

RESEARCH ARTICLE

Genetic diversity and population structure of pearl millet in the Senegalese germplasm

Aliou Ba^{1,2,3*}, Thomas L. Odong⁴, Richard Edema^{1,2}, Arfang Badji^{1,2}, Oumar Diack³, Tony Obua², Mildred Ochwo-Semakula², Paul Gibson², Ndjido. A. Kane³, Phinehas Tukamuhabwa²

1 Makerere University Regional Centre for Crop Improvement (MaRCCI), College of Agricultural and Environmental Sciences, Makerere University, Kampala, Uganda, **2** Department of Crop Science and Horticulture, College of Agricultural and Environmental Sciences, Makerere University, Kampala, Uganda, **3** Institut Sénégalais de Recherches Agricoles (ISRA), Dakar, Sénégal, **4** Department of Soil Science and Land Use Management, College of Agricultural and Environmental Sciences, Makerere University, Kampala, Uganda

* baaliou52@gmail.com



OPEN ACCESS

Citation: Ba A, Odong TL, Edema R, Badji A, Diack O, Obua T, et al. (2026) Genetic diversity and population structure of pearl millet in the Senegalese germplasm. PLoS One 21(2): e0343497. <https://doi.org/10.1371/journal.pone.0343497>

Editor: Somashekhar Mallikarjun Punhuri, Fort Valley State University, UNITED STATES OF AMERICA

Received: June 14, 2025

Accepted: February 6, 2026

Published: February 25, 2026

Copyright: © 2026 Ba et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the paper and its [Supporting information](#) files.

Funding: RSIF-PASET Scholarship.

Abstract

Pearl millet (*Pennisetum glaucum*), a key cereal for food security in arid and semiarid regions, combines drought tolerance, high nutritional value, and adaptability to marginal environments. In Senegal, numerous landraces have been collected to develop breeding populations. However, genetic diversity and population structure of this germplasm remain poorly understood. This study characterized the genetic diversity and population structure of 169 genotypes, including 15 PMiGAP inbreds lines, 58 F₅ ISMI lines, 94 F₃ PLS lines, and 2 introgressions (29AW and ICMR088888). Genomic DNA was extracted using the CTAB protocol and genotyping was conducted on a DArTseq platform. Following quality control, 16,693 SNPs were retained for downstream analyses. Population structure was inferred using the sparse nonnegative matrix factorization (sNMF) algorithm, and phylogenetic relationships were reconstructed using the Neighbor-Joining method. The mean gene diversity was 0.30, and the average minor allele frequency was 0.21. Four genetic clusters were identified: cluster 1 comprised mainly F₃ lines (88%), while cluster 2 included 56% of F₅ ISMI lines. Nei genetic distances ranged from 0.023 (between clusters 2 and 4) to 0.078 (between clusters 3 and 2). Analysis of molecular variance revealed 92.1% of variation within clusters and 7.9% among clusters. Principal component analysis and fixation index (F_{ST}) results were consistent with the structure and phylogenetic analyses. The moderate genetic diversity and clear population structure observed suggest great potential for defining heterotic groups and developing resilient hybrid varieties as well as better conservation strategies.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Pearl millet exhibits unique agronomic advantages among cereals, producing nutrient-dense grains while maintaining low water and energy footprints, and showing resilience in some of the most extreme and resource-limited environments worldwide [1]. In Senegal, it is cultivated mainly in the central, southern, and southeastern parts; these regions accounted for more than 80% of the country's total production in 2021 [2]. Between 2019 and 2022 pearl millet had the second largest harvested area in the country among cereals, followed by sorghum and maize. In both rural and urban areas, pearl millet grains are used into a wide range of nutritious meals, such as porridges, doughs, couscous, and bread [2], the crop is also used as forage. However, average grain yields in Senegal remain very low and unstable, rarely exceeding 1000 kg ha⁻¹, compared with approximately 1400 kg ha⁻¹ in countries like India. The main constraints include the lack of high-performing varieties (both hybrids and open-pollinated), along with multiple biotic and abiotic stresses such as *Striga hermonthica* infestation, low soil fertility, drought, and insect pests [3]. Pearl millet shows broad diversity reservoir, resulting from its allogamous nature and exceptional adaptability to diverse ecological conditions, particularly in drought-prone environments [4]. Earlier studies of the Senegalese pearl millet germplasm revealed high genetic diversity in a collection consisting of 477 accessions, including 353 early-flowering (Souna), 112 late-flowering (Sanio) and 12 improved early flowering varieties [5]. The Senegalese landrace collection consistently presented little population structure and greater diversity than other regions of Africa and Asia; a plausible reason for this being the proximity of the center of domestication of the crop [5]. Moreover, Senegal is considered a pearl millet diversity hotspot, partly due to the high extent of gene flow with local wild populations [6].

In an improvement effort, Senegalese landraces have been used as parents to generate populations of advanced lines representing a useful genetic reservoir for developing new varieties with relatively high yield potential and other desirable agronomic traits (yield, panicle length, flowering time, and striga resistance) [7]. These efforts were coupled with an increased use of molecular tools to gain insights on genetic diversity and structure from local collections [2]. Inferring the molecular diversity and population structure of crop germplasms is therefore a prerequisite for devising efficient selection methods as well as developing conservation and valorization strategies [8]. Markers such as amplified fragment length polymorphisms (AFLPs), restriction fragment length polymorphisms (RFLPs), and simple sequence repeats (SSRs) have been extensively used to provide crucial information on plant genetic diversity [9]. Recently, single nucleotide polymorphisms (SNPs) have become the markers of choice for genetic diversity studies owing to several of their features: they are abundant, feasible for automated high-throughput genotyping, highly reproducible, and increasingly cost effective [10]. These markers allow studies on genetic diversity and population structure to provide clear insights into how selection could be more efficiently carried out. Genetic diversity is the backbone of any breeding program; therefore, determining the variability within germplasms is an indispensable step before engaging in any crop improvement scheme [11]. Multivariate analyses,

such as principal component analysis (PCA), play a crucial role in plant breeding by enabling researchers to explore and interpret complex, high-dimensional datasets effectively. These techniques allow breeders to identify patterns, relationships, and key traits that contribute most to genetic diversity or agronomic performance among genotypes. They also enable breeders to efficiently select genetic variation and develop strategies to incorporate useful diversity in their breeding programs [12].

By reducing the dimensionality of data without significant loss of information, PCA helps visualize genetic variation as well as structuring patterns of different genotypes. Furthermore, insights on population structure can help plant breeders understand the distribution of genetic variation within and between populations. This information is essential for selecting diverse parental lines and minimizing inbreeding in breeding programs. Moreover, through phylogenetic studies, subpopulations constituting heterotic groups [13] with specific adaptive traits can be identified, which can be targeted for improvement or conservation.

Substantial genetic variation was expected both within and among Senegalese pearl millet populations, with population structure expected to reflect breeding history or geographic origins of the studied material. The Senegalese pearl millet breeding program has developed several populations composed of numerous advanced lines (F_3 and F_5) and a few introgressions from international sources [3]; however, the genetic diversity and population structure of these materials remain insufficiently characterized, limiting their optimal use in breeding. To address this gap, the present study assessed the genetic diversity and population structure of Senegalese pearl millet germplasm consisting of F_5 lines (ISRAMIL: ISMI), a collection F_3 lines derived from Senegalese landraces (PLS), a pearl millet inbred germplasm association panel (PMiGAP) [4], and two genotypes (29AW and ICMR088888) sourced from Niger and India respectively.

Materials and methods

Ethics approval and consent to participate

All experimental plant materials were obtained from the ex-situ collection of the Senegalese Agricultural Research Institute (ISRA). The materials used in this study were owned by governmental institutions established for research and were freely available for noncommercial purposes. Data exchange was performed under institutional, national and international plant import/export guidelines.

Plant material

One hundred sixty-nine (169) genotypes obtained from the Senegalese pearl millet germplasm were used in this study. The panel consisted of 94 F_3 lines (PLS) and 58 F_5 lines (ISMI) obtained from several crosses between landraces (S1 Table), 15 Pearl Millet Inbred Germplasm Association Panel (PMiGAP) lines (including cultivated germplasms from Africa and Asia, elite improved open-pollinated cultivars, hybrid parental inbreds and inbred mapping population parents), and 2 introgressions (29 AW and ICMR088888).

Genomic DNA extraction and SNP calling

Planting was done in February 2023 at Bambey (Senegal), and 3-week-old leaves were sampled and placed into 96-well collection plates. Sampling was performed using leafcutters that were always sterilized with 70% alcohol before the next genotype was cut to prevent cross-contamination. Genotyping was performed at Intertek AgriTech (<https://www.intertek.com/agriculture/agritech/>), using the DArTseq™ technology developed by Diversity Arrays Technology Pty Ltd (Canberra, Australia). The DArTseq protocol involves genome complexity reduction using a combination of restriction enzymes (*Pst*I and *Mse*I), ligation of Illumina-compatible adapters carrying unique barcodes, and limited-cycle PCR amplification to generate sequencing libraries. Libraries were pooled and sequenced as single-end reads on an Illumina platform. Sequence data were processed using the proprietary DArTsoft14 pipeline for SNP and silicoDArT marker calling [14].

SNP calling and data filtering

An initial set of 58627 raw SNP markers was obtained from the variant calling pipeline. Imputations were then carried out using the LD KNNi method implemented in TASSEL program until reaching a plateau on missing markers. This decision is informed by Roshyara *et al.*, [15] who reported that little or no SNP filtering prior to imputation appears to be the best strategy for imputing small to moderately sized datasets. Thereafter, quality control was performed, and SNP markers with minor allele frequency (MAF) > 5% and missing rate per sample < 20% and per SNP < 10% were retained for downstream analysis. SNPs with missing data were again imputed using the LD KNNi imputation method, and scaffolds and unmapped SNPs were removed. To avoid redundancy, SNP markers were further pruned based on linkage disequilibrium (LD) with a threshold of 0.5 using PLINK. Finally, structure and diversity analyses were performed on 169 samples using a set of 16693 quality SNP markers.

Data analyses

All analyses were conducted using R statistical software (version 4.4.2). Genetic parameters such as polymorphism information content (PIC), gene diversity (GD), observed (H_o) and expected (H_e) heterozygosity, minor allele frequency (MAF), fixation index (F_{ST}) and genetic distance were computed using the *snprReady* package [16]. Genetic distances between clusters were computed using *adeigenet* package using the Nei method [17]. The analysis of molecular variance (AMOVA) was performed using *poppr* package [18].

To explore the genetic structure of the 169 pearl millet germplasms, the sNMF (sparse nonnegative matrix factorization) statistical method we used to infer population structure and ancestry proportions. This method estimates ancestry proportions and infers the number of ancestral populations from genotype data. We specified a range of 1–10 potential populations (K), and the program used cross-validation to determine the optimal number with the lowest cross-entropy value. The analysis was performed with the following parameters after an optimization process: alpha = 1500, number of iterations = 10000, number of repetitions = 100, tolerance = 0.00001, and percentage of masked data for cross validation = 0.05. The matrices of ancestry proportions were visualized through a bar plot showing different colors, each corresponding to a distinct K. Principal component analysis (PCA) was conducted using PLINK program to summarize the contributions of each component to the variation that existed in the population. PCA plot was visualized using *ggplot2* package and a Neighbor-Joining phylogenetic tree was constructed *ape* package.

Results

SNP variation and marker distribution

The distribution of the 16,693 SNP markers was plotted in a Mb (megabase pair) window across the *P. glaucum* genome (Fig 1). Variant distribution was fairly uniform across the chromosomes. An average of 44 SNPs per Mb window were detected. The highest density (>88 SNPs/Mb) was observed in the telomeric regions of Chr1, Chr3, Chr6 and Chr7. The lowest average SNP density was found in the centromeric regions of Chr1, Chr2, and Chr3 and in the para-telomeric regions of Chr5 and Chr6.

Population structure

The genetic structure of the overall population was analyzed using SNP markers with the sNMF package, which applies a sparse non-negative matrix factorization algorithm to estimate ancestry coefficients for each individual [19]. Cross-entropy values were used to determine the most likely number of genetic clusters (K). The lowest cross-entropy value corresponded to K = 4, indicating that the population could be divided into four distinct subpopulations (Fig 2). The analysis revealed clear genetic stratification among the individuals. Each individual was assigned to the cluster where it showed the highest ancestry coefficient from the Q-matrix. The composition of these subpopulations is summarized in (Table 1), each cluster is represented by a distinct color (Fig 3). Cluster 1 is characterized by a high proportion of ancestry from

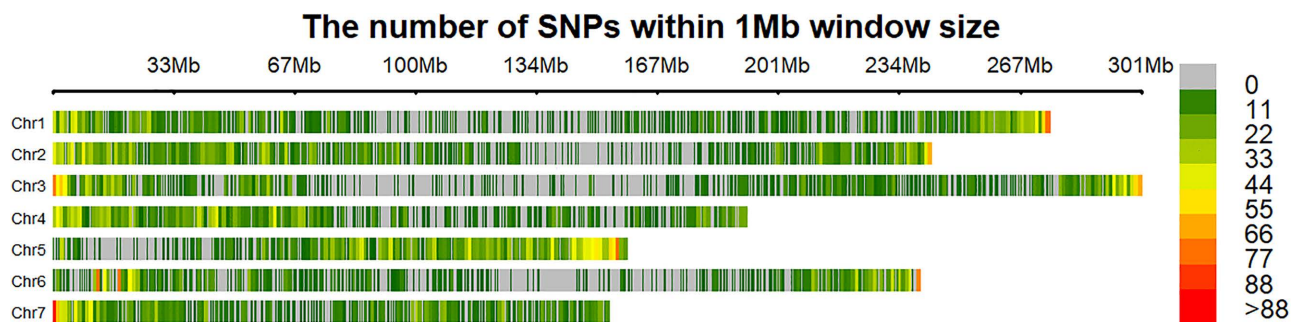


Fig 1. Distribution of SNP markers within 1 Mb window size across seven chromosomes. Colored bars are SNP counts in 1 Mb interval.

<https://doi.org/10.1371/journal.pone.0343497.g001>

a single genetic component (purple). Cluster 2 is dominated by another genetic component (blue). Cluster 3 and 4 are represented by other genetic components, red and yellow respectively. Most individuals in clusters 1, 2, and 4 exhibited mixed ancestry (multiple colors on the bar plot), while cluster 3 displayed the lowest degree of admixture (Fig 3).

Genetic relatedness and cluster composition

Genetic relatedness among individuals was inferred from the clusters identified through population structure analysis. Cluster 2 comprised the largest number of genotypes ($n=85$), followed by Cluster 1 ($n=62$), Cluster 3 ($n=13$), and Cluster 4 ($n=9$) (Table 1). Cluster 1 was mainly composed of PLS genotypes ($n=55$) and ISMI genotypes ($n=7$). Cluster 2 was the most diverse, containing genotypes from all groups: 2 INTRO, 15 PMIGAP, 48 ISMI, and 20 PLS accessions. Cluster 3 consisted exclusively of PLS genotypes ($n=13$), whereas Cluster 4 included 3 ISMI and 6 PLS genotypes.

Phylogenetic analysis

To further understand the population structure, a phylogenetic analysis was performed using the Neighbor-Joining (NJ) method based on Nei's genetic distances computed from SNP marker data. The resulting tree displayed a clustering pattern that was largely consistent with the subpopulations identified through sNMF analysis. Three major clades were observed and they corresponded closely to the two major clusters inferred from the population structure results (Fig 4). The blue clade was composed mainly of PLS lines, closely mirroring the composition of Cluster 1 identified in the structure analysis. In contrast, the red and green clades, which included mostly ISMI lines and a smaller number of PMIGAP accessions, collectively corresponded to Cluster 2. Cluster 3 formed a distinct and well-defined clade, showing minimal evidence of admixture, consistent with the low levels of genetic mixing observed in the structure analysis. Furthermore, individuals assigned to Clusters 2 and 4 were more widely dispersed across the phylogenetic tree, reflecting their higher levels of admixture. Cluster 1 exhibited both compact clades and scattered individuals, further supporting the mixed ancestry pattern inferred from the structure analysis.

Genetic diversity metrics between clusters. Genetic diversity within each cluster was assessed using several parameters, including observed and expected heterozygosity (H_o and H_e), gene diversity (GD), minor allele frequency (MAF), and polymorphic information content (PIC) (Table 2). Gene diversity values ranged from 0.16 to 0.30, with the highest observed in Cluster 4, followed by Cluster 2 (0.29), Cluster 1 (0.28), and Cluster 3 (0.22). Similarly, the highest PIC value was recorded in Cluster 4 (0.24), while Cluster 3 had the lowest (0.18); Clusters 1 and 2 shared an intermediate PIC of 0.23. MAF values varied from 0.16 (Cluster 3) to 0.22 (Cluster 4), with Clusters 1 and 2 both showing 0.20. Across all clusters, observed heterozygosity (H_o) was lower than expected heterozygosity (H_e). The mean H_o and H_e across clusters were 0.10 and 0.29, respectively. The lowest H_o (0.08) was observed in Clusters 2 and 3, and H_e values of 0.28 and 0.21 respectively. Cluster 1 showed $H_o=0.11$ and $H_e=0.28$, whereas Cluster 4 exhibited the highest $H_o=0.24$ and $H_e=0.29$.

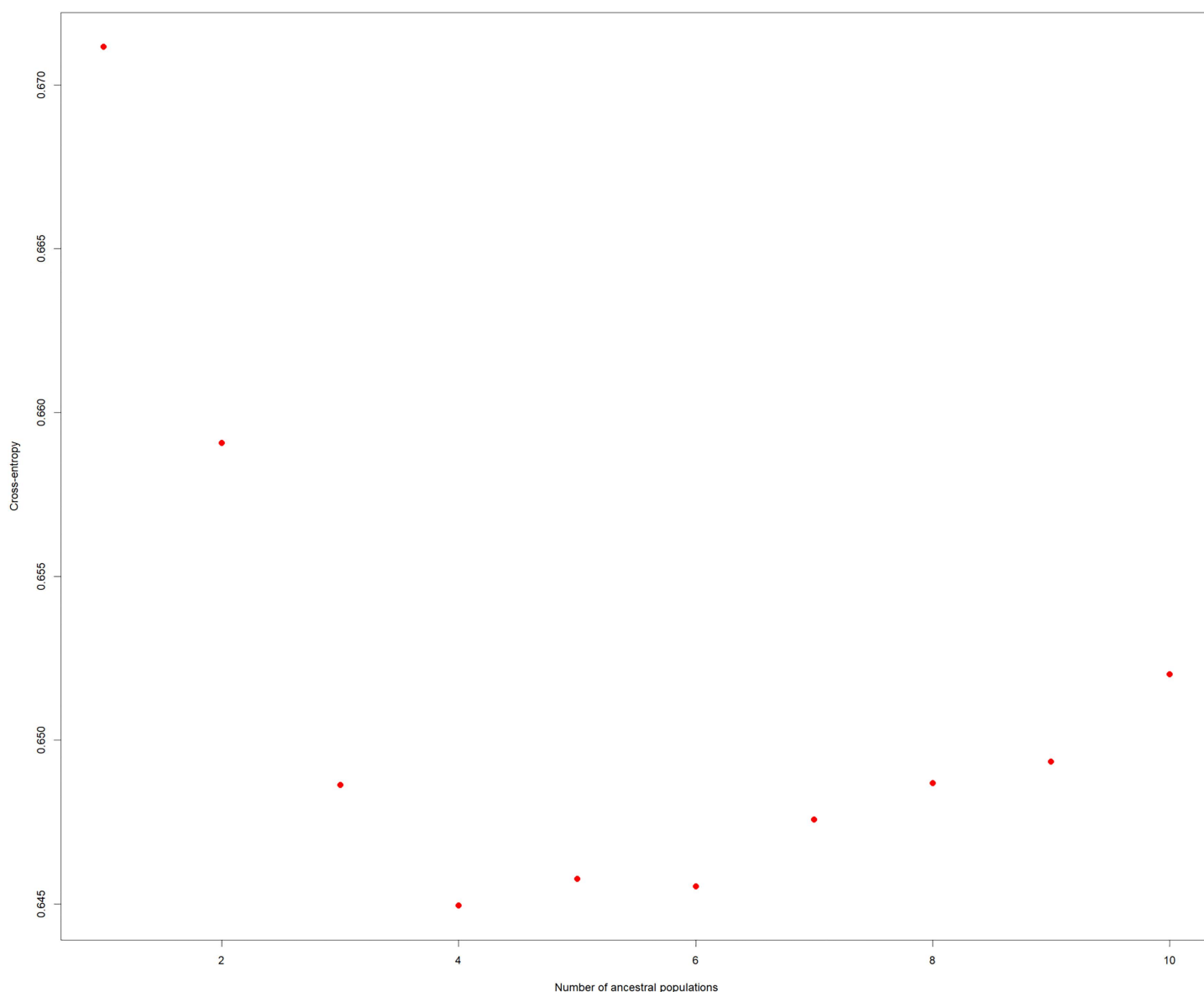


Fig 2. Number of ancestral populations and their corresponding cross entropy values.

<https://doi.org/10.1371/journal.pone.0343497.g002>

Analysis of molecular variance (AMOVA)

Analysis of molecular variance revealed that the proportion of variation within clusters was significantly greater (92.12%, $P < 0.001$) than the variation between clusters (7.88%, $P < 0.001$) ([Table 3](#)).

Genetic differentiation among clusters

Pairwise genetic fixation indices (F_{ST}) among clusters were calculated from SNP markers ([Table 4](#)). The average F_{ST} value across clusters was 0.052. The lowest differentiation was observed between Clusters 2 and 4 ($F_{ST} = 0.028$). In contrast, the highest F_{ST} was detected between Clusters 4 and 3 (0.11), followed by Clusters 3 and 1 (0.071), Clusters 2 and 3 (0.063), Clusters 4 and 1 (0.044), and Clusters 2 and 1 (0.043).

Table 1. Clustering of 169 pearl millet genotypes into four groups using SNP markers.

Group	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
INTRO	0	2	0	0	2
PMIGAP (ICRISAT)	0	15	0	0	15
ISMI (F ₅)	7	48	0	3	58
PLS (F ₃)	55	20	13	6	94
Total	62	85	13	9	169

INTRO: Introgression, PMIGAP: Pearl Millet Inbred Germplasm Association Panel, ISMI: ISRAMIL, PLS: Pearl millet landrace of Senegal.

<https://doi.org/10.1371/journal.pone.0343497.t001>

Nei's genetic distance and identity

Nei's genetic distances were calculated among the different clusters (Fig 5 and Table 5). The largest distances were observed between Clusters 2 and 3 (0.078) and between Clusters 3 and 4 (0.077), while the smallest distance was found between Clusters 2 and 4 (0.023). The highest genetic identity was recorded between Clusters 2 and 4 (0.976), whereas the lowest distance was observed between Clusters 2 and 3 (0.924).

Principal component analysis

To assess genetic variation among genotypes, a principal component analysis (PCA) was performed. The two-dimensional PCA plot revealed four distinct clusters, although the separation between Clusters 2 and 4 was less pronounced. In total, the first 20 PCs were sufficient to capture the overall variation (Fig 6). The first two principal components (PC1 and PC2) explained 13.21% and 9.30% of the total genetic variation, respectively, accounting together for 23% of the overall variation (Fig 7). The clustering pattern closely reflected the genetic structure results, with cluster 1 comprising almost all the PLS genotypes, cluster 2 comprising mostly the ISMI and PMIGAP genotypes, and cluster 3 comprising only the PLS genotypes.

Discussion

The present study provides an up-to-date assessment of genetic structure and diversity within the Senegalese pearl millet germplasm currently used in the national breeding program. Population structure was inferred prior to diversity analysis to ensure that diversity estimates accurately reflected the underlying population history [20]. Following cluster identification, the calculation of diversity parameters both within and across clusters allowed for a more meaningful evaluation of genetic variation at multiple levels, capturing patterns of differentiation as well as shared diversity among well-defined groups.

The population structure revealed substantial admixture among clusters, a pattern commonly observed in allogamous crop species like pearl millet [21]. Such admixture is typically driven by gene flow resulting from natural cross-pollination, farmer-mediated seed exchange, and intentional hybridization during variety development [22]. The structure analysis identified four sub-populations within the overall collection, a result that was consistent with the patterns observed in PCA. Genotypes were broadly admixed across clusters irrespective of geographic origin, with several accessions in clusters 1, 2, and 4 exhibiting high proportions of mixed ancestry, indicating extensive gene flow among these groups. These findings support previous reports of widespread germplasm exchange across geographic regions in pearl millet [23].

On the other hand, the clustering pattern partially reflected known pedigrees and agronomic characteristics of the accessions. Cluster 1 comprised 62 accessions, primarily originating from Senegal, including three ISMI lines (ISMI21207, ISMI21211, and ISMI21508) that shared at least one common parent. For instance, ISMI21207 was derived from THI-ALACK2019B 01-59-RB-10-1-1, while ISMI21508 originated from the cross SOUNA 3 × THIALACK2-8-2. Cluster 2 contained 85 accessions, including 15 PMiGAP inbred lines developed at ICRISAT [4], and was characterized predominantly by early-flowering genotypes. Cluster 3 consisted of 13 lines derived from Senegalese landraces, all of which were

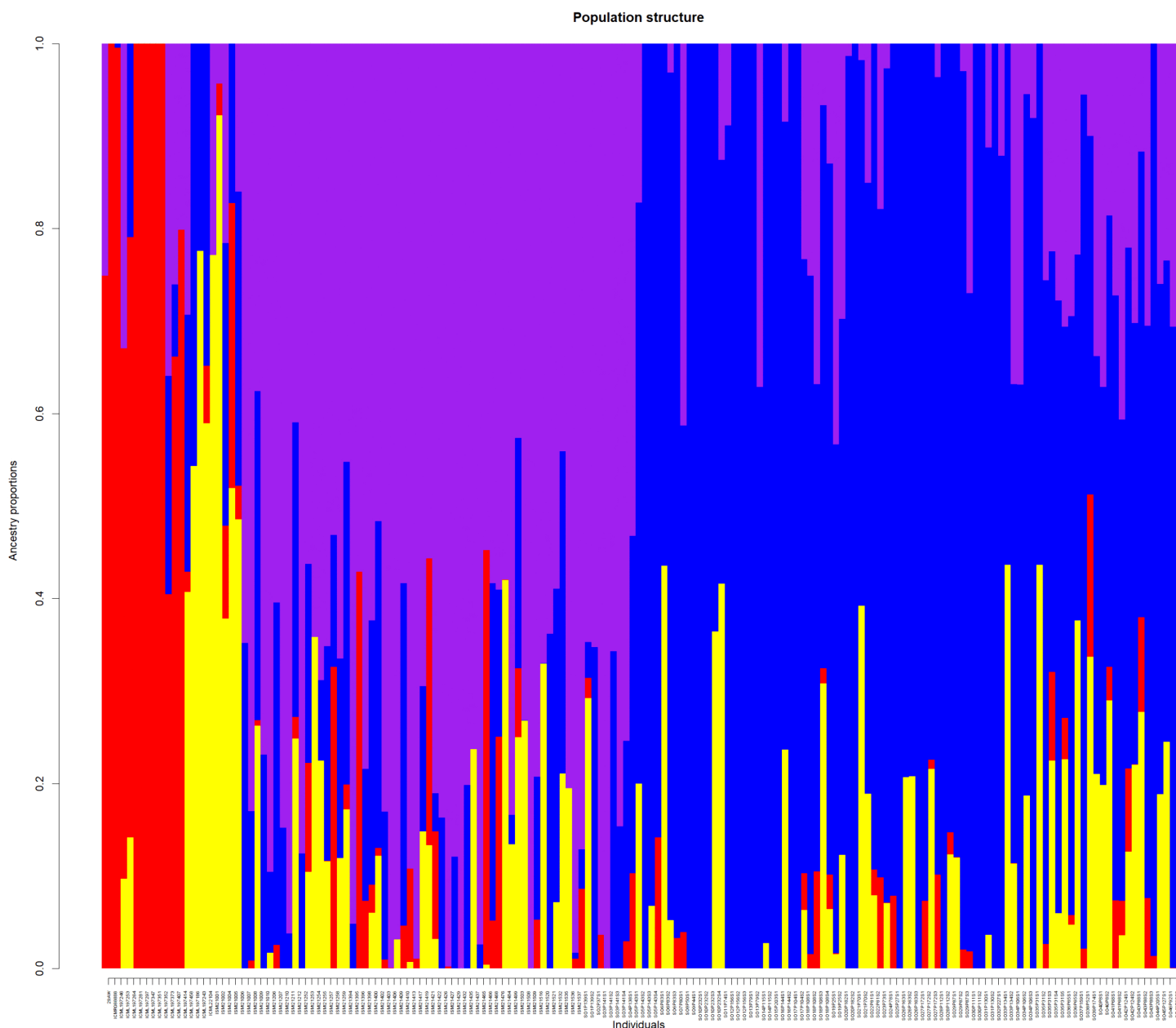


Fig 3. Population structure of the 169 pearl millet genotypes ($K = 4$). Each vertical bar represents genotypes that are divided into K -colored segments. Cluster 1(purple), Cluster 2 (blue), Cluster 3 (red), Cluster 4 (yellow).

<https://doi.org/10.1371/journal.pone.0343497.g003>

late-flowering, with several accessions sharing parental origins, such as SO42P124S2 and SO44P205S1, derived from PLS312_S1_3-1-2 and PLS312-2, respectively. Cluster 4 included nine accessions, three of which belonged to the ISMI collection, with two lines (ISMI21252 and ISMI21259) sharing parental backgrounds from SALAM2019_01-78-RB-7-1-2 and SALAM2019_01-72-RB-8-2-2, respectively. These results indicate that while geographic origin contributes to genetic differentiation, population structure in Senegalese germplasm is shaped predominantly by breeding history, shared pedigrees, flowering time, and human-mediated selection and exchange. This pattern is consistent with previous studies

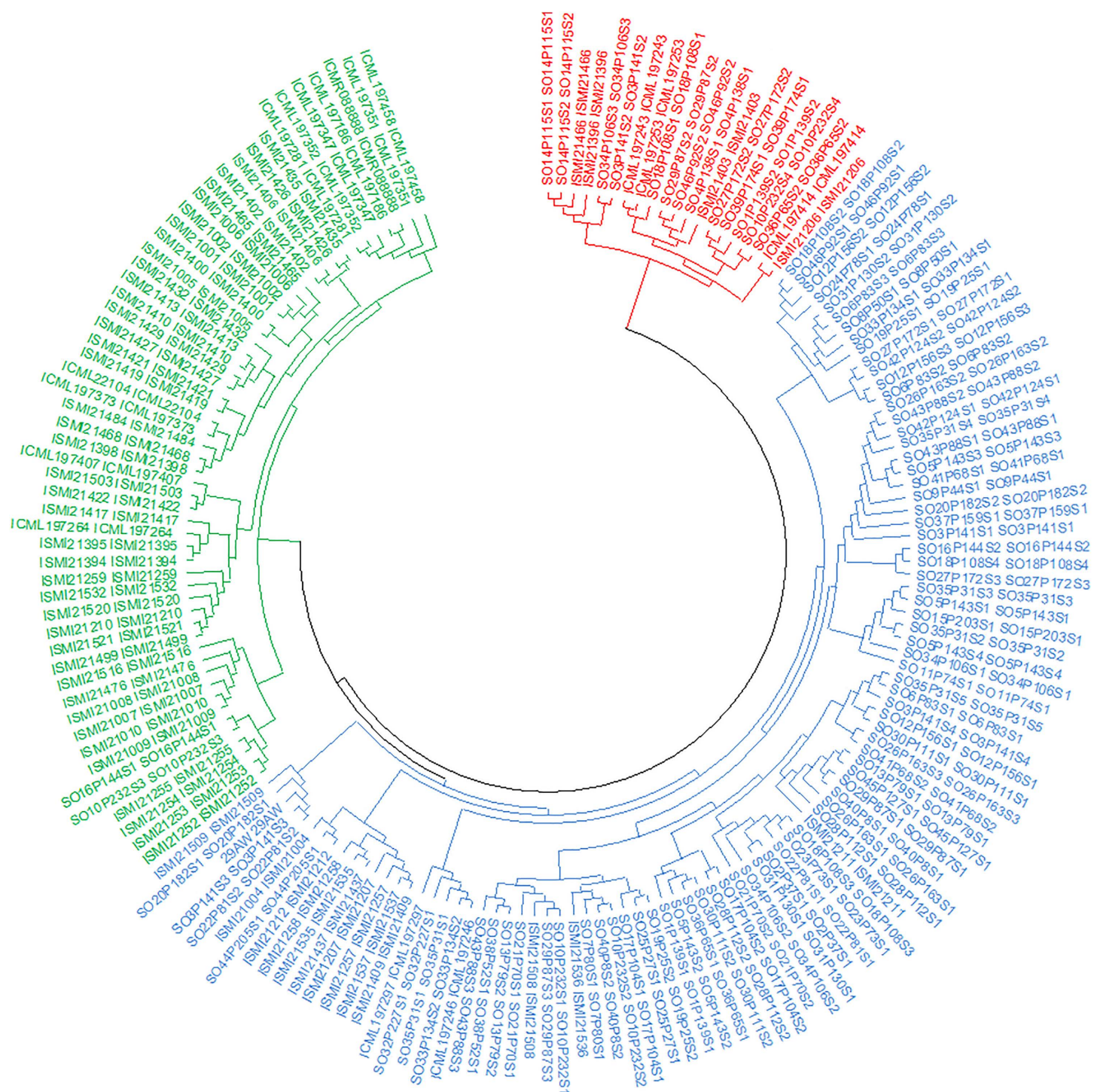


Fig 4. Phylogenetic tree of the 169 pearl millet genotypes using SNP markers. SO_xP_yS_z are the PLS lines.

<https://doi.org/10.1371/journal.pone.0343497.g004>

Table 2. Gene diversity (GD), observed heterozygosity (Ho), expected heterozygosity (He), minor allele frequency (MAF) and polymorphism information content (PIC) of the 169 pearl millet genotypes.

Group	Genotypes	GD	Ho	He	MAF	PIC
Cluster 1	62	0.28	0.11	0.28	0.20	0.23
Cluster 2	85	0.29	0.08	0.28	0.20	0.23
Cluster 3	13	0.22	0.08	0.21	0.16	0.18
Cluster 4	9	0.30	0.24	0.29	0.22	0.24
Overall population	169	0.30	0.10	0.29	0.21	0.24

<https://doi.org/10.1371/journal.pone.0343497.t002>

Table 3. Analysis of molecular variance (AMOVA) of the population based on SNP markers.

Source of variation	Df	Mean Sq	Proportion of variance (%)	Probability <i>P</i> value
Between clusters	3	15230.603	7.88	***
Within clusters	165	3899.772	92.12	***
Total	168	4102.108	100	

Significance codes: 0 '***'.

<https://doi.org/10.1371/journal.pone.0343497.t003>

Table 4. Pairwise genetic differentiation (F_{ST}) among pearl millet populations using SNP markers.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	---			
Cluster 2	0.043	---		
Cluster 3	0.071	0.063	---	
Cluster 4	0.044	0.028	0.11	---
Average F_{ST} value	0.052			

<https://doi.org/10.1371/journal.pone.0343497.t004>

showing that pearl millet genetic diversity is influenced more strongly by diversifying selection and germplasm movement than by geographic separation alone [24].

STRUCTURE-based and phylogenetic analyses produced partially contrasting patterns of genetic relationships, which can be explained by the distinct methodological frameworks and assumptions underlying these approaches. The sNMF analysis identified four genetic clusters, whereas the Neighbor-Joining (NJ) tree grouped individuals into three major clades, one of which was larger and more admixed. This discrepancy likely arises from the fact that NJ relies on pairwise genetic distances to infer hierarchical relationships, a framework that may not fully capture complex admixture patterns [25]. In contrast, sNMF directly estimates individual ancestry proportions and is therefore more sensitive to subtle population structure and admixture [19]. The presence of a large, admixed clade in the NJ tree suggests that individuals within this group share mixed ancestry, resulting in elevated genetic variability that prevents their clear separation into distinct phylogenetic branches. By assigning fractional ancestry to individuals, sNMF is better able to resolve fine-scale genetic structure and detect subpopulations that may remain unresolved in distance-based phylogenetic reconstructions [26].

These results highlight the complementary nature of the two methods, with NJ providing insights into overall evolutionary relationships and divergence, while sNMF offers a more detailed representation of population structure and admixture.

The average gene diversity estimate was 0.30, reflecting the significant diversity and adaptability potential of the genotypes constituting the germplasm. This value is, however, lower than the value obtained by Sehgal *et al.*, [4] in a pearl millet inbred germplasm association panel (PMiGAP) comprising 250 inbred lines, which is representative of cultivated

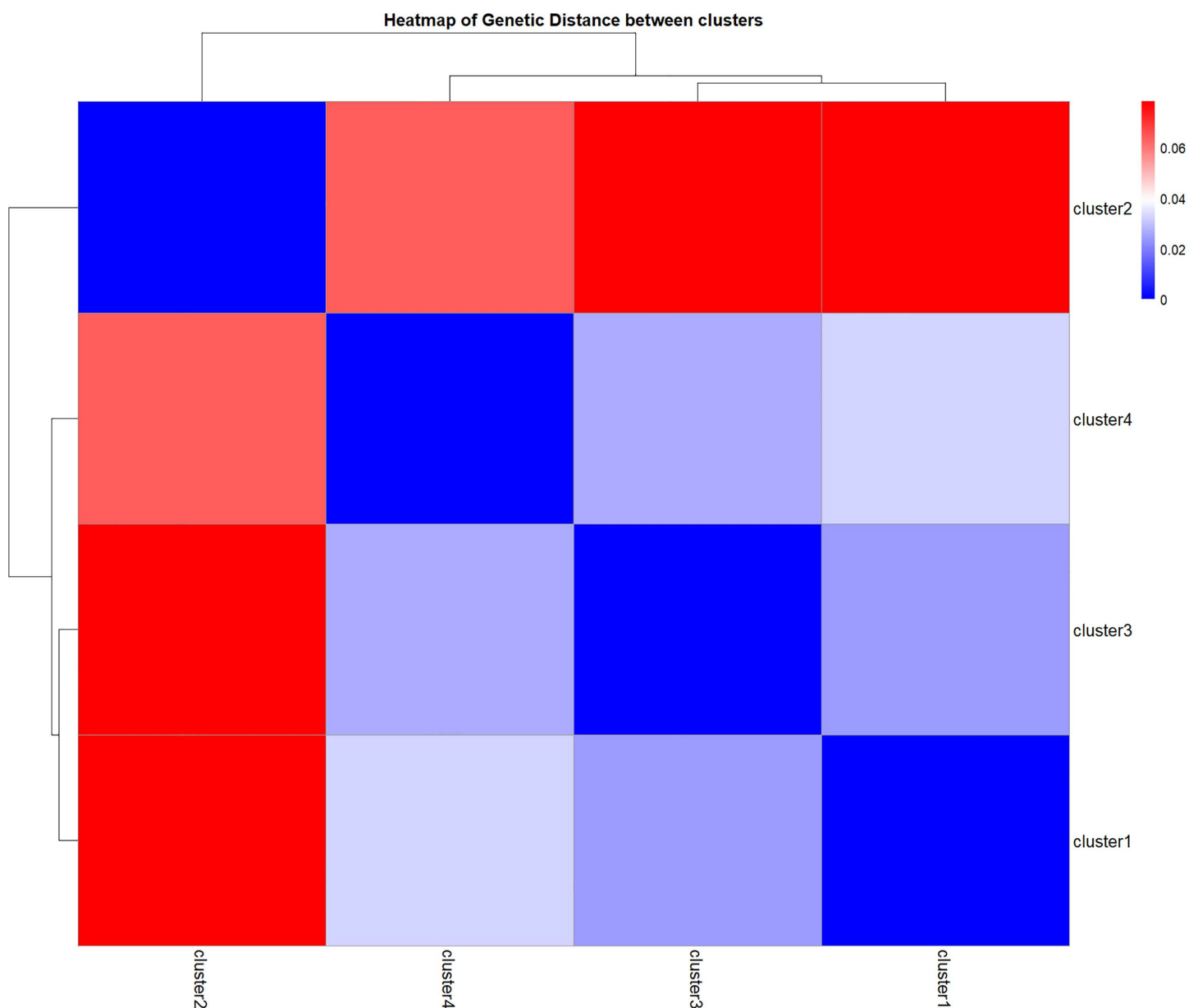


Fig 5. Heatmap of Nei genetic distances between clusters.

<https://doi.org/10.1371/journal.pone.0343497.g005>

Table 5. Nei genetic distance (below diagonal) and genetic identity (above diagonal) between pearl millet clusters based on SNP markers.

Cluster	cluster2	cluster1	cluster4	cluster3
Cluster2	---	0.974	0.976	0.924
Cluster1	0.026	---	0.967	0.938
Cluster4	0.023	0.032	---	0.925
Cluster3	0.078	0.063	0.077	---

<https://doi.org/10.1371/journal.pone.0343497.t005>

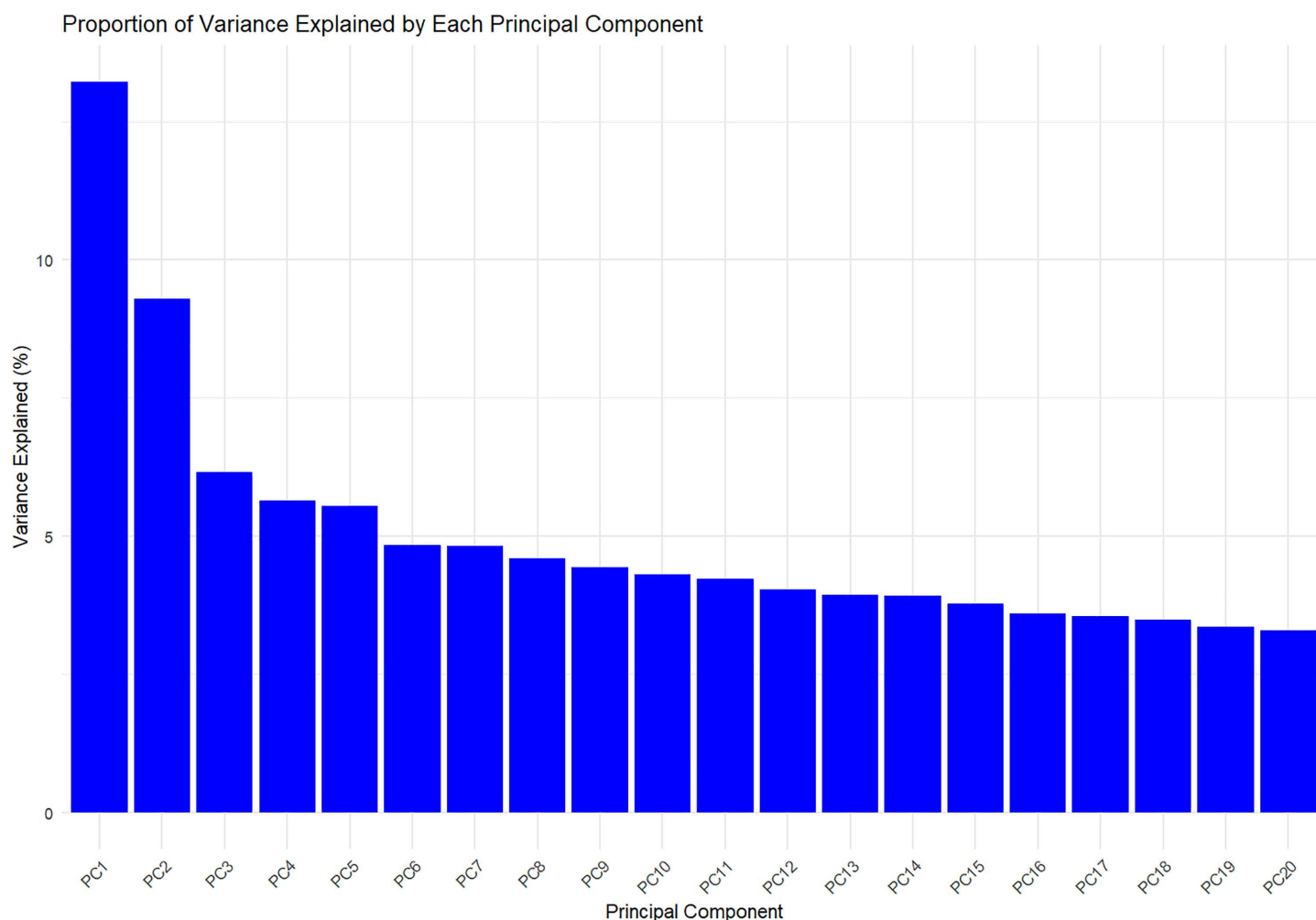


Fig 6. Percentages of explained variances for pearl millet genotypes using SNP markers.

<https://doi.org/10.1371/journal.pone.0343497.g006>

germplasms from Africa and Asia. The gene diversity estimates in the present collection are lower than those obtained for pearl millet landraces derived from West Africa and Central Africa by Stich *et al.*, [24]. This could be explained by the fact that moderately intense selection has been made on the material, narrowing the genetic base of genotypes and therefore removing some alleles from the gene pool.

Both observed heterozygosity (H_o) and expected heterozygosity (H_e) were relatively low. In contrast, Diack *et al.*, [7] reported a higher expected heterozygosity ($H_e = 0.516$) in Senegalese pearl millet germplasm. The low observed heterozygosity ($H_o = 0.10$) observed here indicates reduced heterozygosity within the panel and a high level of inbreeding, which is consistent with the nature of the genetic material used in this study, as most genotypes consisted of advanced inbred lines (F_3 and F_5). The average expected heterozygosity ($H_e = 0.29$) obtained in this study suggests low to moderate genetic diversity within clusters. This estimate is slightly lower than that reported by Kandarkar *et al.*, [27], who observed an observed heterozygosity of 0.334, although their study reported an even lower H_o (0.031) than that observed in the present collection. Together, these results indicate that breeding history and inbreeding level strongly influence heterozygosity estimates, with advanced selection and line development leading to reduce within-line heterozygosity while retaining moderate levels of allelic diversity across the panel. These findings underscore the need to balance line development

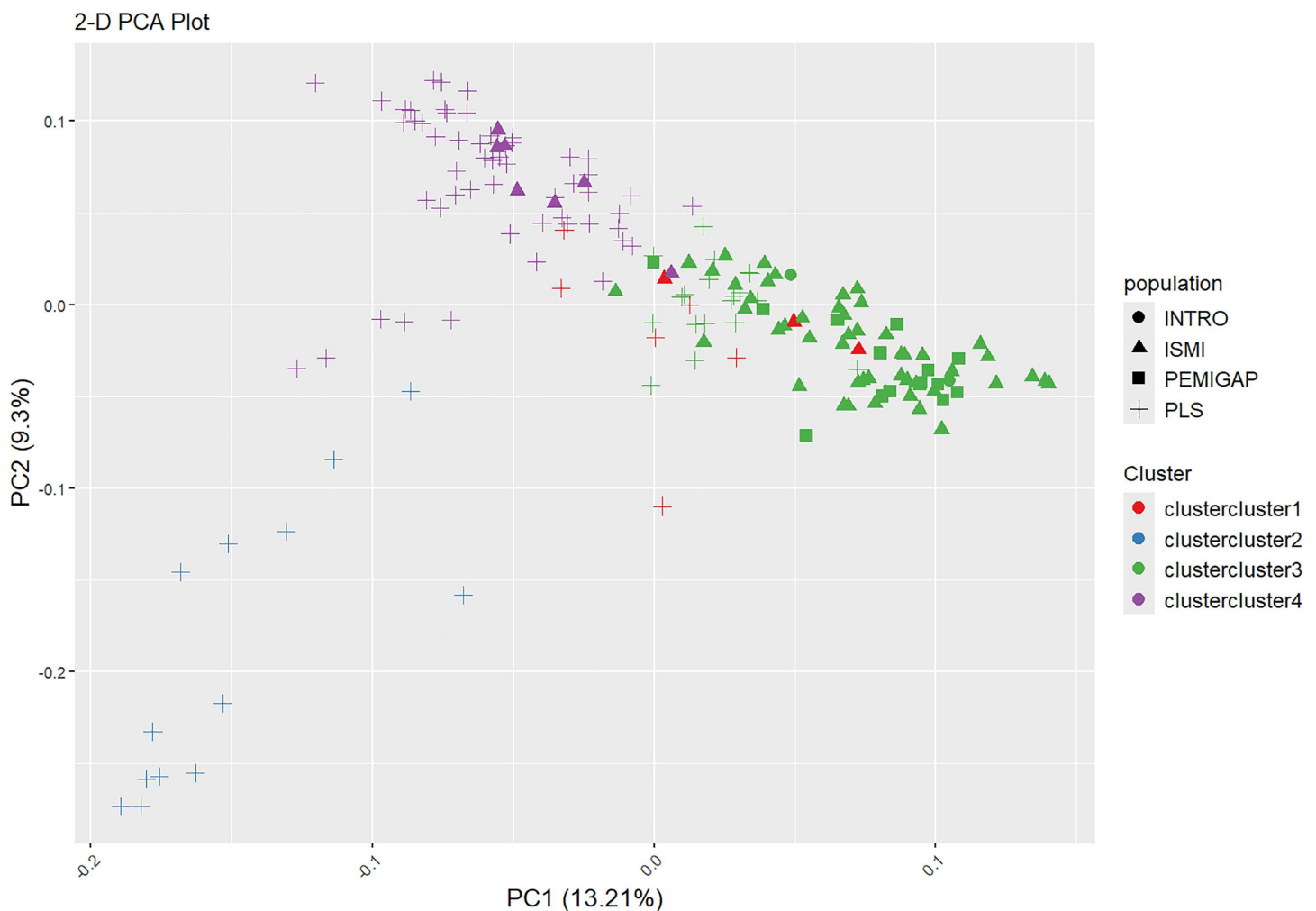


Fig 7. Principal component analysis showing the clustering among the 169 pearl millet genotypes using SNP markers. The different shapes refer to the population of origin for each genotype and the different colors refer the different clusters obtained from structure analysis.

<https://doi.org/10.1371/journal.pone.0343497.g007>

with the conservation of genetic variability in the Senegalese pearl germplasm, as excessive narrowing of the genetic base may limit future gains in adaptation and genetic improvement [28].

The average PIC of 0.24 obtained in this study indicates that this set of markers is useful to differentiate between individuals and subpopulations as discussed by Serrote *et al.*, [29]. An average MAF of 0.21 indicates the absence of rare alleles, which might have contributed to a less pronounced population structure. This could be explained by ongoing selection that can remove some of the rare alleles during the parent selection process. AMOVA revealed that 92.12% of the total genetic variation was within clusters, whereas 7.88% was attributed to differences among clusters. This suggests that the majority of genetic diversity is distributed within clusters, indicating gene flow across the studied populations. This result is similar to the findings of Kandarkar *et al.*, [27] who reported 7% of the variation among subpopulations and 93% of the variation within subpopulations in a collection of pearl millet hybrid parental lines.

Cluster 3 was the less admixed group and presented the highest genetic distance from the other clusters. This cluster was made up of only PLS lines, suggesting that these lines have fewer allele combinations shared with other clusters. The lines in cluster 3 could therefore constitute a heterotic group, which would be useful if there is heterosis with other groups. This could be useful when integrated into a hybrid breeding program, as suggested by Sehgal *et al.*, [4].

The overall average F_{ST} value of 0.052 indicates a moderate level of genetic differentiation among clusters, suggesting some population structure but also considerable gene flow across genotypes. Cluster 3, composed mainly of F_3 lines derived from late-flowering Senegalese landrace parents, showed the highest F_{ST} values, making it the most genetically distinct group. This distinctiveness, particularly the strong differentiation between Cluster 3 and Cluster 4 ($F_{ST}=0.11$) reflects differences in flowering time and parental origin, as the other clusters consisted mostly of early-flowering genotypes [4,7]. These results highlight that phenological variation and breeding history contribute significantly to the observed population structure, with Cluster 3 potentially serving as an important reservoir of unique alleles for improving adaptation and expanding the genetic base in Senegalese pearl millet breeding programs.

Conclusion

This study provides key insights for pearl millet improvement in Senegal by integrating population structure, genetic diversity, and pedigree information. The extensive admixture, combined with the moderate genetic differentiation between clusters, and the low observed heterozygosity reflect the combined effects of allogamy, germplasm exchange, and advanced line development, highlighting the need to balance population advancement with diversity conservation. The identification of four genetic clusters, particularly the distinctiveness of Cluster 3 derived from late-flowering landraces offers clear opportunities for informed parental selection, heterotic group development, and genetic-based germplasm management. These efforts can help optimize crossing strategies, in order to facilitate association studies as well as QTL discovery, enhance breeding efficiency and promote long-term genetic gain.

Supporting information

S1 Table. List of the genotypes used in the study.

(XLSX)

S2 Table. List of individual and their clusters.

(TXT)

S3 Table. Raw genotypic data.

(TXT)

S4 Table. Eigenvectors table.

(EIGENVEC)

S5 Table. Eigenvalues table.

(EIGENVAL)

Acknowledgments

The authors would like to acknowledge the Senegalese Agricultural Research Institute (ISRA) for providing the plant materials for this study and CIMMYT for providing the raw genotypic data. Aliou BA would also like to extend his gratitude to Makerere University for hosting him as a Ph.D. fellow. The support and contribution of the Makerere University Regional Centre for Crop Improvement (MaRCCI) are also acknowledged.

Author contributions

Conceptualization: Aliou Ba, Arfang Badji, Oumar Diack.

Data curation: Aliou Ba.

Formal analysis: Aliou Ba, Oumar Diack.

Investigation: Aliou Ba.

Methodology: Aliou Ba, Arfang Badji, Oumar Diack, Tony Obua.

Software: Aliou Ba.

Supervision: Thomas L. Odong, Oumar Diack, Njido A. Kane, Phinehas Tukamuhabwa.

Writing – original draft: Aliou Ba, Oumar Diack, Tony Obua, Mildred Ochwo-Semakula, Paul Gibson.

Writing – review & editing: Thomas L. Odong, Richard Edema, Arfang Badji, Oumar Diack, Tony Obua, Mildred Ochwo-Semakula, Paul Gibson.

References

1. Srivastava RK, Singh RB, Pujarula VL, Bollam S, Pusuluri M, Chellapilla TS, et al. Genome-wide association studies and genomic selection in pearl millet: advances and prospects. *Front Genet.* 2020;10:1389. <https://doi.org/10.3389/fgene.2019.01389> PMID: [32180790](#)
2. Kane NA, Foncéka D, Dalton TJ. Crop adaptation and improvement for drought-prone environments. 2022.
3. Kanfany G, Diack O, Kane NA, Gangashetty P. Implications of farmer perceived production constraints and. *Afr Crop Sci J.* 2020;28:411–20.
4. Sehgal D, Skot L, Singh R, Srivastava RK, Das SP, Taunk J, et al. Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. *PLoS One.* 2015;10(5):e0122165. <https://doi.org/10.1371/journal.pone.0122165> PMID: [25970600](#)
5. Hu Z, Mbacké B, Perumal R, Guèye MC, Sy O, Bouchet S, et al. Population genomics of pearl millet (*Pennisetum glaucum* (L.) R. Br.): Comparative analysis of global accessions and Senegalese landraces. *BMC Genomics.* 2015;16:1048. <https://doi.org/10.1186/s12864-015-2255-0> PMID: [26654432](#)
6. Olodo KF, Barnaud A, Kane NA, Mariac C, Faye A, Couderc M, et al. Abandonment of pearl millet cropping and homogenization of its diversity over a 40 year period in Senegal. *PLoS One.* 2020;15(9):e0239123. <https://doi.org/10.1371/journal.pone.0239123> PMID: [32925982](#)
7. Diack O, Kane NA, Berthouly-Salazar C, Gueye MC, Diop BM, Fofana A, et al. New genetic insights into pearl millet diversity as revealed by characterization of early- and late-flowering landraces from Senegal. *Front Plant Sci.* 2017;8:818. <https://doi.org/10.3389/fpls.2017.00818> PMID: [28567050](#)
8. Yirgu M, Kebede M, Feyissa T, Lakew B, Woldeyohannes AB, Fikere M. Single nucleotide polymorphism (SNP) markers for genetic diversity and population structure study in Ethiopian barley (*Hordeum vulgare* L.) germplasm. *BMC Genom Data.* 2023;24(1):7. <https://doi.org/10.1186/s12863-023-01109-6> PMID: [36788500](#)
9. Garcia AAF, Benchimol LL, Barbosa AMM, Geraldi IO, Souza Jr CL, de Souza AP. Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. *Genet Mol Biol.* 2004;27(4):579–88. <https://doi.org/10.1590/s1415-47572004000400019>
10. Shehzad M, Ditta A, Sajid Iqbal M, Jarwar AH. Role of molecular markers and importance of SNP for the development of cotton programs. *J Biol Agric Healthc.* 2017;7:61–73.
11. Aleem S, Tahir M, Sharif I, Aleem M, Najeebullah M, Nawaz A, et al. Principal component and cluster analyses as tools in the assessment of genetic diversity for late season cauliflower genotypes. *Pak J Agric Res.* 2021;34(1). <https://doi.org/10.17582/journal.pjar/2021/34.1.176.183>
12. Das S, Sawarkar A, Saha S, Raman RB, Dasgupta T. Principal component and cluster analysis in Mungbean [*Vigna radiata* (L.) Wilczek]. *Legum Res - Int J.* 2024;1–10. <https://doi.org/10.18805/lr-5305>
13. Werner CR, Gaynor RC, Gorjanc G, Hickey JM, Kox T, Abbadi A, et al. How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. *Front Plant Sci.* 2020;11:592977. <https://doi.org/10.3389/fpls.2020.592977> PMID: [33391305](#)
14. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al. Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol.* 2012;888:67–89. https://doi.org/10.1007/978-1-61779-870-2_5 PMID: [22665276](#)
15. Roshayara NR, Kirsten H, Horn K, Ahnert P, Scholz M. Impact of pre-imputation SNP-filtering on genotype imputation results. *BMC Genet.* 2014;15:88. <https://doi.org/10.1186/s12863-014-0088-5> PMID: [25112433](#)
16. Granato ISC, Galli G, de Oliveira Couto EG, e Souza MB, Mendonça LF, Fritsche-Neto R. snpReady: a tool to assist breeders in genomic analysis. *Mol Breeding.* 2018;38(8). <https://doi.org/10.1007/s11032-018-0844-8>
17. Jombart T. Package “ade4genet” Title Exploratory Analysis of Genetic and Genomic Data. *Bioinformatics.* 2023;1403–5.
18. Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet.* 2015;6:208. <https://doi.org/10.3389/fgene.2015.00208> PMID: [26113860](#)
19. Frichot E, François O. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol.* 2015;6(8):925–9. <https://doi.org/10.1111/2041-210x.12382>
20. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945–59. <https://doi.org/10.1093/genetics/155.2.945> PMID: [10835412](#)
21. Charlesworth D, Wright SI. Breeding systems and genome evolution. *Curr Opin Genet Dev.* 2001;11(6):685–90. [https://doi.org/10.1016/s0959-437x\(00\)00254-9](https://doi.org/10.1016/s0959-437x(00)00254-9) PMID: [11682314](#)

22. Bøhn T, Aheto DW, Mwangala FS, Fischer K, Bones IL, Simoloka C, et al. Pollen-mediated gene flow and seed exchange in small-scale Zambian maize farming, implications for biosafety assessment. *Sci Rep*. 2016;6:34483. <https://doi.org/10.1038/srep34483> PMID: [27694819](#)
23. Budak LR. In a collection of pearl millet germplasm. *Crop Sci*. 2002;43:2284–90.
24. Stich B, Haussmann BI, Pasam R, Bhosale S, Hash CT, Melchinger AE, et al. Patterns of molecular and phenotypic diversity in pearl millet [*Pennisetum glaucum* (L.) R. Br.] from West and Central Africa and their relation to geographical and environmental parameters. *BMC Plant Biol*. 2010;10(1):216. <https://doi.org/10.1186/1471-2229-10-216>
25. Zou Y, Zhang Z, Zeng Y, Hu H, Hao Y, Huang S, et al. Common methods for phylogenetic tree construction and their implementation in R. *Bioengineering* (Basel). 2024;11(5):480. <https://doi.org/10.3390/bioengineering11050480> PMID: [38790347](#)
26. Rousset F, Peyrin F, Ducros N. A semi nonnegative matrix factorization technique for pattern generalization in single-pixel imaging. *IEEE Trans Comput Imaging*. 2018;4(2):284–94. <https://doi.org/10.1109/tci.2018.2811910>
27. Kandarkar K, Palaniappan V, Satpathy S, Vemula A, Rajasekaran R, Jeyakumar P, et al. Understanding genetic diversity in drought-adaptive hybrid parental lines in pearl millet. *PLoS One*. 2024;19(2):e0298636. <https://doi.org/10.1371/journal.pone.0298636> PMID: [38394324](#)
28. Wallace JG, Rodgers-Melnick E, Buckler ES. On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annu Rev Genet*. 2018;52:421–44. <https://doi.org/10.1146/annurev-genet-120116-024846> PMID: [30285496](#)
29. Serrote CML, Reiniger LRS, Silva KB, Rabaioli SMD, Stefanel CM. Determining the polymorphism information content of a molecular marker. *Gene*. 2020;726:144175. <https://doi.org/10.1016/j.gene.2019.144175> PMID: [31726084](#)