adaptive basis of materials. Consequently, the negative f and $H_O > H_E$, congruent with cross-mating and hybrid origin, help to explain the reported productive stability for clones adapted to the Western Amazon.

This study aimed to characterize the genetic diversity of new accessions incorporated into the germplasm bank of *C. canephora* from Embrapa Rondônia, based on genetic structure analysis. The genotypes distributed in four clusters showed distinct levels of genetic diversity, although lower than those observed in wild populations and improved collections of *C. canephora* in the northeastern Democratic Republic of Congo [56]. This reduction may be related to the effects of the artificial selection process, which tends to restrict genetic variability. The observed (H_O) and expected (H_E) heterozygosity coefficients and negative values for fixation indices (f) indicate high genetic diversity [57], according to the excess of heterozygotes in the clusters. The low f and high values of H_O (Table 5) also indicate that the result may be influenced by the reproductive system of the species, that is, by allogamy [54]. It should be noted that the values of f and H_O were contrasting in the research with wild populations and improved collections in the Congo, being very high values for an allogamy (Aglaonema) plant [56]. Heterozygosity estimates how much variation exists in a population and how this variation is distributed depending on the alleles present in each locus [58]. The values of heterozygosity expected in clusters from *C. canephora* genotypes, with a mean of 0.602, were lower than those obtained for the observed heterozygosity, with a mean of 0.683, for all loci in this study. Low F_{IS} values in all the analyzed loci allowed the interpretation that the studied genotypes are subject to strong allogamy.

Recent studies with *C. canephora* in Rondônia and Espírito Santo have documented wide genetic diversity and gains with the combination of morphophysiological metrics and markers (SSR/SNP), reinforcing that the organization in divergent subsets is recurrent and useful for pre-improvement. In parallel, the combined application of STRUCTURE (with Evanno's ΔK) and DAPC is recommended for admixture scenarios, exactly as observed here. In terms of interpretation, the updated classification of F_{ST} bands supports the reading that the differentiation found is not trivial and should be explored to assemble heterogeneous groups and reduce co-ancestry in production blocks [23,52].

The high genetic diversity observed in *C. canephora* genotypes is relevant data for the selection process of morphological and genetic characteristics conducted by human action [59]. Thus, conservation and genetic management should be incorporated into breeding programs as essential strategies to ensure the adaptation and evolution of cultivars in different cultivation areas [60]. The adherence of the clusters to the Hardy–Weinberg model (Table 5) indicates allelic stability and possible adaptation of genotypes to the specific

environmental conditions of the regions of origin [61]. However, the lack of adherence of some loci to the model may be related to the presence of active evolutionary forces, such as natural selection, or the influence of the reproductive system [62].

By integrating these findings with the SDGs and climate adaptation of the Amazon, the genetic structure detected in "robust Amazonian" supports the recommendation of clones by environments and the design of crosses for key attributes (nutritional efficiency, thermal and water tolerance, and drink quality) under regional technological intensification scenarios. Sectoral and technical evidence highlights that the intraspecific diversity and population organization of *C. canephora* materials have translated into productivity, quality, and resilience in Rondônia, providing the genetic basis for advancing from SSRs to SNP/GS platforms when the objective is complex genomic prediction [63]. In this sense, the results presented here, with 10 informative SSRs, F_{ST} ~ 0.16, and 23 private alleles, constitute a strategic core to increase variability, consolidate elite lines, and accelerate the delivery of cultivars/clones adapted to Amazonian edaphoclimatic gradients.

The genetic characterization of genotypes is essential for supporting breeding programs for the species, particularly in regions with specific edaphoclimatic conditions, such as the Amazon. Integrating ex situ and on-farm conservation strategies, with the application of molecular tools, should continue to be encouraged to ensure the sustainability of coffee production in the Amazon region, thereby promoting advances both in the conservation of genetic variability and in the generation of superior cultivars. Thus, the evidence sets the "robust amazonian" as a strategic core for modern programs of *C. canephora*, able to reconcile productivity, stability, and quality in a data-driven improvement pipeline aligned to future goals of sustainability and climate adaptation.

5. Conclusions

This study showed that the evaluated genotypes of *C. canephora* have high genetic diversity, with significant distribution of this variability within the identified groups. The analyses showed the formation of four genetically distinct groups, corroborated by the observed heterozygosity parameters superior to the expected heterozygosity and negative inbreeding coefficients. These findings reinforce the occurrence of allogamy and indicate the absence of endogamous crosses, in addition to revealing an excess of heterozygotes in the evaluated population. The analysis of molecular variance (AMOVA) and the estimates of F_{ST} pair to pair confirmed the existence of genetic structure between the clusters, with the gene flow being influenced by the geographic origin of the genotypes and possibly limited by the spatial distance between accesses.

The genetic arrangement provides a map for the formation of heterotic groups, the design of crossings between more distant subsets to maximize useful recombination and heterosis, and the consolidation of specific attributes by crossings between closer subsets, with direct gains in selection efficiency under heterogeneous tropical environments.

The high genetic diversity observed supports the identification of heterotic groups and potential parents for directed crossbreeding, contributing to the development of more productive, adapted, and resistant cultivars. The results support immediate developments, such as the organization of diallels and crossing blocks guided by the population structure, with control of co-ancestry; the directed incorporation of private alleles to expand the adaptive base (water/thermal tolerance, nutrient efficiency, and health), mitigating performance risks under G×E in the Amazon; and the progressive integration of molecular platforms to accelerate gains in complex characters.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/horticulturae12020250/s1; Table S1. Binary data of 10 SSR markers in 43 Genotypes.

Author Contributions: Conceptualization, S.H.F.F., C.d.S.B., S.L.F.R., and M.T.G.L.; methodology, S.H.F.F., S.L.F.R., R.L., M.C.E., and R.R.M.; software, S.L.F.R., C.H.S.G.M., T.S.R., and M.J.M.; validation, M.T.G.L., R.L., M.C.E., and S.H.F.F.; formal analysis, S.L.F.R.; investigation, C.d.S.B., M.J.M., M.C.E., R.L., and M.T.G.L.; resources, T.S.R., C.H.S.G.M., and M.T.G.L.; data curation, S.H.F.F., S.L.F.R., and R.R.M.; writing—original draft preparation, S.H.F.F., C.d.S.B., S.L.F.R., R.L., M.J.M., and M.T.G.L.; writing—review and editing, C.d.S.B., R.L., M.C.E., C.H.S.G.M., T.S.R., R.R.M., and M.T.G.L.; supervision, C.H.S.G.M. and M.T.G.L.; project administration, S.L.F.R., M.C.E., and M.T.G.L.; funding acquisition, C.H.S.G.M. and M.T.G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by Paraiba State University, grant #01/2026; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES): Finance Code 001; and FAPEAM-Edital No. 004/2018—"Amazonas Estratégico" for research funding.

Data Availability Statement: The original contributions presented in this study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Acknowledgments: Shayenne Hevelyn Farias Fernandes was supported by a scholarship from CNPq. The authors wish to thank CNPq for providing the research productivity grant that enabled the following to complete this research: Carlos Henrique Salvino Gadelha Meneses (Process no. 306943/2025-5); Santiago Linorio Ferreyra Ramos (Process no. 305280/2022-8); Ricardo Lopes (Process no. 308815/2023-8); and Maria Teresa Gomes Lopes (Process no. 306943/2025-5).

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

Abbreviations

The following abbreviations are used in this manuscript:

UFAM	Universidade Federal do Amazonas
UEPB	Universidade Estadual da Paraíba
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
FAPEAM	Fundação de Amparo à Pesquisa do Estado do Amazonas
IBGE	Instituto Brasileiro de Geografia e Estatística
SIDRA	Sistema IBGE de Recuperação Automática
SSR	Simple sequence repeat
SNP	Single-nucleotide polymorphism
DNA	Deoxyribonucleic acid
PCR	Polymerase chain reaction
dNTP	Deoxynucleotide triphosphate
CTAB	Cationic hexadecyl trimethyl ammonium bromide
TBE	Tris-Borate-EDTA buffer
ROXR	ROX™ Size Standard (Applied Biosystems)
ABI	Applied biosystems
HWE	Hardy–Weinberg equilibrium
LD	Linkage disequilibrium
AMOVA	Analysis of molecular variance
MCMC	Markov Chain Monte Carlo

H_O	Observed heterozygosity
H_E	Expected heterozygosity
f/F_{IS}	Inbreeding coefficient
F_{ST}	Fixation index
F_{IT}	Total inbreeding coefficient
DAPC	Discriminant analysis of principal components
BIC	Bayesian information criterion
K	Number of genetic clusters in the software STRUCTURE
A	Mean number of alleles per locus
AT	Total number of alleles
AR	Allelic richness
AP	Private alleles
LI	Linkage imbalance
R	R statistical environment
GDA	Genetic data analysis software
CLUMPP	Cluster matching and permutation program
Poppr	R package for clonal and sexual populations
Ade4	R package for ecology and genetics
ΦCP	Population differentiation (among clusters)
ΦCAP	Sample differentiation within populations (among genotypes within clusters)
ΦCA	Total genetic differentiation (among genotypes)
BRS	Cultivar released by Embrapa (Brazilian Agricultural Research Corporation)

References

1. Freitas, V.V.; Borges, L.L.R.; Vidigal, M.C.T.R.; Santos, M.H.; Stringheta, P.C. Coffee: A comprehensive overview of origin, market, and the quality process. *Trends Food Sci. Technol.* **2024**, *146*, 104411. [CrossRef]
2. Adunola, P.; Ferrão, M.A.G.; Ferrão, R.G.; Da Fonseca, A.F.A.; Volpi, P.S.; Comério, M.; Verdin Filho, A.C.; Munoz, P.R.; Ferrão, L.F.V. Genomic selection for genotype performance and environmental stability in *Coffea canephora*. *G3 Genes Genomes Genet.* **2023**, *13*, jkad062. [CrossRef]
3. Davis, A.P.; Rakotonasolo, F. Six new species of coffee (*Coffea*) from northern Madagascar. *Kew Bull.* **2021**, *76*, 497–511. [CrossRef]
4. Mihai, R.A.; Ortiz-Pillajo, D.C.; Iturralde-Proaño, K.M.; Vinueza-Pullotasig, M.Y.; Sisa-Tolagasí, L.A.; Villares-Ledesma, M.L.; Catana, R.D. Comprehensive assessment of coffee varieties (*Coffea arabica* L.; *Coffea canephora* L.) from Coastal, Andean, and Amazonian Regions of Ecuador: A holistic evaluation of metabolism, antioxidant capacity and sensory attributes. *Horticulturae* **2024**, *10*, 200. [CrossRef]
5. Martins, S.C.V.; Ramalho, J.D.C. Ecophysiology of coffee growth and production in a context of climate changes. *Adv. Bot. Res.* **2025**, *114*, 97–139. [CrossRef]
6. Partelli, F.L.; Vieira, H.D. Coffee Breeding and Stress Biology. *Plants* **2024**, *13*, 1912. [CrossRef] [PubMed]
7. Salvador, H.P.; Berilli, A.P.C.G.; Rodrigues, W.P.; Mazzafera, P.; Partelli, F.L. A climate change perspective on the selection, development, and management of *Coffea canephora* genotypes. *Adv. Bot. Res.* **2025**, *114*, 283–315. [CrossRef]
8. Acre, L.B.; Viencz, T.; Francisco, J.S.; Rocha, R.B.; Alves, E.A.; Benassi, M.T. Composition of *Coffea canephora* Varieties from the Western Amazon. *J. Braz. Chem. Soc.* **2024**, *35*, e20240031. [CrossRef]
9. Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira de café- safra 2024- quarto levantamento- CONAB. Available online: http://www.conab.gov.br (accessed on 16 July 2025).
10. Ferrão, M.A.G.; Da Fonseca, A.F.A.; Volpi, P.S.; de Souza, L.C.; Comério, M.; Verdin Filho, A.C.; Riva-Souza, E.M.; Munoz, P.R.; Ferrão, R.G.; Ferrão, L.F.V. Genomic-assisted breeding for climate-smart coffee. *Plant Genome* **2023**, *17*, e20321. [CrossRef] [PubMed]
11. Teixeira, A.L.; Rocha, R.B.; Espindula, M.C.; Ramalho, A.R.; Vieira, J.R.; Alves, E.A.; Fernandes, C.D.F. Amazonian Robustas—New *Coffea canephora* Coffee Cultivars for the Western Brazilian Amazon. *Crop Breed. Appl. Biotechnol.* **2020**, *20*, e323420318. [CrossRef]
12. Instituto Brasileiro de Geografia e Estatística (IBGE). Sistema IBGE de Recuperação Automática (SIDRA): Produção Agrícola Municipal. 2023. Available online: https://sidra.ibge.gov.br/tabela/1613 (accessed on 15 February 2025).
13. Ferrão, M.A.G.; de Mendonça, R.F.; Fonseca, A.F.A.; Ferrão, R.G.; Senra, J.F.B.; Volpi, P.S.; Verdin Filho, A.C.; Comério, M. Characterization and Genetic Diversity of *Coffea canephora* Accessions in a Germplasm Bank in Espírito Santo, Brazil. *Crop. Breed. Appl. Biotechnol.* **2021**, *21*, e36132123. [CrossRef]

14. Espindula, M.C.; Dalazen, J.R.; Rocha, R.B.; Teixeira, A.L.; Diocleciano, J.M.; Dias, J.R.M.; Schmidt, R.; Lima, P.P.D.; Lima, G.M.; Gama, W. *Robustas Amazônicos: Os Cafeeiros Cultivados em Rondônia*; Embrapa: Brasília, Brazil, 2022; p. 144.
15. Moraes, M.S.; Rocha, R.B.; Teixeira, A.L.; Espindula, M.C.; Silva, C.A.; Lunz, A.M.P. Adaptability and stability of *Coffea canephora* Pierre ex Froehner genotypes in the Western Amazon. *Cienc. Rural* **2020**, *50*, e20190087. [CrossRef]
16. Salgotra, R.K.; Chauhan, B.S. Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. *Genes* **2023**, *14*, 174. [CrossRef]
17. Gan, S.T.; Teo, C.J.; Manirasa, S.; Wong, W.C.; Wong, C.K. Assessment of Genetic Diversity and Population Structure of Oil Palm (*Elaeis guineensis* Jacq.) Field Genebank: A Step towards Molecular-Assisted Germplasm Conservation. *PLoS ONE* **2021**, *16*, e0255418. [CrossRef] [PubMed]
18. Sharma, C.K.; Gupta, A.; Sharma, M. Molecular marker: Genetic improvement and conservation of industrial crops. In *Industrial Crop Plants. Interdisciplinary Biotechnological Advances*; Kumar, N., Ed.; Springer: Singapore, 2024; pp. 101–122. [CrossRef]
19. Kanaka, K.K.; Sukhija, N.; Goli, R.C.; Singh, S.; Ganguly, I.; Dixit, S.P.; Dash, A.; Malik, A.A. On the concepts and measures of diversity in the genomics era. *Curr. Plant Biol.* **2023**, *33*, 100278. [CrossRef]
20. Jones, M.R.; Winkler, D.E.; Massatti, R. Demographic modeling informs functional connectivity and management interventions in Graham's beardtongue. *Conserv. Genet.* **2021**, *22*, 993–1003. [CrossRef]
21. Bidyananda, N.; Jamir, I.; Nowakowska, K.; Varte, V.; Vendrame, W.A.; Devi, R.S.; Nongdam, P. Plant Genetic Diversity Studies: Insights from DNA Marker Analyses. *Int. J. Plant Biol.* **2024**, *15*, 607–640. [CrossRef]
22. Dagnon, Y.D.; Palanga, K.K.; Bammite, D.; Bodian, A.; Akabassi, G.C.; Foncéka, A.; Tozo, K. Genetic Diversity and Population Structure of Cowpea (*Vigna unguiculata* (L.) Walp.) Accessions from Togo Using SSR Markers. *PLoS ONE* **2022**, *17*, e0252362. [CrossRef]
23. Al-Yasi, H.M.; Al-Qthanin, R. Comparing Genetic Differentiation and Variation Using ISSR and SCoT among Juniper Plant Markers in Saudi Arabia. *Front. Plant Sci.* **2024**, *15*, 1356917. [CrossRef]
24. Ramos, S.L.F.; Dequigiovanni, G.; Lopes, M.T.G.; Veasey, E.A.; de Macedo, J.L.V.; Batista, J.S.; Formiga, K.M.; Kageyama, P.Y. Microsatellite Markers for *Euterpe precatoria* Mart. (Arecaceae), a Palm Species Used by Extractive Traditional Farmers of the Amazon. *Conserv. Genet. Resour.* **2016**, *8*, 43–81.
25. Dequigiovanni, G.; Ramos, S.L.F.; Lopes, M.T.G.; Clement, C.R.; Rodrigues, D.P.; Fabri, E.G.; Zucchi, M.I.; Veasey, E.A. New Microsatellite Loci for Annatto (*Bixa orellana*), a Source of Natural Dyes from Brazilian Amazonia. *Crop Breed. Appl. Biotechnol.* **2018**, *18*, 116–122. [CrossRef]
26. Ramos, S.L.F.; Dequigiovanni, G.; Lopes, M.T.G.; Aguiar, A.V.; Lopes, R.; Veasey, E.A.; Macêdo, J.L.V.; Alves-Pereira, A.; Fraxe, T.J.P.; Wrege, M.S.; et al. Genetic Structure in Populations of *Euterpe precatoria* Mart. in the Brazilian Amazon. *Front. Ecol. Evol.* **2021**, *8*, 603448. [CrossRef]
27. Doyle, J.J.; Doyle, J.L. Isolation of Plant DNA from Fresh Tissue. *Focus* **1990**, *12*, 13–15.
28. Hendre, P.S.; Aggarwal, R.K. Development of genic and genomic SSR markers of robusta coffee (*Coffea canephora* Pierre Ex A. Froehner). *PloS ONE* **2014**, *9*, e113661. [CrossRef]
29. Schuelke, M. An Economic Method for the Fluorescent Labeling of PCR Fragments. *Nat. Biotechnol.* **2000**, *18*, 233–234. [CrossRef]
30. Pritchard, J.K.; Stephens, M.; Donnelly, P.J. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [CrossRef] [PubMed]
31. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef] [PubMed]
32. Behr, A.A.; Liu, K.Z.; Liu-Fang, G.; Nakka, P.; Ramachandran, S. Pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **2016**, *32*, 2817–2823. [CrossRef]
33. González-Toral, C.; Cuesta, C.; Cires, E. Genetic Diversity and Population Structure Assessed Using Microsatellite (SSR) Markers from Relict Populations of *Nuphar pumila* (Nymphaeaceae). *Plants* **2023**, *12*, 1771. [CrossRef]
34. Thia, J.A. Guidelines for standardizing the application of discriminant analysis of principal components to genotype data. *Mol. Ecol. Resour.* **2023**, *23*, 523–538. [CrossRef] [PubMed]
35. Depecker, J.; Verleysen, L.; Asimonyio, J.A.; Hatangi, Y.; Kambale, J.L.; Mwanga Mwanga, I.; Ebele, T.; Dhed'a, B.; Bawin, Y.; Staelens, A.; et al. Genetic Diversity and Structure in Wild Robusta Coffee (*Coffea canephora* A. Froehner) Populations in Yangambi (DR Congo) and Their Relation to Forest Disturbance. *Heredity* **2023**, *130*, 145–153. [CrossRef]
36. Jombart, T.; Ahmed, I. Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data. *Bioinformatics* **2011**, *27*, 3070–3071. [CrossRef]
37. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023; Available online: https://www.R-project.org/ (accessed on 15 August 2025).

38. Wright, S. The Genetical Structure of Populations. *Ann. Eugen.* **1951**, *15*, 323–354. [CrossRef]
39. Weir, B.S.; Cockerham, C.C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **1984**, *38*, 1358–1370. [CrossRef]
40. Goudet, J. HIERFSTAT: A Package for R to Compute and Test Hierarchical F-Statistics. *Mol. Ecol. Notes* **2005**, *5*, 184–186. [CrossRef]
41. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. Poppr: An R Package for Genetic Analysis of Populations with Clonal, Partially Clonal, and/or Sexual Reproduction. *PeerJ* **2014**, *2*, e281. [CrossRef]
42. Dray, S.; Dufour, A.B.; Thioulouse, J. *ade4: Analysis of Ecological Data*, R Package Version 1.7-26; 2024. Available online: https://cran.r-project.org/package=ade4 (accessed on 16 July 2025).
43. Rice, W.R. Analyzing Tables of Statistical Tests. *Evolution* **1989**, *43*, 223–225. [CrossRef] [PubMed]
44. Wright, S. *Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations*; University of Chicago Press: Chicago, IL, USA, 1978; pp. 1–580.
45. Santos, M.M.; da Silva, C.A.; Oza, E.F.; Gontijo, I.; do Amaral, J.F.T.; Partelli, F.L. Concentration of Nutrients in Leaves, Flowers, and Fruits of Genotypes of *Coffea canephora*. *Plants* **2021**, *10*, 2661. [CrossRef]
46. Silva, M.O.; Honfoga, J.N.B.; Medeiros, L.L.; Madruga, M.S.; Bezerra, T.K.A. Obtaining Bioactive Compounds from the Coffee Husk (*Coffea arabica* L.) Using Different Extraction Methods. *Molecules* **2021**, *26*, 46. [CrossRef] [PubMed]
47. Rodrigues, M.J.L.; da Silva, C.A.; Braun, H.; Partelli, F.L. Nutritional Balance and Genetic Diversity of *Coffea canephora* Genotypes. *Plants* **2023**, *12*, 1451. [CrossRef] [PubMed]
48. Bezerra, C.d.S.; Tomaz, J.S.; Valente, M.S.F.; Espindula, M.C.; Marques, R.L.S.; Tadeu, H.C.; Ferreira, F.M.; Silva, G.d.S.; Meneses, C.H.S.G.; Lopes, M.T.G. Phenotypic Diversity and Genetic Parameters of *Coffea canephora* Clones. *Plants* **2023**, *12*, 4052. [CrossRef]
49. 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. [CrossRef]
50. Li, F.; Fang, H.; Zhou, J.; Hu, S.; Cao, F.; Guo, Q. Genomics reveal population structure, genetic diversity and evolutionary history of *Phyllostachys edulis* (moso bamboo) in global natural distribution. *Front. Plant Sci.* **2025**, *16*, 1532058. [CrossRef] [PubMed]
51. Wu, B.; Wen, J.; Lu, R.; Zhou, W. Genetic Diversity, Population Structure, and Phylogenetic Relationships of a Widespread East Asia Herb, *Cryptotaenia japonica* Hassk. (Apiaceae) Based on Genomic SNP Data Generated by dd-RAD Sequencing. *Front. Genet.* **2024**, *15*, 1368760. [CrossRef] [PubMed]
52. Dossa, A.F.; Tchokponhoué, D.A.; Houdegbe, A.C.; Achigan-Dako, E.G. SNP markers revealed the genetic diversity and population structure of *Mesosphaerum suaveolens* (L.) Kuntze Syn. *Hyptis suaveolens* (L.) Poit. accessions collected in Benin. *PLoS ONE* **2025**, *20*, e0331702. [CrossRef]
53. Kasule, F.; Alladassi, B.M.E.; Aru, C.J.; Adikini, S.; Biruma, M.; Ugen, M.A.; Kakeeto, R.; Esuma, W. Genetic diversity, population structure, and a genome-wide association study of sorghum lines assembled for breeding in Uganda. *Front. Plant Sci.* **2024**, *15*, 1458179. [CrossRef]
54. Silva, A.N.R.; Rocha, R.B.; Teixeira, A.L.; Espindula, M.C.; Partelli, F.L.; Caixeta, E.T. Self-Incompatibility and Pollination Efficiency in *Coffea canephora* Using Fluorescence Microscopy. *Agronomy* **2024**, *14*, 1564. [CrossRef]
55. Sousa, P.; Vieira, H.; Santos, E.; Viana, A.; Partelli, F. Genetic diversity in *Coffea canephora* genotypes via digital phenotyping. *Plants* **2025**, *14*, 2814. [CrossRef]
56. Vanden Abeele, S.; Janssens, S.B.; Asimonyio Anio, J.; Bawin, Y.; Depecker, J.; Kambale, B.; Mwanga Mwanga, I.; Ebele, T.; Ntore, S.; Stoffelen, P.; et al. Genetic Diversity of Wild and Cultivated *Coffea canephora* in Northeastern DR Congo and the Implications for Conservation. *Am. J. Bot.* **2021**, *108*, 2425–2434. [CrossRef]
57. Zhou, Y.; Tan, J.; Huang, L.; Luo, Y.; Huang, S.; Ye, Y.; Xu, Y. Unveiling the Genetic Diversity and Population Structure of the Endangered Fern Angiopteris fokiensis Through Genome Survey and Genomic SSR Markers. *Biomolecules* **2025**, *15*, 1649. [CrossRef] [PubMed]
58. Peralta, W.; Nestares, A.; Gamarra, J.; Rojas, M.; Sullca, J.; Estrada, R. Genetic Diversity and Population Structure of Alpacas (Vicugna pacos) in Peru: A Microsatellite Analysis. *Diversity* **2025**, *17*, 353. [CrossRef]
59. Swarup, S.; Cargill, E.J.; Crosby, K.; Flagel, L.; Kniskern, J.; Glenn, K.C. Genetic diversity is indispensable for plant breeding to improve crops. *Crop Sci.* **2021**, *61*, 839–852. [CrossRef]
60. Pierson, J.C.; Berry, L.; Alexander, L.; Anson, J.; Birkett, M.; Kemp, L.; Pascoe, B.A.; Farquharson, K.A.; Hogg, C.J. Adaptive Genetic Management of a Reintroduction Program from Captive Breeding to Metapopulation Management of an Arboreal Marsupial. *Diversity* **2023**, *15*, 848. [CrossRef]
61. Neamatzadeh, H.; Dastgheib, S.A.; Mazaheri, M.; Masoudi, A.; Shiri, A.; Omidi, A.; Rahmani, A.; Golshan-Tafti, A.; Aghasipour, M.; Yeganegi, M.; et al. Hary-Weinberg Equilibrium in Meta-Analysis Studies and Large-Scale Genomic Sequencing Era. *Asian Pac. J. Cancer Prev.* **2024**, *25*, 2229. [CrossRef] [PubMed]

62. Burgarella, C.; Brémaud, M.F.; Hirschheydt, G.V.; Viader, V.; Ardisson, M.; Santoni, S.; Ranwez, V.; Navascués, M.; David, J.; Glémin, S. Mating systems and recombination landscape strongly shape genetic diversity and selection in wheat relatives. *Evol. Lett.* **2024**, *8*, 866–880. [CrossRef]
63. Junior, H.L.; Rocha, R.B.; Kolln, A.M.; Silva, R.N.P.; Alves, E.A.; Teixeira, A.L.; Espíndula, M.C. Genetic Variability in the Physicochemical Characteristics of Cultivated *Coffea canephora* Genotypes. *Plants* **2024**, *13*, 2780. [CrossRef] [PubMed]