# Genetic Diversity and Population Structure of Mozambican Lowland Rainfed Rice (*Oryza sativa* L.) Using DArTseq SNP Markers

Kajale George Warioba , Celsa Mondlane Macandza * , Leonel Domingos Moiana

Posted Date: 22 April 2026

doi: 10.20944/preprints202604.1491.v1

Keywords: DArTseq; genetic diversity; lowland rainfed rice; population structure; SNP markers

 doi:10.20944/preprints202604.1491.v1



*Article*

# Genetic Diversity and Population Structure of Mozambican Lowland Rainfed Rice (*Oryza sativa* L.) Using DArTseq SNP Markers

**Kajale George Warioba [1,2,3], Celsa Mondlane Macandza [1,2,*] and Leonel Domingos Moiana [4,5]**

[1] Department of Plant Production, Faculty of Agronomy and Forestry Engineering, Eduardo Mondlane University, P.O. Box 257, Maputo 1100, Mozambique

[2] Centre of Excellence in Agri-Food Systems and Nutrition (CE-AFSN), Eduardo Mondlane University, 5th Floor, Rectory Building, 25th June Square, Maputo 1100, Mozambique

[3] Department of Research and Innovation, Tanzania Agricultural Research Institute (TARI), Arusha Road, Dodoma, P.O. Box 1571, Tanzania

[4] Regional Research Centre of Leadership for Rice, Mozambique Agricultural Research Institute (IIAM), National Road No. 1, Namacurra, Zambezia 2400, Mozambique

[5] Department of Agriculture and Animal Health, School of Agriculture and Life Sciences, University of South Africa, P.O. Box 392, Roodepoort 0003, South Africa

* Correspondence: celsa.macandza@uem.mz

**Abstract**

Rice (*Oryza sativa* L.) is an important staple crop in Mozambique, and understanding its genetic diversity is essential for crop improvement, genetic resources management and conservation. However, molecular characterization of Mozambican rice germplasm remains limited. This study assessed genetic diversity and population structure of 40 lowland rainfed rice genotypes using 3473 high-quality single nucleotide polymorphism (SNP) markers generated through DArTseq™ genotyping-by-sequencing platform. Results revealed moderate genetic diversity with a mean polymorphism information content of 0.25, indicating moderate marker informativeness. Unbiased expected heterozygosity (uHe = 0.314) was higher than observed heterozygosity (Ho = 0.125), reflecting the inbreeding nature of rice ($F_{IS}$ = 0.357). Model-based admixture analysis identified four subpopulations, with 20% of genotypes classified as admixed. Substantial genetic differentiation was observed among these subpopulations ($F_{ST}$ = 0.267), which was broadly consistent with the principal coordinate analysis and the neighbor-joining tree. Furthermore, a high mean Manhattan dissimilarity index (0.70), indicated strong genetic divergence across the panel. Analysis of molecular variance revealed significant variation among subpopulations (32.90%) and within subpopulations (67.10%). These findings provide foundational genetic insights to guide Mozambican rice breeding programs and support the long-term conservation of local germplasm.

**Keywords:** DArTseq; genetic diversity; lowland rainfed rice; population structure; SNP markers

## 1. Introduction

Rice (Oryza sativa L.) is a strategic food and commercial crop, important for food security and employment across sub-Saharan Africa (SSA) [1]. It contributes to caloric intake and supports the livelihoods of millions, with growing role as more households participate in the rice value chain [2,3]. The harvested rice area in SSA increased from 12 to 18 million hectares between 2014 and 2024, with production rising from 26 to 38 million tons [4]. Yet, productivity stagnated at about 2t/ha over the same period [4]. Amid rising per capita consumption, this highlights the need to close the yield gap through improved breeding and genetic resource management [5].

 doi:10.20944/preprints202604.1491.v1



In Mozambique, rice is a staple crop [6]. In 2024, production reached 169,311 tons, second to maize (2.1 million tons) among cereal crops, ranking fifth in Southern and Eastern Africa, behind Madagascar, Tanzania, Uganda and Kenya [4]. Cultivation is dominated by smallholder farmers in rainfed lowland systems of the Zambezi river basin in the central region [7]. Farms are typically small (<1 ha), with limited fertilizers use and poor soil and water management. Reliance of unimproved local landraces contributes to low overall rice productivity [7–9].

Mozambican consumers prefer the aroma and grain texture of local landraces [9]. However, these landraces are underutilized in breeding due to limited genetic characterization and the need for extensive pre-breeding, a pattern observed across sub-Saharan Africa [10,11]. Breeding programs therefore rely on germplasms from the International Rice Research Institute and AfricaRice. While these varieties offer uniformity and proven performance, they may lack specific attributes preferred by local farmers [12,13]. Research on rice genetic diversity in Mozambique remains limited, with most data derived from regional analyses rather than country-specific investigations [14,15]. Recent morphological evaluations [16,17] are addressing this gap, yet molecular marker studies are still scarce.

Quantifying genetic diversity is essential for informing targeted breeding. While diversity can be assessed using morphological and molecular approaches, morphological markers alone are often constrained by growth stage and the complexity of gene interactions such as epistasis and pleiotropy [18]. In contrast, molecular markers such as single nucleotide polymorphisms (SNP), provide a robust germplasm characterization, minimizing environmental effects and offering higher resolution [19]. SNPs detect genome-wide variations, enabling precise genotype discrimination, trait mapping and population structure analysis [20]. Diversity is commonly quantified using parameters such as heterozygosity, while population structure and genetic differentiation are assessed using fixation indices, analysis of molecular variance, and model-based admixture [21]. Genetic relationships are further explored using principal coordinate analysis and neighbor-joining trees to identify distinct groups [22]. High diversity indicates strong breeding potential and supports the identification of suitable parental lines for crossing. These approaches have been successful applied in rice genetic diversity studies [23,24]. Building on these analytical approaches, this study assessed the genetic diversity and population structure of 40 lowland rainfed rice genotypes from Mozambique using single nucleotide polymorphism markers. It provides insights into the extent and distribution of genetic variation, supporting parental selection, genetic improvement, and conservation.

## 2. Results

### *2.1. SNP Markers Distribution and Quality*

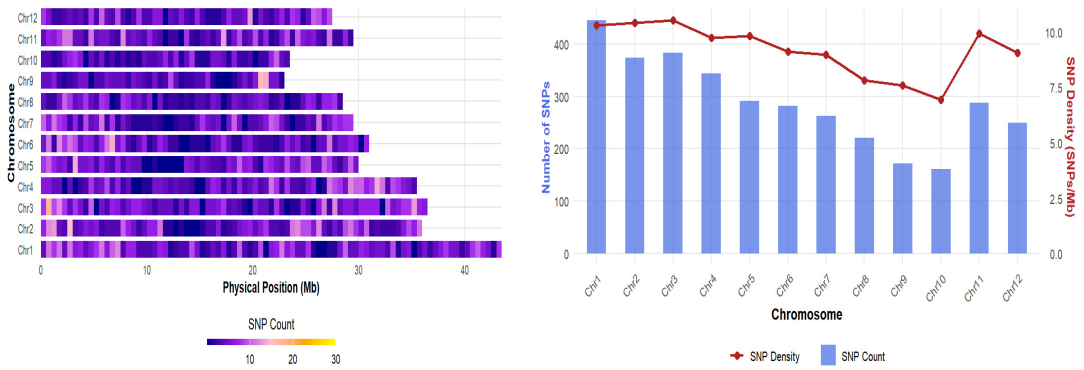
A total of 22046 SNPs was initially obtained from 40 rice genotypes. After filtering criteria, a total of 3473 SNP markers spanning a cumulative physical length of 371.39 Mb across the 12 rice chromosomes (Chr) were retained for genetic diversity analyses (Figure 1A). The distribution of SNP markers showed a moderate heterogeneity, ranging from 6.94 SNP/Mb (Chr10) to 10.53 SNP/Mb (Chr3) (Figure 1B). The highest marker densities were observed on Chr3 (10.53), Chr2 (10.41) and Chr1 (10.29), indicating relatively dense marker coverage in these regions. In contrast, lower densities were observed on Chr10 (6.94), Chr9 (7.59) and Chr8 (7.82). Chr1 harbored the highest number of SNP markers 445, while Chr10 had the lowest (161). Overall, SNPs were well distributed across the genome.

Among the retained SNP markers, 65.8% had a call rate above 98%, while 34.2% had a call rate of 97% (Figure 2).

Analysis of rice genome revealed that 63.2% of mutations were transitions (A/G and C/T) and 36.8% were transversions (A/C, A/T, G/C, G/T), resulting in a Ts/Tv ratio of 1.72 (Figure 3). Among transitions for A/G and C/T accounted for 31.4% and 31.8% of polymorphisms, respectively, while transversions were distributed as 10.2% (A/C), 9.7% (G/T), 9.4% (A/T) and 7.5% (G/C).

 doi:10.20944/preprints202604.1491.v1






(**A**) (**B**)

**Figure 1.** Distribution of SNP markers across the 12 rice chromosomes showing variation in marker count and density.

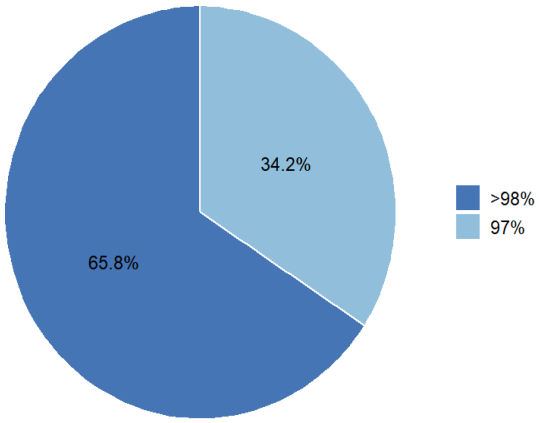



**Figure 2.** SNP markers by call rate proportion after applying filtering criteria (call rate >95% and MAF >5%).

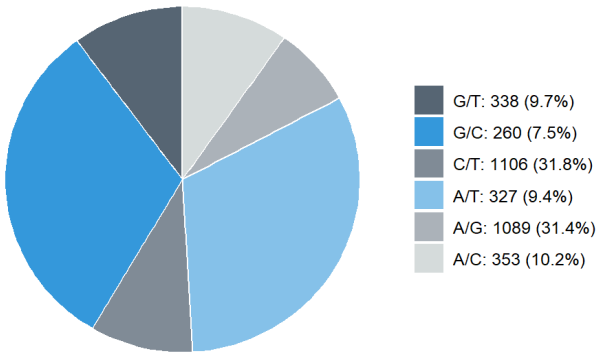



**Figure 3.** The distribution of SNP transitions and transversions mutation types across the genomes of the 40 rice genotypes.

 doi:10.20944/preprints202604.1491.v1



*2.2. Genetic Diversity Parameters and Visualization*

A total of 98.04% of retained loci were polymorphic (Table 1). Polymorphic information content (PIC) ranged from 0.000 to 0.375 averaging 0.250. Unbiased expected heterozygosity (uHe) varied from 0.000 to 0.506 with a mean of 0.314, whereas, observed heterozygosity (Ho) was lower, ranging from 0.000 to 0.750, with mean of 0.125 (Supplementary File S1; Figures S3–S5).

**Table 1.** Genetic diversity parameters across 40 rice genotypes using 3473 SNP markers.

| Parameter | Minimum | Maximum | Mean ± SE |
|---|---|---|---|
| Observed heterozygosity (Ho) | 0.000 | 0.750 | 0.125 ± 0.001 |
| Unbiased expected heterozygosity (uHe) | 0.000 | 0.506 | 0.314 ± 0.003 |
| Polymorphic information content (PIC) | 0.000 | 0.375 | 0.250 ± 0.002 |
| Total number of SNP | | 3473 | |
| % of polymorphic loci | | 98.04 | |

- Principal coordinate analysis (PCoA) and Neighbor-joining (NJ) tree

Principal coordinate analysis (PCoA) and neighbor-joining (NJ) tree constructed based on a Manhattan dissimilarity index derived from 3473 SNP markers revealed clear genetic structuring among the 40 rice genotypes. The pairwise dissimilarity index ranged from 0.05 to 1.00, with a mean value of 0.70 (Supplementary File S1; Figure S7). The lowest dissimilarity (0.05) was observed between genotypes G33 and G52, whereas the highest dissimilarity (1.00) was recorded between genotypes G36 and G65. For PCoA (Figure 4), the first two principal coordinates both had eigenvalues greater than 1 and together explained 48.29% of the total variation (PCo1 = 34.21%; PCo2 = 14.08%) and were therefore used to visualize genetic relationships among genotypes (Supplementary File S1; Table S2). The resulting scatterplot resolved the genotypes into three distinct clusters (Supplementary File S1; Figure S6). Landraces were distributed across all clusters, whereas improved lines were confined to cluster 1. This cluster was clearly separated from clusters 2 and 3, indicating greater genetic divergence. Differences in within-cluster dispersion were also observed, with cluster 2 forming a compact grouping, while cluster 1 showed a wider spread (Supplementary File S1; Table S1). The NJ tree clustering pattern was mostly consistent with the principal coordinate analysis (Figure 5).

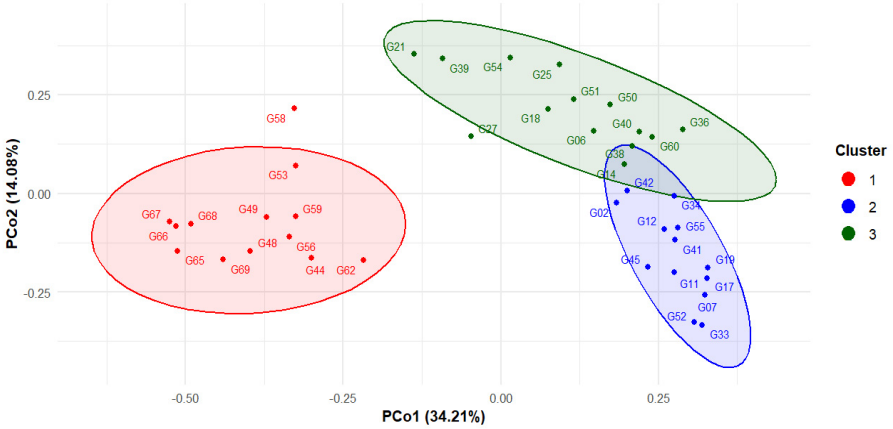



**Figure 4.** Principal coordinate analysis, illustrating the genetic relationships among rice genotypes. Ellipses represent 95% confidence intervals of the spread. Genotypes names are listed in Table 5.

 doi:10.20944/preprints202604.1491.v1



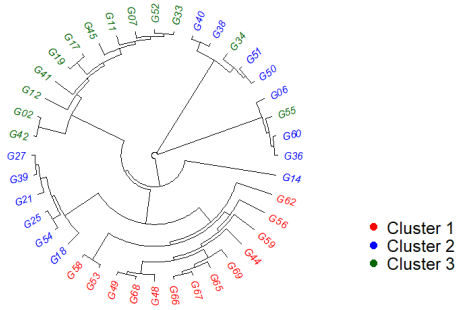



**Figure 5.** SNP-based Neighbor-Joining tree showing the genetic relationships among 40 rice genotypes. The major clusters are indicated by different colors. Genotypes names are listed in Table 5.

*2.3. Population Structure Analysis*

- Admixture

The optimal number of subpopulations was K = 4, as determined by the minimum cross-entropy criterion (Supplementary File S1; Figure S8). At this level, four distinct subpopulations and an admixed group were identified (Figure 6). Subpopulation I comprised 10 genotypes (G44, G48, G49, G56, G59, G65, G66, G67, G68 and G69), while Subpopulation II included 11 genotypes (G06, G14, G18, G25, G36, G38, G40, G50, G51, G54 and G60). Subpopulation III consisted of 3 genotypes (G27, G39 and G58), and Subpopulation IV contained 8 genotypes (G07, G11, G12, G17, G19, G33, G45 and G52). The admixed group comprised 8 genotypes (G02, G21, G34, G41, G42, G53, G55 and G62) (Supplementary File S1; Table S3). No completely pure subpopulation was observed, as all four subpopulations exhibited average ancestry coefficients ranging from 80% to 90%, indicating varying degree of shared ancestry. In contrast, the admixed group showed a lower average ancestry coefficient of about 50%, reflecting substantial genetic admixture. The inferred population structure did not fully correspond with the clustering patterns observed in the neighbor-joining tree and principal coordinate analysis. However, broadly consistent patterns were evident, as most genetically related individuals were grouped together across the different analytical approaches.

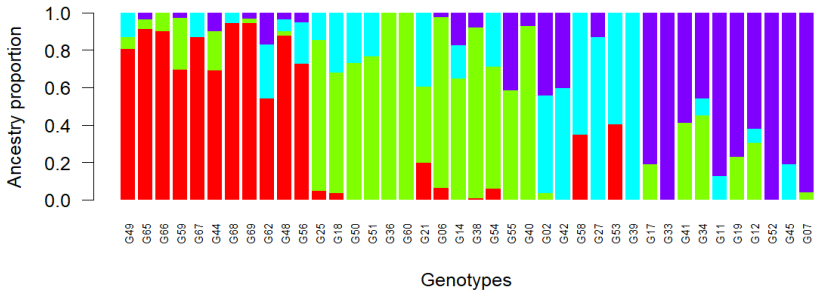



**Figure 6.** Population structure inferred from admixture analysis at K = 4. Each individual is represented by a vertical bar portioned into color segments corresponding to estimated ancestry proportions (Q-values) across the four clusters.

 doi:10.20944/preprints202604.1491.v1



- Population genetic diversity parameters

Subpopulation III was the most genetically diverse, recording the highest values for observed heterozygosity (Ho = 0.320), unbiased expected heterozygosity (uHe = 0.313), and nucleotide diversity ($\pi$ = 0.313), however this subpopulation consists of only three genotypes (Table 2). In contrast, Subpopulation IV was the least diverse exhibiting the lowest values for Ho (0.061), uHe (0.132), $\pi$ (0.132), and percentage of polymorphic loci (45.93%). Polymorphic information content (PIC) values were generally low across subpopulations, ranging from 0.082 in Subpopulation IV to 0.161 in Subpopulation I, while the admixed group exhibited slightly higher value (0.172). Across all subpopulations and the admixed group, unbiased expected heterozygosity (uHe) generally exceeded observed heterozygosity (Ho), resulting in overall positive inbreeding coefficient ($F_{IS}$ = 0.357). Subpopulation II exhibited the highest level of inbreeding ($F_{IS}$ = 0.551), whereas Subpopulation III had the lowest ($F_{IS}$ = -0.064).

**Table 2.** Population genetic diversity parameters (mean ± SE) calculated for each subpopulation and admixed group.

| Parameter | n | Ho | uHe | PIC | %P | $F_{IS}$ | $\pi$ |
|---|---|---|---|---|---|---|---|
| Subpop I | 10 | 0.120 ± 0.002 | 0.283 ± 0.003 | 0.161 ± 0.002 | 78.52 ± 0.70 | 0.528 ± 0.006 | 0.283 ± 0.003 |
| Subpop II | 11 | 0.081 ± 0.002 | 0.226 ± 0.004 | 0.130 ± 0.002 | 65.65 ± 0.81 | 0.551 ± 0.006 | 0.226 ± 0.004 |
| Subpop III | 3 | 0.320 ± 0.005 | 0.313 ± 0.004 | 0.156 ± 0.002 | 67.81 ± 0.79 | -0.064 ± 0.008 | 0.313 ± 0.004 |
| Subpop IV | 8 | 0.061 ± 0.002 | 0.132 ± 0.003 | 0.082 ± 0.002 | 45. 93 ± 0.85 | 0.414 ± 0.008 | 0.132 ± 0.003 |
| Admixed | 8 | 0.183 ± 0.003 | 0.309 ± 0.003 | 0.172 ± 0.002 | 81.57 ± 0.66 | 0.355 ± 0.006 | 0.309 ± 0.003 |
| Overall | 40 | 0.153 ± 0.003 | 0.253 ± 0.004 | 0.140 ± 0.002 | 67.90 ± 0.76 | 0.357 ± 0.007 | 0.253 ± 0.003 |

Key: n = number of genotypes; Ho = observed heterozygosity; uHe = unbiased expected heterozygosity; PIC = polymorphic information content; %P = percentage of polymorphic loci; $F_{IS}$ = inbreeding coefficient; $\pi$ = nucleotide diversity; Subpop = subpopulation.

- Pairwise fixation index ($F_{ST}$) and Nei's genetic distance

Genetic differentiation among the four subpopulations was further assessed using pairwise fixation index ($F_{ST}$) and Nei's genetic distance (Table 3). The overall $F_{ST}$ across all groups was 0.267, indicating a substantial level of population structure. The pairwise $F_{ST}$ values ranged from 0.170 to 0.458, with the highest genetic differentiation between Subpopulation I and IV ($F_{ST}$ = 0.458), and the lowest between Subpopulation I and III ($F_{ST}$ = 0.170). The admixed group showed low levels of differentiation from Subpopulation II and III. These results were largely consistent with the pattern obtained for Nei's genetic distance analysis which ranged from 0.131 to 0.284, with an overall value of 0.154. The greatest divergence was again detected between Subpopulations I and IV (0.284), whereas the lowest was between II and III (0.131).

**Table 3.** Pairwise estimates of genetic differentiation among the four subpopulations and the admixed group, with $F_{ST}$ values presented below the diagonal and Nei's genetic distances shown above the diagonal.

| Subpopulation | | Nei's genetic distance | | | | |
|---|---|---|---|---|---|---|
| | | I | II | III | IV | Admixed |
| $F_{ST}$ | I | 0 | 0.215 | 0.155 | 0.284 | 0.146 |
| | II | 0.336 | 0 | 0.131 | 0.148 | 0.058 |
| | III | 0.170 | 0.184 | 0 | 0.229 | 0.094 |
| | IV | 0.458 | 0.347 | 0.443 | 0 | 0.077 |
| | Admixed | 0.200 | 0.077 | 0.056 | 0.163 | 0 |

Key: $F_{ST}$ = pairwise fixation index.

 doi:10.20944/preprints202604.1491.v1



- Analysis of molecular variance

Analysis of Molecular Variance (AMOVA) revealed that 32.90% of the total genetic variation was attributable to differences among subpopulations (variance component = 311.36), whereas 67.10% occurred within subpopulations (variance component = 635.11) (Table 4). The variation among subpopulations was significance based on Monte Carlo permutation test with 999 replicates (p = 0.001), with the observed variance component (311.36) exceeding the expected variance under random distribution (-0.22).

**Table 4.** Analysis of molecular variance (AMOVA) showing the partitioning of genetic variation among subpopulations and within subpopulations.

| Source | df | SS | MS | Est. var | % variation | Φ-statistics | P-value |
|---|---|---|---|---|---|---|---|
| Among subpopulations | 4 | 12208.30 | 3052.07 | 311.36 | 32.90 | 0.329 | 0.001 |
| Within subpopulations | 35 | 22228.81 | 635.11 | 635.11 | 67.10 | | |
| Total | 39 | 34437.10 | 883.00 | 946.47 | 100 | | |

Key: df = degree of freedom; SS = sum squares; MS = mean squares; Est. var = estimated variance.

## 3. Discussion

This study used single nucleotide polymorphism (SNP) markers to assess genetic diversity and population structure among 40 rice genotypes. SNP markers estimate genetic relatedness and allelic variation to inform breeding decisions. The observed genetic diversity can be leveraged to broaden the genetic base and improve germplasm utilization by guiding parental selection.

Stringent filtering retained 3473 SNPs across 12 rice chromosomes (Chr) and 371 Mb, providing good genome coverage. Moderate SNP density, higher on Chr1 – Chr3 and lower on Chr8 – Chr10, likely reflected differences in recombination, gene distribution, or selection pressures. However, this variation suggested no major chromosomal bias affecting diversity estimates. Similarly, author [25], reported higher SNP density on Chr1 and lower on Chr9 using 15020 SNPs across 373 Mb. Differences in SNP numbers may arise from genotyping platforms, sample quality, and filtering criteria, influencing diversity and population structure estimates [23]. High call rates confirm reliability, while the predominant transition over transversion mutations aligns with DArTseq studies [26], further supporting marker quality.

Polymorphisms are allelic or chromosomal variations that underpin genetic diversity [27]. The high proportion of polymorphic loci indicated substantial genetic variability within the panel. Mean polymorphic information content showed moderate informativeness, consistent with bi-allelic SNP expectations [28], supporting their reliability for diversity and population structure analyses. The heterozygosity deficit (uHe>Ho) may be attributed to inbreeding, population structure, or self-pollinating nature of rice. Although autogamy reduces individual heterozygosity, *Oryza sativa* maintains considerable allelic diversity at the population level due to its complex domestication and evolutionary history [29,30]. These findings align with previous reports [31,32], confirming the expected genetic structure of self-pollinating rice.

Principal coordinate analysis (PCoA) and the neighbor-joining (NJ) tree are used to assess genetic relationships among genotypes [33]. PCoA reduced SNP dimensionality and grouped genotypes into three distinct clusters, supported by a high average Manhattan dissimilarity index. These clusters likely reflected evolutionary divergence, historical selection, or breeding practices shaping the panel's genetic structure. The NJ tree confirmed the PCoA grouping, indicating robust genetic relationships. These findings align with author [26], who identified three groups among 94 rice genotypes using principal component analysis, and author [34], who classified 365 rice landraces into two groups using PCoA. Despite differences in methods and population size, the results highlight the reliability of ordination-based approaches in revealing population structure. The clusters represent distinct genetic pools useful for parental selection and broadening the genetic base.

 doi:10.20944/preprints202604.1491.v1



Model-based admixture analysis estimates ancestry and allows partial membership across populations, making it effective for resolving germplasm with mixed breeding histories [35]. Using a 60% ancestry threshold (Q > 60%), genotypes were grouped into four subpopulations with 20% classified as admixed. The fourth subpopulation, undetected by PCoA and the NJ tree, highlighted the method's sensitivity to subtle genetic structure. The admixed group reflected mixed ancestry from historical gene flow and recombination, making these genotypes valuable for breeding and association studies due to enhanced detection of loci controlling key agronomic traits [36]. Similar thresholds (Q > 60%) have been used to classify rice genotypes into subpopulations and admixed groups [21,23], supporting this approach for population structure analysis.

Differences between distance and model-based approaches are expected. Distance methods capture broad genetic divergence, whereas admixture analysis partitions ancestry to detect finer-scale structure [37]. Similar inconsistencies have been reported in rice [38]. Thus, these methods are complementary: distance-based clustering identifies divergent parents, while admixture analysis reveals ancestry, gene flow, and stratification, essential for introgression and association mapping [39,40].

Clustering of most improved lines within Subpopulation I suggested a narrow genetic base among breeding materials, although overlap with landraces indicated potential introgression. Subpopulation III showed the highest heterozygosity and nucleotide diversity, indicating potential as an allelic reservoir; however, these estimates should be interpreted cautiously due to the small sample size (n = 3), which may increase sampling bias as noted by authors [41] and [42]. Nevertheless, evidence from maize suggests that heterozygosity and genetic differentiation ($F_{ST}$) can be reliably estimated using high-density SNP datasets (>1000), with few individuals, although inbreeding coefficients remain less reliable [43].

Based on conventional $F_{ST}$ thresholds proposed by author [44], values of 0.15 – 0.25 and >0.25 indicate moderate and strong genetic differentiation, respectively [45]. The observed $F_{ST}$ value (0.267) therefore indicated strong differentiation among subpopulations, confirming substantial genetic structure within the panel. Pairwise comparisons further revealed high differentiation between Subpopulations I and IV, suggesting their suitability as divergent parental pools for breeding. In contrast, the low differentiation among Subpopulations II, III, and the admixed group indicated closer genetic relatedness, with most genotypes from these groups clustering together in the PCoA and NJ tree. Overall, these results confirm structured genetic diversity within the panel, a characteristic feature of *Oryza sativa* germplasm that has been widely reported in previous studies [26,46,47].

The population structure was further supported by the analysis of molecular variance, which revealed that most genetic variation occurred within rather than among subpopulations. Although *Oryza sativa* is predominantly self-pollinating with low heterozygosity, the high within subpopulations variation likely reflected allelic richness rather than heterozygosity per se. This pattern is typical of selfing species, where diversity is partitioned into homozygous lineages, and population differentiation is mostly driven by allele frequency differences [48]. The high within-subpopulation variation may also reflect historical admixture, introgression, and the use of diverse germplasm in breeding [49]. Nevertheless, the significant among subpopulations variance indicated meaningful differentiation, consistent with the observed pairwise fixation index. Overall, these patterns highlight structured diversity of the panel, providing divergent parents and a broad genetic base for selection. These findings align with trends reported in other rice germplasm studies [34,37].

## 4. Materials and Methods

### *4.1. Plant Materials*

Forty lowland rainfed rice genotypes were evaluated in this study (Table 5). Among these, 34 were landraces cultivated for many years in the Zambezi valley of central Mozambique, while six were improved lines introduced by the International Rice Research Institute and AfricaRice. All genotypes are conserved at the Regional Centre for Rice Leadership and Research (CLIPA) of the

 doi:10.20944/preprints202604.1491.v1



Mozambique Agricultural Research Institute (IIAM), located in Namacurra, Zambezia, Mozambique (17°29′35.1″ S, 37°00′34.0″ E).

**Table 5.** List of forty rice genotypes evaluated in the present study.

| Entry No. | Genotype Name | Origin | Type | Notes |
|---|---|---|---|---|
| G02 | Fardamento | IIAM | Landrace | Rainfed lowland |
| G06 | Mpulo | IIAM | Landrace | Rainfed lowland |
| G07 | Mamima | IIAM | Landrace | Rainfed lowland |
| G11 | Mucabo | IIAM | Landrace | Rainfed lowland |
| G12 | Nhacungo | IIAM | Landrace | Rainfed lowland |
| G14 | Muindeia | IIAM | Landrace | Rainfed lowland |
| G17 | Paulo | IIAM | Landrace | Rainfed lowland |
| G18 | Chinchurica | IIAM | Landrace | Rainfed lowland |
| G19 | Ercidji | IIAM | Landrace | Rainfed lowland |
| G21 | Muluabo | IIAM | Landrace | Rainfed lowland |
| G25 | Sabuadigae | IIAM | Landrace | Rainfed lowland |
| G27 | Mucamba | IIAM | Landrace | Rainfed lowland |
| G33 | Nene | IIAM | Landrace | Rainfed lowland |
| G34 | Canduacafri | IIAM | Landrace | Rainfed lowland |
| G36 | Angelo | IIAM | Landrace | Rainfed lowland |
| G38 | Mucandra | IIAM | Landrace | Rainfed lowland |
| G39 | Nasoco | IIAM | Landrace | Rainfed lowland |
| G40 | Nasaia | IIAM | Landrace | Rainfed lowland |
| G41 | Mwenhe | IIAM | Landrace | Rainfed lowland |
| G42 | Mutanzania | IIAM | Landrace | Rainfed lowland |
| G44 | Mexoeira | IIAM | Landrace | Rainfed lowland |
| G45 | Bridhan P-14 | IIAM | Landrace | Rainfed lowland |
| G48 | Sinabibi | IIAM | Landrace | Rainfed lowland |
| G49 | Simao | IIAM | Landrace | Rainfed lowland |
| G50 | Namapupa | IIAM | Landrace | Rainfed lowland |
| G51 | Tacabina | IIAM | Landrace | Rainfed lowland |
| G52 | Chupa | IIAM | Landrace | Rainfed lowland |
| G53 | Agulha | IIAM | Landrace | Rainfed lowland |
| G54 | Carrungo | IIAM | Landrace | Rainfed lowland |
| G55 | Indamula | IIAM | Landrace | Rainfed lowland |
| G56 | Balachao | IIAM | Landrace | Rainfed lowland |
| G58 | Vitinho | IIAM | Landrace | Rainfed lowland |
| G59 | Aviao Branco | IIAM | Landrace | Rainfed lowland |
| G60 | Namurawani | IIAM | Landrace | Rainfed lowland |
| G62 | B1P15 | Africa rice | Line | Rainfed lowland |
| G65 | B1P02 | Africa rice | Line | Rainfed lowland |
| G66 | B1P11 | Africa rice | Line | Rainfed lowland |
| G67 | B1P01 | Africa rice | Line | Rainfed lowland |
| G68 | IRB1P21 | IRRI | Line | Rainfed lowland |
| G69 | IRB1P26 | IRRI | Line | Rainfed lowland |

*4.2. Genotyping*

- Sample Preparation and DNA Extraction

Five grams of rice seeds per genotype were submitted to SEQART AFRICA for genotyping (International Livestock Research Institute, Nairobi, Kenya). Genomic DNA was extracted from young leaf tissue using the NucleoMag Plant DNA extraction kit (Macherey-Nagel, Düren, Germany)

 doi:10.20944/preprints202604.1491.v1



per manufacturer protocols. The concentration of the extracted DNA ranged between 50 and 100 ng/μL. DNA quality and quantity were assessed by electrophoresis on 0.8% agarose gel.

- Library preparation and sequencing

Genotyping libraries were prepared using the DArTseq™ Genotyping-by-Sequencing platform, based on the complexity reduction method [50]. Genomic DNA was digested with PstI and MseI restriction enzymes, followed by ligation of barcoded adapters to the digested fragments. Adapter-ligated fragments were amplified by polymerase chain reaction (PCR) to selectively enrich fragments and generate sequencing-ready libraries. The amplified libraries were sequenced using single-read runs (138 cycles) on the NovaSeq X platform (Illumina Inc.).

- SNP calling and alignment to rice reference genome

Raw reads were processed using DArTsoft14 [50]. After filtering and demultiplexing, identical sequences were collapsed into unique tags and aligned to identify SilicoDArT and SNP markers. SilicoDArT and SNP Markers were scored in binary format (1 = presence; 0 = absence) of the restriction fragment with the marker sequence in genomic representation of the sample. Additionally, markers were aligned to the Rice_v9 reference genome, originally obtained from the Phytozome database, to determine their chromosomal positions and distribution.

- SNP filtering and quality control

Raw SNP data in Variant Call Format (VCF) were imported into R using the vcfR package version 1.15.0 [51]. The VCF file contained 70 genotypes; however, only 40 genotypes (Table 5), previously used in a drought tolerance screening study by authors [17] were extracted for further analysis. Criteria for quality control included removal of unknown chromosomes, non-informative monomorphic loci, and loci with more than 10% missing data. SNP markers with minor allele frequency >0.05 and call rate >95% were retained (Supplementary File S1; Figures S1 and S2).

*4.3. Data Analysis*

SNP markers distribution across the 12 rice chromosomes was characterized, and marker quality was evaluated based on polymorphic information content (PIC), and call rate using custom R functions. Genetic diversity was assessed using standard parameters, including observed heterozygosity ($H_o$), unbiased expected heterozygosity (uHe) [52], minor allele frequency, and percentage of polymorphic loci using the adegenet package version 2.1.11 [53]. Genetic relationships among genotypes were examined using the Manhattan dissimilarity index, followed by principal coordinate analysis [54] and construction of a neighbor-joining tree [55] using the ape package version 5.8.1 [56].

Population structure and admixture were inferred using sparse non-negative matrix factorization to estimate the ancestry of the 40 rice genotypes, implemented in the LEA package version 3.22.0 [57]. Genotypes were assigned to subpopulations based on their maximum ancestry coefficient (Q); those with Q > 60% were classified into a specific subpopulation, while genotypes with Q ≤ 60% were considered admixed. The optimum number of subpopulation (K) was determined by testing K = 1 – 10, with 10 iterations per K value, and selecting the K with the lowest cross-entropy criterion.

The following parameters were estimated for each subpopulation: observed heterozygosity and unbiased expected heterozygosity using the adegenet package, inbreeding coefficient ($F_{IS}$) and pairwise differentiation ($F_{ST}$) [58], Nei's standard genetic distance [59] using the hierfstat package version 0.5.11 [60], while nucleotide diversity [61] was calculated using the pegas package version 1.4 [62]. Genetic differentiation among populations was assessed using analysis of molecular variance [63], implemented in the poppr version 2.9.8 [64]. Data analyses were performed in R version 4.5.1 [65] (Supplementary File S2; R scripts).

 doi:10.20944/preprints202604.1491.v1



## 5. Conclusions

This study demonstrated that DArTseq SNP markers effectively revealed moderate genetic diversity among Mozambican rice genotypes, defining four distinct subpopulations with 20% admixture. Significant differentiation among subpopulations, together with substantial allelic variation within genotypes, provides a rich foundation for crop improvement. Subpopulations I and IV represent promising parental pools for enhancing heterosis and broadening the genetic base. The observed population structure likely reflects the combined effects of selection pressure, historical divergence, and gene flow. Overall, this study provides a framework for the conservation, management, and strategic use of rice germplasm in Mozambique. The findings offer practical guidance for parent selection and the development of cultivars adapted to lowland rainfed systems, supporting efforts to improve rice productivity and food security in Mozambique.

**Supplementary Materials:** The following supporting information can be downloaded at: Preprints.org, Supplementary File S1: Additional Results; Supplementary File S2: R Scripts used for data analysis.

**Author Contributions:** Conceptualization, methodology, data curation, software, formal analysis, and visualization, K.G.W.; validation, resources, and supervision, C.M.M. and L.D.M.; investigation, writing—original draft preparation, and writing—review and editing, K.G.W; C.M.M. and L.D.M; project administration, and funding acquisition, C.M.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Centre of Excellence in Agri-Food Systems and Nutrition (CE-AFSN), Eduardo Mondlane University, Maputo, Mozambique, through the World Bank, Grant number E089-MZ.

**Data Availability Statement:** All data presented or analyzed in the study are included in the article and Supplementary Material. Further inquiries can be directed at the corresponding author.

**Acknowledgments:** The authors thank the Mozambican Agricultural Research Institute (IIAM), Namacurra centre, for technical and infrastructure support. We extend our appreciation to Eng. Blessing Eduardo, and Ms. Otilia Jose, for logistic support.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| SNP | Single Nucleotide Polymorphism |
| AMOVA | Analysis of Molecular Variance |
| Mb | Megabase |

## References

1. Zilberman, D.; Goetz, R.; Garrido, A.; Otsuka, K.; Mano, Y.; Takahashi, K. *Natural Resource Management and Policy Series Editors: Rice Green Revolution in Sub-Saharan Africa.*; Otsuka, K., Mano, Y., Takahashi, K., Eds.; Springer Nature: Singapore, **2023**; Vol. 56; ISBN 978-981-19-8045-9, doi.org/10.1007/978-981-19-8046-6.
2. Traoré, F.; Debucquet, D.L.; Del Prete, D.; Sánchez, M.; Diop, I. *The Rice Value Chain in Africa. In Africa Agriculture Trade Monitor 2025, Eds. Sunday Odjo, Fousseini Traore, and Chahir Zaki. Chapter 3, Pp. 85-118.*; Kigali, **2025**; doi.org/10.54067/9798991636919.
3. Arouna, A.; Fatognon, I.A.; Saito, K.; Futakuchi, K. Moving toward Rice Self-Sufficiency in Sub-Saharan Africa by 2030: Lessons Learned from 10 Years of the Coalition for African Rice Development. *World Dev. Perspect.* **2021**, *21*, doi:10.1016/j.wdp.2021.100291.
4. FAOSTAT Statistical Database. Food and Agriculture Organization of the United Nations, Rome. Available online: https://www.fao.org/faostat/en/#data/QCL (accessed on 10 January 2026).

 doi:10.20944/preprints202604.1491.v1




5. Adjah, K.L.; Asante, M.D.; Toure, A.; Aziadekey, M.; Amoako-Andoh, F.O.; Frei, M.; Diallo, Y.; Agboka, K. Improvement of Rice Production under Drought Conditions in West Africa: Application of QTLs in Breeding for Drought Resistance. *Rice Sci.* **2022**, *29*, 512–521, doi:10.1016/j.rsci.2022.06.002.
6. Ismael, F.; Ndayiragije, A.; Fangueiro, D. New Fertilizer Strategies Combining Manure and Urea for Improved Rice Growth in Mozambique. **2021**, doi:10.3390/agronomy.
7. Kajisa, K.; Vu, T.T. The Importance of Farm Management Training for the African Rice Green Revolution: Experimental Evidence from Rainfed Lowland Areas in Mozambique. *Food Policy* **2023**, *114*, doi:10.1016/j.foodpol.2022.102401.
8. Saito, K.; Senthilkumar, K.; Dossou-Yovo, E.R.; Ali, I.; Johnson, J.M.; Mujawamariya, G.; Rodenburg, J. Status Quo and Challenges of Rice Production in Sub-Saharan Africa. *Plant Prod. Sci.* **2023**, *26*, 320–333, doi:10.1080/1343943X.2023.2241712.
9. Kodama, W.; Pede, V.O.; Mishra, A.K.; Cuevas, R.P.O.; Ndayiragije, A.; Cabrera, E.R.; Langa, M.; Ali, J. Assessing the Benefits of Green Super Rice in Sub-Saharan Africa: Evidence from Mozambique. *Q Open* **2022**, *2*, doi:10.1093/qopen/qoac006.
10. Kehinde, B.O.; Xie, L.; Song, B.-K.; Zheng, X.; Fan, L. African Cultivated, Wild and Weedy Rice (Oryza Spp.): Anticipating Further Genomic Studies. *Biology (Basel).* **2024**, *13*, 697, doi:10.3390/biology13090697.
11. Wambugu, P.W.; Furtado, A.; Le Waters, D.; Nyamongo, D.O.; Henry, R.J. Conservation and Utilization of African Oryza Genetic Resources. *Rice* **2013**, *6*, 1–13, doi:10.1186/1939-8433-6-29.
12. Fan, J.; Venuprasad, R.; Xia, S.; Yang, Z.; Zheng, X.; Chen, F. Strengthening Global Rice Germplasm Sharing: Insights from the INGER Platform **2025**, doi.org/10.21203/rs.3.rs-7477756/v1.
13. IRRI International Rice Research Institute (IRRI) and AfricaRice Collaborate on Developing and Delivering Improved and Localized Rice Varieties to Smallholder Farmers in Africa Available online: https://www.irri.org/news-and-events/news/irri-and-africarice-collaborate-developing-and-delivering-improved-and? (accessed on 18 March 2026).
14. Nangombe, S.; Macharia, M.; Solemanegy, M.; Caproni, L.; Takele, R.; Munisse, P.; Amane, M.; Buizza, R. *Report on the Characterization of Climate-Ready Rice and Cowpea Varieties.*; Research and Innovation Action, **2023**, 66 pages. CSIR.
15. APPSA Facilitates Collection and Characterization of Rice Landraces as Genetic Resources for Enhancing Resilience to Climate Change and Biotic Stresses in Mozambique Available online: https://www.ccardesa.org/appsa-facilitates-collection-and-characterization-rice-landraces-genetic-resources-enhancing? (accessed on 18 March 2026).
16. Mropes, S.K.M.N.; Moiana, L.D.; Alberto, L.A.; Andreque, J.M.; Abade, H. Morphological Characterization of 18 Varieties of Local Rice in Mozambique. *Aust. J. Crop Sci.* **2026**, *20*, 188–197, doi:10.21475/ajcs.26.20.02.p383x.
17. Warioba, K.G.; Macandza, C.M.; Moiana, L.D. Screening Rice (Oryza Sativa L.) Genotypes for Seedling-Stage Drought Tolerance. *Stresses* **2026**, *6*, 13, doi:10.3390/stresses6010013.
18. 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. *International Journal of Plant Biology* **2024**, *15*, 607–640, doi.org/10.3390/ijpb15030046.
19. Sarif, H.M.; Rafii, M.Y.; Ramli, A.; Oladosu, Y.; Musa, H.M.; Rahim, H.A.; Zuki, Z.M.; Chukwu, S.C. Genetic Diversity and Variability among Pigmented Rice Germplasm Using Molecular Marker and Morphological Traits. *Biotechnology and Biotechnological Equipment* **2020**, *34*, 747–762, doi:10.1080/13102818.2020.1804451.
20. Yu, R.; Liu, J.; Niu, Y.; Han, X.; Wang, X.; Yang, Y. A Simple and Flexible Approach for Detecting Small Numbers of SNPs. *Front. Plant Sci.* **2025**, *16*, doi:10.3389/fpls.2025.1748099.
21. Berdugo-Cely, J.A.; Pérez-Pazos, J. V.; Perez-Cantero, S.P.; Morales-Angulo, J.G.; Romero-Ferrer, J.L. Genetic Diversity and Population Structure of Regional Rice Genotypes from Colombia's Caribbean and Pacific Regions: Differentiation and Ancestry in Relation to the 3000 Rice Genomes Project. *Genet. Resour. Crop Evol.* **2025**, doi:10.1007/s10722-025-02549-y.
22. Raza, Q.; Riaz, A.; Saher, H.; Bibi, A.; Raza, M.A.; Ali, S.S.; Sabar, M. Grain Fe and Zn Contents Linked SSR Markers Based Genetic Diversity in Rice. *PLoS One* **2020**, *15*, doi:10.1371/journal.pone.0239739.

 doi:10.20944/preprints202604.1491.v1




23. Gouda, A.C.; Sangare, J.R.; Gnikoua, K.; Wambugu, P.; Huggins, T.D.; Ndjiondjop, M.N. Genetic Variation and Population Structure of the Rice Accessions Maintained in the AfricaRice Genebank Using DArTseq. *Crop Sci.* **2024**, doi:10.1002/csc2.21395.
24. Ndjiondjop, M.N.; Gouda, A.C.; Eizenga, G.C.; Warburton, M.L.; Kpeki, S.B.; Wambugu, P.W.; Gnikoua, K.; Tia, D.D.; Bachabi, F. Genetic Variation and Population Structure of Oryza Sativa Accessions in the AfricaRice Collection and Development of the AfricaRice O. Sativa Core Collection. *Crop Sci.* **2023**, *63*, 724–739, doi:10.1002/csc2.20898.
25. Ndjiondjop, M.N.; Semagn, K.; Sow, M.; Manneh, B.; Gouda, A.C.; Kpeki, S.B.; Pegalepo, E.; Wambugu, P.; Sié, M.; Warburton, M.L. Assessment of Genetic Variation and Population Structure of Diverse Rice Genotypes Adapted to Lowland and Upland Ecologies in Africa Using SNPs. *Front. Plant Sci.* **2018**, *9*, doi:10.3389/fpls.2018.00446.
26. Kimwemwe, P.K.; Bukomarhe, C.B.; Mamati, E.G.; Githiri, S.M.; Civava, R.M.; Mignouna, J.; Kimani, W.; Fofana, M. Population Structure and Genetic Diversity of Rice (Oryza Sativa L.) Germplasm from the Democratic Republic of Congo (DRC) Using DArTseq-Derived Single Nucleotide Polymorphism (SNP). *Agronomy* **2023**, *13*, doi:10.3390/agronomy13071906.
27. Serrote, C.M.L.; Reiniger, L.R.S.; Silva, K.B.; Rabaiolli, S.M. dos S.; Stefanel, C.M. Determining the Polymorphism Information Content of a Molecular Marker. *Gene* **2020**, *726*, doi:10.1016/j.gene.2019.144175.
28. Razak, S.A.; Azman, N.H.E.N.; Kamaruzaman, R.; Saidon, S.A.; Yusof, M.F.M.; Ismail, S.N.; Jaafar, M.A.; Abdullah, N. Genetic Diversity of Released Malaysian Rice Varieties Based on Single Nucleotide Polymorphism Markers. *Czech Journal of Genetics and Plant Breeding* **2020**, *56*, 62–70, doi:10.17221/58/2019-CJGPB.
29. Hodgins, K.A.; Yeaman, S. Mating System Impacts the Genetic Architecture of Adaptation to Heterogeneous Environments. *New Phytologist* **2019**, *224*, 1201–1214, doi:10.1111/nph.16186.
30. Choi, J.Y.; Purugganan, M.D. Multiple Origin but Single Domestication Led to Oryza Sativa. *G3: Genes, Genomes, Genetics* **2018**, *8*, 797–803, doi:10.1534/g3.117.300334.
31. Choudhury, D.R.; Kumar, R.; Vimala Devi, S.; Singh, K.; Singh, N.K.; Singh, R. Identification of a Diverse Core Set Panel of Rice From the East Coast Region of India Using SNP Markers. *Front. Genet.* **2021**, *12*, doi:10.3389/fgene.2021.726152.
32. Mvuyekure, S.M.; Sibiya, J.; Derera, J.; Nzungize, J.; Nkima, G. Assessment of Genetic Diversity of Rice Based on SNP Markers for Selection of Parents for Sheath Rot (Sarocladium Oryzae) Resistance Breeding. *South African Journal of Plant and Soil* **2018**, *35*, 51–59, doi:10.1080/02571862.2017.1333636.
33. Baksh, S.K.Y.; Donde, R.; Kumar, J.; Mukherjee, M.; Meher, J.; Behera, • Lambodar; Sushanta, •; Dash, K. Genetic Relationship, Population Structure Analysis and Pheno-Molecular Characterization of Rice (Oryza Sativa L.) Cultivars for Bacterial Leaf Blight Resistance and Submergence Tolerance Using Trait Specific STS Markers. *Physiology and Molecular Biology of Plants* **2021**, *27*, 543–562, doi:10.1007/s12298.
34. Aesomnuk, W.; Ruengphayak, S.; Ruanjaichon, V.; Sreewongchai, T.; Malumpong, C.; Vanavichit, A.; Toojinda, T.; Wanchana, S.; Arikit, S. Estimation of the Genetic Diversity and Population Structure of Thailand's Rice Landraces Using Snp Markers. *Agronomy* **2021**, *11*, doi:10.3390/agronomy11050995.
35. Skotte, L.; Korneliussen, T.S.; Albrechtsen, A. Estimating Individual Admixture Proportions from next Generation Sequencing Data. *Genetics* **2013**, *195*, 693–702, doi:10.1534/genetics.113.154138.
36. Ghazy, M.I.; EL-Naem, S.A.; Hefeina, A.G.; Sallam, A.; Eltaher, S. Genome-Wide Association Study of Rice Diversity Panel Reveals New QTLs for Tolerance to Water Deficit Under the Egyptian Conditions. *Rice* **2024**, *17*, doi:10.1186/s12284-024-00703-1.
37. Choudhury, D.R.; Kumar, R.; Maurya, A.; Semwal, D.P.; Rathi, R.S.; Gautam, R.K.; Trivedi, A.K.; Bishnoi, S.K.; Ahlawat, S.P.; Singh, K.; et al. SSR and SNP Marker-Based Investigation of Indian Rice Landraces in Relation to Their Genetic Diversity, Population Structure, and Geographical Isolation. *Agriculture (Switzerland)* **2023**, *13*, doi:10.3390/agriculture13040823.
38. Kumar, K.P.; Pushpam, R.; Manonmani, S.; Raveendran, M.; Santhiya, S.; Senthil, A. Enhancing Stress Resilience in Rice (Oryza Sativa L.) through Profiling Early-Stage Morpho-Physiological and Molecular Responses to Multiple Abiotic Stress Tolerance. *Front. Plant Sci.* **2024**, *15*, doi:10.3389/fpls.2024.1342441.

 doi:10.20944/preprints202604.1491.v1




39. Sun, Z.; Peng, J.; Lv, Q.; Ding, J.; Chen, S.; Duan, M.; He, Q.; Wu, J.; Tian, Y.; Yu, D.; et al. Dissecting the Genetic Basis of Heterosis in Elite Super-Hybrid Rice. *Plant Physiol.* **2023**, *192*, 307–325, doi:10.1093/plphys/kiad078.
40. McCouch, S.R.; Wright, M.H.; Tung, C.W.; Maron, L.G.; McNally, K.L.; Fitzgerald, M.; Singh, N.; DeClerck, G.; Agosto-Perez, F.; Korniliev, P.; et al. Open Access Resources for Genome-Wide Association Mapping in Rice. *Nat. Commun.* **2016**, *7*, doi:10.1038/ncomms10532.
41. Gouda, A.C.; Ndjiondjop, M.N.; Djedatin, G.L.; Warburton, M.L.; Goungoulou, A.; Kpeki, S.B.; N'Diaye, A.; Semagn, K. Comparisons of Sampling Methods for Assessing Intra- and Inter-Accession Genetic Diversity in Three Rice Species Using Genotyping by Sequencing. *Sci. Rep.* **2020**, *10*, doi:10.1038/s41598-020-70842-0.
42. Fumagalli, M. Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. *PLoS One* **2013**, *8*, doi:10.1371/journal.pone.0079667.
43. Aguirre-Liguori, J.A.; Luna-Sánchez, J.A.; Gasca-Pineda, J.; Eguiarte, L.E. Evaluation of the Minimum Sampling Design for Population Genomic and Microsatellite Studies: An Analysis Based on Wild Maize. *Front. Genet.* **2020**, *11*, doi:10.3389/fgene.2020.00870.
44. Wright, S. *Evolution and the Genetics of Populations: Variability within and among Natural Populations*; University of Chicago Press, **1978**; Vol. 4;
45. Thant, A.A.; Zaw, H.; Kalousova, M.; Singh, R.K.; Lojka, B. Genetic Diversity and Population Structure of Myanmar Rice (Oryza Sativa L.) Varieties Using DArTseq-Based SNP and SilicoDArT Markers. *Plants* **2021**, *10*, doi:10.3390/plants10122564.
46. Singh, A.K.; Kumar, D.; Gemmati, D.; Ellur, R.K.; Singh, A.; Tisato, V.; Dwivedi, D.K.; Singh, S.K.; Kumar, K.; Khan, N.A.; et al. Investigating Genetic Diversity and Population Structure in Rice Breeding from Association Mapping of 116 Accessions Using 64 Polymorphic SSR Markers. *Crops* **2024**, *4*, 180–194, doi:10.3390/crops4020014.
47. Suvi, W.T.; Shimelis, H.; Laing, M.; Mathew, I.; Shayanowako, A.I.T. Assessment of the Genetic Diversity and Population Structure of Rice Genotypes Using SSR Markers. *Acta Agric. Scand. B Soil Plant Sci.* **2020**, *70*, 76–86, doi:10.1080/09064710.2019.1670859.
48. Hartfield, M.; Bataillon, T.; Glémin, S. The Evolutionary Interplay between Adaptation and Self-Fertilization. *Trends in Genetics* **2017**, *33*, 420–431, doi.org/10.1016/j.tig.2017.04.002.
49. Anwar, A.; Tabassum, J.; Ahmad, S.; Ashfaq, M.; Hussain, A.; Ullah, M.A.; Saad, N.S.B.M.; Ghazy, A.I.; Javed, M.A. Screening and Assessment of Genetic Diversity of Rice (Oryza Sativa L.) Germplasm in Response to Soil Salinity Stress at Germination Stage. *Agronomy* **2025**, *15*, doi:10.3390/agronomy15020376.
50. Kilian, A.; Wenzl, P.; Huttner, E.; Carling, J.; Xia, L.; Blois, H.; Caig; Hok, P.; Uszynski, G. Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In *Data Production and Analysis in Population Genomics* ; Pompanon, F., Bonin, A., Eds.; **2012**; Vol. 888, pp. 67–89, doi.org/10.1007/978-1-61779-870-2_5.
51. Knaus, B.J.; Grünwald, N.J. VCFR: A Package to Manipulate and Visualize Variant Call Format Data in R. *Mol. Ecol. Resour.* **2017**, *17*, 44–53, doi:10.1111/1755-0998.12549.
52. Nei, M. Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics* **1978**, 583–590, doi:10.1093/genetics/89.3.583.
53. Jombart, T.; Ahmed, I. Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data.Bioinformatics. **2011**, doi:10.1093/bioinformatics/btr52.
54. Torgerson, W.S. Multidimensional Scaling: I. Theory and Method. *Psychometrika* **1952**, *17*, 401–419, doi:10.1007/BF02288916.
55. Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* **1987**, doi:10.1093/oxfordjournals.molbev.a040454.
56. Paradis, E.; Schliep, K. Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R. **2019**, *35*, 526–528, doi:10.1093/bioinformatics/bty633.
57. Frichot, E.,; Francois, O. LEA: Package for Landscape and Ecological Association Studies. Methods in Ecology and Evolution 2015, http://membres-timc.imag.fr/Olivier.Francois/lea.html.

 doi:10.20944/preprints202604.1491.v1




58. Weir, B.S.; Cockerham, C.C. Estimating *F* -Statistics for the Analysis of Population Structure. *Evolution (N. Y).* **1984**, *38*, 1358–1370, doi:10.1111/j.1558-5646.1984.tb05657.x.
59. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, **1987**; ISBN 0231063544.
60. Goudet, J.; Jombart, T. Hierfstat: Estimation and Tests of Hierarchical F-Statistics **2022**, https://CRAN.R-project.org/package=hierfsta.
61. Nei, M.; Li, W.H. Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases. *Proceedings of the National Academy of Sciences* **1979**, *76*, 5269–5273, doi:10.1073/pnas.76.10.5269.
62. Paradis, E. Pegas: An R Package for Population Genetics with an Integrated–Modular Approach. *Bioinformatics* **2010**, *26*, 419–420, doi:10.1093/bioinformatics/btp696.
63. Excoffier, L.; Smouse, P.E.; Quattro, J.M. Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* **1992**, *131*, 479–491, doi:10.1093/genetics/131.2.479.
64. Kamvar, Z.N.; Brooks, J.C.; GrÃ¼nwald, N.J. Novel R Tools for Analysis of Genome-Wide Population Genetic Data with Emphasis on Clonality. *Front. Genet.* **2015**, *6*, doi:10.3389/fgene.2015.00208.
65. R Core Team. R: A Language and Environment for Statistical Computing (Version 4.5.1). R Foundation for Statistical Computing, Vienna, Austria. Available online: https://www.R-project.org/ (accessed on 19 October 2025).