Received: 19 February 2024 | Accepted: 26 September 2024

DOI: 10.1002/csc2.21395

Crop Science

ORIGINAL ARTICLE

Plant Genetic Resources

Genetic variation and population structure of the rice accessions maintained in the AfricaRice genebank using DArTseq

Arnaud Comlan Gouda[1] | **Jean Rodrigue Sangare**[2] | **Karlin Gnikoua**[1] | **Peterson Wambugu**[3] | **Trevis D. Huggins**[4] | **Marie Noelle Ndjiondjop**[1]

[1]Africa Rice Center (AfricaRice), M'bé Research Station, Bouaké, Cote d'Ivoire

[2]Institut d'Economie Rural, Station de Sikasso, Sikasso, Mali

[3]Kenya Agricultural and Livestock Research Organization (KALRO), Nairobi, Kenya

[4]USDA-ARS, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas, USA

Correspondence
Marie Noelle Ndjiondjop, Africa Rice Center (AfricaRice), M'bé Research Station, 01 B.P. 2551, Bouaké 01, Cote d'Ivoire.
Email: m.ndjiondjop@cgiar.org

Assigned to Associate Editor Laura Shannon.

Funding information
German Federal Ministry for Economic Cooperation and Development (BMZ) under the project titled "Rapid Action Plan of the Genebank Platform", and by the CGIAR System Organization

Abstract

Utilizing the full potential of rice collections mainly depends on an in-depth exploration and understanding of the vast diversity in its germplasm. The AfricaRice genebank holds the largest collection of rice germplasm originating from the African continent. In the present study, we comprehensively characterized a collection of 9013 accessions, including *Oryza barthii* A. Chev., *Oryza glaberrima* Steud., *Oryza longistaminata* A. Chev. & Roehr., *Oryza sativa* L. ssp. *indica*, and *Oryza sativa* L. ssp. *japonica*, for genetic diversity and population structure using genotyping-by-sequencing through DArTseq analysis. We identified 27,718 high-quality single nucleotide polymorphism markers after the genotypic data were filtered. Based on the analyses, the collection has extensive genetic diversity, and the average genetic distance of the entire set was 0.267 (range 0.001–0.469), with 45.1% of pairs of accessions being highly distant and 40.1% moderately distant from each other. Neighbor-joining tree, principal component, and Bayesian population structure analyses clustered the 9013 accessions into six groups, based roughly on their taxonomic and biological status. The first, second, and third groups consisted of accessions belonging to *O. glaberrima*, *O. barthii*, and *O. longistaminata*, respectively. The fourth, fifth, and sixth groups were improved-*indica*, *japonica*, and traditional-*indica* accessions, respectively. The highest value of genetic variance proportion (*PhiPT*) was found in the species group followed by groups based on cluster analysis and on Bayesian population structure at $K = 6$. These results allow us to better understand the genetic diversity present in 9013 rice accessions maintained in the AfricaRice genebank and offer a valuable tool for pre breeding, breeding, and further genetic applications.

Abbreviations: AMOVA, analysis of molecular variance; DArT, diversity array technology; IBS, identity by state; IRRI, International Rice Research Institute; LEA, Landscape and Ecological Association; MDS, multidimensional scaling; NJ, neighbor joining; PCA, principal component analysis; SNP, single nucleotide polymorphism.

Plain Language Summary

To fully utilize rice collections, understanding their genetic diversity is essential. The AfricaRice genebank, which holds the largest African rice collection, characterized 9013 accessions from *Oryza barthii*, *Oryza glaberrima*, *Oryza longistaminata*, *Oryza sativa* ssp. indica, and *Oryza sativa* ssp. japonica using DArTseq genotyping-by-sequencing. A total of 27,718 high-quality single nucleotide polymorphism markers were identified, revealing significant genetic diversity with an average genetic distance of 0.267, and 45.1% of pairs being highly distant. Six clusters emerged from analyses, corresponding to species and biological classifications. These findings enhance the understanding of genetic diversity in the AfricaRice collection and provide valuable tools for breeding and genetic studies.

1 | INTRODUCTION

Rice is primarily a staple food in Asia, the Caribbean, and Latin America, and is becoming progressively popular in Africa (OECD-FAO, 2021). It has superior nutrient content compared to maize, wheat, and potatoes (Mohidem et al., 2022). It is also recognized as a great source of vitamin E and B5, as well as carbohydrates, thiamine, calcium, folate, and iron (Mohidem et al., 2022). For more than 3000 years, rice has been cultivated in parts of Africa and is grown in about 40 out of 54 African countries (Arouna et al., 2020). Despite comprising only 13% of the global population, Africa constitutes 32% of worldwide rice imports, establishing it as a significant participant in the global rice trade. For instance, the growth in annual demand for rice consumption is faster than for any other staple food in sub-Saharan Africa with a rate of about 6% (Soullier et al., 2020). However, levels of domestic rice production lag behind the growth in consumption for various reasons, including population growth, urbanization, and changes in consumer preferences (Sekiya et al., 2020). Thus, the continent fills the expanding gap between supply and demand through massive imports from Asia (Falcon et al., 2022). To ensure food and nutrition security in Africa, it is imperative that research and breeding efforts not only focus on developing climate-resilient rice varieties but also on enhancing yield and addressing consumer preferences in quality and nutrition. Developing resilient varieties requires the availability of and accessibility to diverse and ecologically adapted germplasm with a broad genetic base (Mba & Ogbonnaya, 2022; Whitfield et al., 2021).

Genetic diversity is a basic component of biodiversity and its conservation is essential for the long-term survival of any species in changing environments (J. Kumar & Agrawal, 2019; Quiones-Prez et al., 2014). Large amounts of plant genetic resources of different plant species are maintained in various genebank collections around the world (Mohidem et al., 2022). Genebanks play a key role in the conservation and use of untapped plant genetic diversity for crop improvement for food and nutrition security (FAO, 2014). They complement the conservation of genetic diversity in farmers' fields and in nature through long-term conservation (Weise et al., 2020).

AfricaRice genebank holds the largest collection of African rice in the world and the largest rice collection in Africa, with about 20,681 rice accessions. Nearly 83.4% of the accessions in the AfricaRice collection were collected within Africa, 14% (3130 accessions) of them being *Oryza glaberrima* while 64% (14,480 accessions) are African *Oryza sativa* (Ndjiondjop et al., 2022). This rich source of genetic variation is available and accessible to public and private plant breeding programs as well as other interested germplasm users (Halewood et al., 2020). The AfricaRice genebank germplasm may contain useful traits that could be exploited for future crop improvement programs. However, its deployment in rice improvement is constrained by lack of information on its potential genetic value due to inadequate characterization.

The value of plant genetic resources in crop genetic improvement depends on the level of genetic variation present in the germplasm and this can be assessed through a variety of approaches (MacKay et al., 2009). Assessment of genetic diversity, relationships, and structure within a given set of germplasm is an essential tool for crop improvement through (1) elucidating the diversity between parental lines before hybridization and introgression of desirable genes into elite genotypes, (2) determining the level of genetic variability when defining core subsets selected for specific traits, and (3) estimating possible loss of genetic diversity during conservation or through selection (Chang et al., 2022; Gouda et al., 2020; A. Kumar et al., 2020; Mohammadi & Prasanna, 2003; Mommer et al., 2011; Ndjiondjop et al., 2022; Reif et al., 2005).

Genetic diversity in rice has been investigated using morphological, biochemical, and DNA markers (Vanlalsanga & Singh, 2019). However, both morphological and biochemical

traits are highly influenced by the environment through genotype × environment interactions and may not allow accurate genetic analysis of the germplasm (Chesnokov & Kosolapov, 2020; S un et al., 2022). To address these challenges and to ensure effective conservation and sustainable use, molecular tools are excellent for analyzing and characterizing germplasm collections. To date, molecular markers have been extensively used to measure genetic diversity, population structure, and phylogenetic relationships, and assess population evolutionary history and conduct association studies (Konishi et al., 2006; Lakew et al., 2021; Liu et al., 2021; Montcho et al., 2021; Ndjiondjop et al., 2018a; Phung et al., 2014; Vilayheuang et al., 2020; Wang et al., 2018). Although molecular markers have proven effective, novel tools with greater capacity are needed to support the conservation and accelerated use of plant genetic diversity. Diversity arrays technology (DArT) and single-nucleotide polymorphisms (SNPs), based on next-generation sequencing, have been utilized extensively in genetic diversity studies in numerous crops including rice (Adeboye et al., 2020; Baloch et al., 2017; Gouda et al., 2020; Ndjiondjop et al., 2018a, 2018b, 2022; Ramirez-Villegas et al., 2022; Sansaloni et al., 2020).

In the present study, we characterized released cultivars, improved/breeding lines, traditional/landraces, and wild species and identified useful variations that can be efficiently utilized in rice breeding. The main goal of this study was to assess the genetic diversity and population structure of 9013 rice accessions maintained by AfricaRice genebank using high-density DArTseq SNPs.

Core Ideas

- The 9013 genotyped accessions are 42% of the AfricaRice collection and represent four species, six ecologies, and 76 countries.
- Genetic distances between pairs of accessions were variable but 58% were highly distant, implying the uniqueness of most accessions of the AfricaRice collection.
- Single nucleotide polymorphism based cluster analysis revealed six distinct groups corresponding to two *indica* groups and one *japonica* group and one individual group for each of the other species included in the study (*Oryza glaberrima*, *Oryza barthii*, and *Oryza longistaminata*).

2 | MATERIALS AND METHODS

2.1 | Genotyping

A total of 9013 rice accessions comprising improved cultivars, advanced lines, landraces/ traditional cultivar, and rice wild species were selected in a random manner from the entire collection maintained in AfricaRice genebank and used in this study. These included *O. glaberrima* (3130), *O. sativa* ssp. *indica* (4928), *O. sativa* ssp. *japonica* (707), the wild species *Oryza barthii* (163), and *Oryza longistaminata* (85), with origins from 75 countries and 11 geographical regions. They represent varied ecosystems, upland and irrigated lowland, rainfed lowland, flooding, hydromorphic, mangrove, and shallow and deep forest swamp (Table S1; Figure 1). DNA was extracted according to the protocol described previously by Ndjiondjop et al. (2022). Briefly, approximately 5 mg of fresh leaf samples were harvested from a single 3- to 4-week-old seedling and placed in a 1.1 mL MicroTubes (Bioquote Limited). The tissue was dried at 57˚C using a Binder FD53 E2 Drying oven (Akribis Scientific Limited), covered with micronic sealing mats (NBS Scientific), and shipped to DArT Pty Ltd (http://www.diversityarrays.com/) for genotyping. A total of 49,685 SNPs markers were generated from the genotyping of these 9013 *Oryza* sp accessions.

2.2 | Data filtering

Data were filtered using the procedure previously described in Ndjiondjop et al. (2022). SNPs with unknown chromosome positions were removed from the dataset. The resulting genotypic dataset of the 12 chromosomes were filtered for (1) missingness ≤25%, (2) max heterozygosity fixed at 0.5, and (3) minor allele frequency (MAF) ≥1%. The filtering returned 27,718 SNP markers that captured the diversity of the 9013 *Oryza* sp accessions.

2.3 | Statistical analysis

Statistical analyses were performed as previously described in Ndjiondjop et al. (2017, 2022). Briefly, heterozygosity, identity by state (IBS)-based genetic distance matrices, and principal component analysis (PCA) were computed using TASSEL v.5.2.58 (Bradbury et al., 2007). The first three principal components (PCs) from the PCA were plotted for visual examination in CurlyWhirly version 1.19.09.04 (The James Hutton Institute, Information & Computational Sciences) using the seven categorical variables: subspecies, germplasm type, country of origin, adapted ecology, region, continent, and predicted group memberships from phylogenetic and population structure analyses. The R package "poppr" version 2.9.3 (Kamvar et al., 2014) was used to perform the analysis of molecular variance (AMOVA) and to detect the genetic variance within and among populations using the

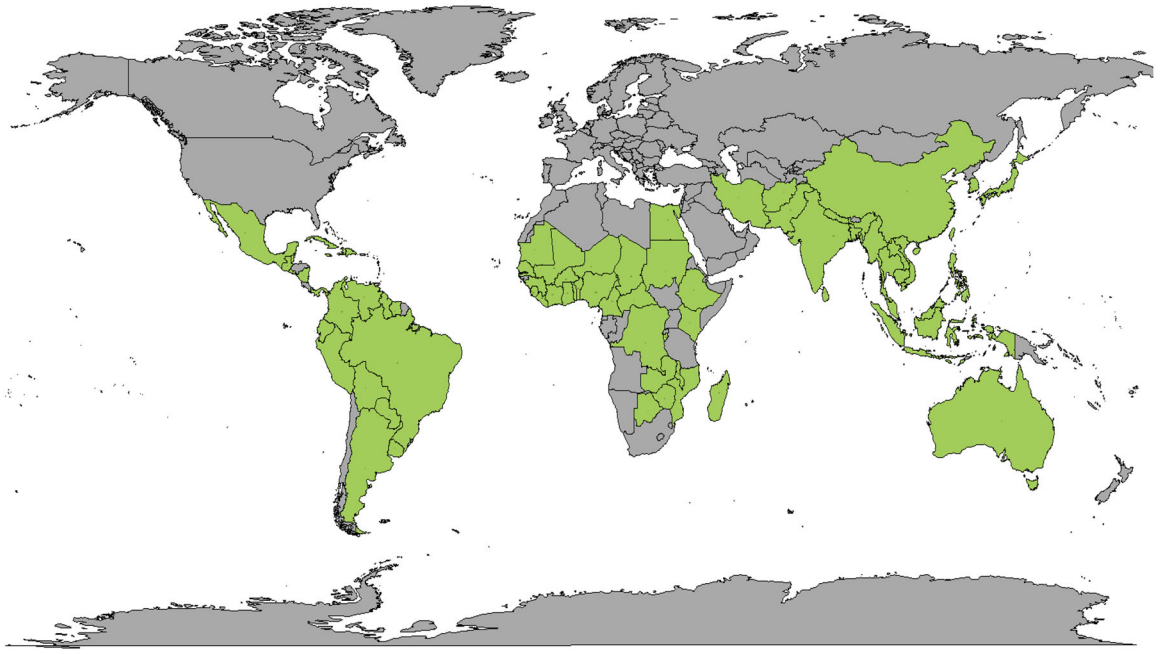

FIGURE 1 World map showing in green color, the countries of origin where the 9013 *Oryza* sp. accessions were collected.

PhiPT value (analogue of F_{ST} fixation index). Based on continent, region, country of origin, ecology, biological status, species, and group membership predicted from the phylogenetic and population structure analyses, accessions were assigned to six groups (populations). Neighbor-joining (NJ) trees were built using TASSEL v.5.2.58 software. The trees were exported in the "Newick" format and further refined in iTOL v4 online program (Letunic & Bork, 2019). The HapMap format of each dataset was exported to PHYLIP interleaved format using TASSEL v.5.2.57 and then converted to Molecular Evolutionary Genetics Analysis (MEGA) X (S. Kumar et al., 2018) format using PGDSpider v.2.1.1.3 (Lischer & Excoffier, 2012). The MEGA X software was used to measure the number of segregating sites (S), the proportion of polymorphic sites (Ps), theta (θ), and nucleotide diversity (π). A site (SNP) was considered segregating if it had two or more nucleotides at that site; π refers to the average number of pairwise nucleotide differences between two sequences (samples), while θ was used as another estimator of diversity based on the number of segregating sites in the samples. Population structure was analyzed using the R package "LEA" STRUCTURE-like method and the Admixture model (Frichot & François, 2015). A plot of cross-entropy against the number of ancestral populations was generated to determine the optimal number of ancestral populations or the best run for a fixed value of K. DNA samples and accessions with membership probabilities >60% were assigned to the same clusters (group), while those with probabilities <60% in any group were assigned to a "mixed" group (Gouda et al., 2020; Ndjiondjop et al., 2018a, 2022).

3 | RESULTS AND DISCUSSION

3.1 | Distribution of SNP markers

Genomic DNA from all 9013 rice accessions in this study was analyzed and revealed a total of 49,685 polymorphic SNPs. These single nucleotide variations were further subjected to hard filtering and resulted in a final number of 27,718 polymorphic SNPs markers (about 56% of the 49,685 SNPs) distributed across the 12 chromosomes (Table S2). The number of SNPs ranged from 1427 on chromosome 10 to 3196 on chromosome 1, with an overall average of 2309 SNPs per chromosome (Table 1). The physical length of each chromosome ranged from 16,820 kb on chromosome 10 to 33,040 kb on chromosome 3, with an overall genome size of 284,104 Mb. The average map length ranged from 9.3 (chromosome 12) to 11.7 kb (chromosome 10) indicating that the SNPs provide sufficient coverage of the whole genome (Figure 2; Table 1).

Analysis of SNPs within individual species/subspecies indicated that *O. sativa* ssp. *indica* had the highest number of polymorphic SNPs, 18,902, whereas *O. longistaminata* had the lowest number of polymorphic SNPs, 7220 (Figure 3). SNP markers unique to each species/subspecies were identified for *O. longistaminata* (466), *O. glaberrima* (324), *O. barthii* (726), *O. sativa japonica* (202), and *O. sativa indica* (2888) (Figure 3). The MAF values of the SNP markers ranged from 0.010 to 0.500 with an average of 0.191 (Table S2). These values were similar to those previously reported in earlier studies using DArTseq SNPs markers (Ndjiondjop

TABLE 1 The chromosomal distribution of 27,718 polymorphic single nucleotide polymorphisms (SNPs) used for genotyping 9013 *O. sativa* accessions, the physical length of each chromosome covered by the SNPs in kilobase (kb) pairs and average map length per SNP (kb).

Chromosome	Physical length based on 27,718 SNPs (Mb)	Number of SNPs polymorphic in 9013 *Oryza* sp	Average map length per SNP (kb)
Chr1	32,598	3196	10.1
Chr2	29,072	2984	9.7
Chr3	33,040	3023	10.9
Chr4	26,231	2526	10.3
Chr5	23,188	2045	11.3
Chr6	23,814	2356	10.1
Chr7	21,754	2225	9.7
Chr8	20,272	2101	9.6
Chr9	17,390	1625	10.7
Chr10	16,820	1427	11.7
Chr11	20,770	2167	9.5
Chr12	19,148	2043	9.3
Total	284,104	27,718	–

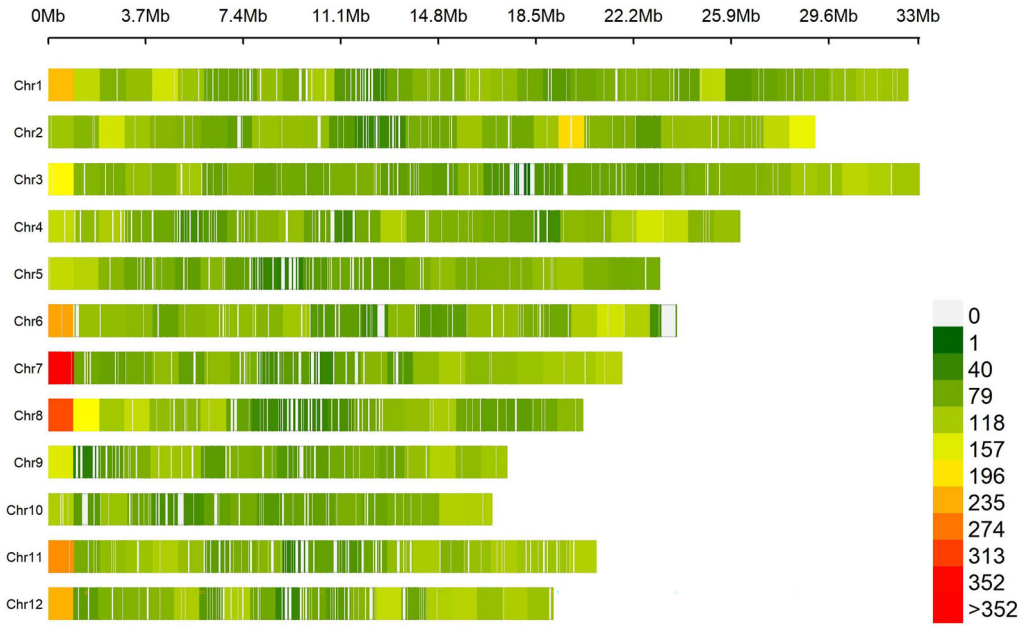

FIGURE 2 Distribution of the 27,718 high-quality single nucleotide polymorphism (SNP) markers within a 1 Mb window for each of the 12 chromosomes of the *Oryza* sp genome.

et al., 2017, 2018b, 2022). The identified SNP markers are an important resource to further investigate rice genetic diversity and for genome-wide associations studies (Razak et al., 2020; Reig-Valiente et al., 2016).

3.2 | Genetic purity

Heterozygosity is one of the most popular parameters to determine the proportion of heterozygous individuals at a locus in a population. Heterozygosity can be a useful source of genetic variation in parental lines to develop new segregating populations (Belefant-Miller et al., 2012). Accessions with high levels of heterozygosity are of great conservation value due to the high amount of genetic variation that they possess. In addition, understanding of heterozygosity is important as it helps inform conservation management decisions. Assessing patterns of heterozygosity can, for example, help in determining the suitability and effectiveness of a regeneration method in maintaining genetic diversity and

TABLE 2 Observed heterozygosity calculated across the 9013 Oryza accessions using the 27,718 polymorphic single nucleotide polymorphism (SNP) markers.

Species/subspecies	No. of accessions	Range (%)	Average (%)	No. of accessions with Het ≤5%	No. of accessions with Het >5%
Oryza sativa spp. *indica*	4928	0–15.4	1.5	4593	335
Oryza sativa spp. *japonica*	707	0–15.4	1.5	651	56
Oryza glaberrima	3130	0–9.3	0.7	3127	3
Oryza barthii	163	0–2.6	0.6	163	0
Oryza longistaminata	85	0–2.3	0.9	85	0

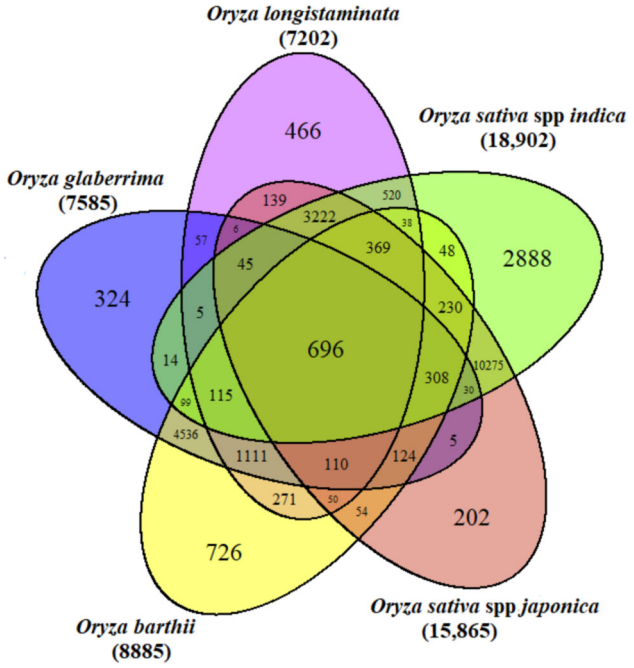

FIGURE 3 Venn diagram showing the number of unique and common divergent single nucleotide polymorphism (SNP) markers among the four *Oryza* species and two *Oryza sativa* subspecies. The values in the parentheses represent the total number of polymorphic markers for the species and subspecies used in this study.

integrity (McCouch et al., 2012). Rice is a self-pollinated crop with an outcrossing rate of between 2% and 5% (Semon et al., 2005). Heterozygosity analysis was performed on the 9013 accessions and is presented in Table S3 and Figure S1. The observed heterozygosity per accession within *O. sativa* ranged from 0% to 15.4%, with an average of 1.5%, and confirmed the recent observations published by Ndjiondjop et al. (2022). The number of *indica* and *japonica* accessions that exceeded the normal rate of heterozygosity were 335 (7.2%) and 56 (7.9%), respectively (Table 2). For the *O. glaberrima* species, the observed heterozygosity was between 0% and 9.3%, with an average of 0.7%. Three *O. glaberrima* accessions were observed to have heterozygosity greater than 5% (Table 2). The average observed heterozygosity of *O. glaberrima* accessions (0.7%) was lower than the 2.3% previously reported by Ndjiondjop et al. (2017). The difference observed between the study conducted by Ndjiondjop et al. (2017) and this study may be due to (1) the large difference in the number of accessions genotyped in each study, with 3130 accessions (this study) compared to 2179 accessions in Ndjiondjop et al. (2017); (2) the low heterozygosity rate in the 951 new accessions used in the present study, thus affecting MAF, and (3) the number of polymorphic markers used to assess the observed heterozygosity within *O. glaberrima* accessions, with 27,560 SNPs compared to 7585 SNPs in the previous study. The occurrence of low levels of heterozygosity (0.002–0.010) has also been reported in other studies focused on rice (Choudhury et al., 2014; Nachimuthu et al., 2015; Singh et al., 2013). Nachimuthu et al. (2015) suggested that low levels of heterozygosity may be due to the self-fertilizing nature of rice. Heterozygosity in accessions belonging to *O. barthii* and *O. longistaminata* species ranged from 0 to 2.6 and 0 to 2.3, respectively (Table 2). Veltman et al. (2019) reported that *O. barthii* has higher heterozygosity than *O. glaberrima*. However, the present results do not agree with Veltman et al. (2019) and can be explained by the low number of *O. barthii* accessions genotyped compared to *O. glaberrima*. All *O. longistaminata* accessions exhibited low heterozygosity and agreed with the study by Melaku et al. (2019) on *O. longistaminata* collected from Ethiopia. In sum, observed heterozygosity (Ho = 0.00–0.154) was found to be lower than expected (He = 0.241–0.289) due to the high inbreeding (F_{is} = 0.424–1) nature of rice. These findings are consistent with the conclusions reported by Kimwemwe et al. (2023), who assessed the genetic diversity of 94 rice accessions, 43 of which were provided by the AfricaRice genebank.

A genetic bottleneck occurs when a population undergoes a significant reduction in size, resulting in the loss of genetic diversity (Londo et al., 2006; Maruyama & Fuerst, 1985; Zachariah Peery et al., 2012). Wild rice species typically have restricted gene flow due to factors such as geographic isolation, ecological barriers, or differences in flowering time (Gao, 2004; Z. P. Song et al., 2003). Hence, limited gene flow and/or genetic bottlenecks within wild rice species populations can restrict the introduction of new genetic variants,

leading to reduced heterozygosity (Veltman et al., 2019). Additionally, wild rice populations may have experienced past genetic bottlenecks contributing to the low heterozygosity levels observed. These results demonstrate the urgent need for genebanks, such as AfricaRice, to collect and preserve more wild species worldwide to avoid/reduce loss of diversity over time. About 4.4% (394 out of 9013 genotypes) of the genotyped *Oryza* sp accessions had residual heterozygosity (Table S3; Figure S1). According to previous studies, residual heterozygosity is common in many rice cultivars and can occur because of pollen flow during germplasm regeneration (Belefant-Miller et al., 2012; Gouda et al., 2020; McCouch et al., 1998). The availability of genetically pure seeds is extremely important in breeding and commercial seed production (Ndjiondjop et al., 2018a). In genebanks, genetic purity is essential in maintaining and preserving genetic resources. Genetic purity in genebanks usually refers to how an accession is maintained through time, sustaining it in its introduced state and avoiding seed mixture and contamination. In cases where accessions are landraces, it could mean maintaining the accession as a mixture of different genotypes if it is mixed. Heterozygosity and seed homogeneity are both essential for sustaining genetic purity in genebanks. Heterozygosity within and across accessions provides breeders and researchers with valuable alleles with the potential to address a current or future stress or infection. The data generated in the current study will help AfricaRice identify and address segregating accessions so that genetic purity of such accessions can be maintained. This process is currently in progress at AfricaRice with the goal to have pure seeds for distribution in the year 2024.

3.3 | Genetic divergence between pairs of accessions

The IBS values between pairs of genotypes were used to estimate average genetic similarity between the 9013 *Oryza* sp accessions. Genetic distance is classified as low, medium, high, or very high when the pairwise genetic distance values are <0.1, 0.101–0.200, 0.0201–0.3, and >0.3, respectively. (Gouda et al., 2020; Liu et al., 2021; Ndjiondjop et al., 2017, 2022). The minimum and maximum values of the paired genetic distance observed were between 0.002–0.035 and 0.394–0.529, respectively, with an average of 0.180–0.269. The 9013 *Oryza* accessions showed high genetic variability, with 58% of pairs of accessions exhibiting very high genetic distances. Pairs of genetic distance computed from 7220 SNPs across the 85 *O. longistaminata* accessions varied from 0.002 to 0.448 with an average distance of 0.200 (Table 3; Figure 4a; Table S4). The majority (79.5%) of the pairs of *O. longistaminata* accessions had genetic distance values between 0.101 and 0.200 as shown in Figure 4a. About 8.82% of pairs of accessions had genetic distance greater than 0.300. The highest genetic distance was observed between WABT230362 from Mali and the following accessions from Kenya: WABTMP18013, WABTMP18006, and WABTMP18016 (Tables S1 and S4). In wild rice, high genetic distance can be beneficial for breeding, as it can lead to increased genetic diversity and the introgression of desirable traits into cultivated rice varieties (Z. Song et al., 2005). Thus, these four wild accessions can be screened under different environmental conditions for disease resistance, drought, salinity tolerance, and iron toxicity and determine their fit as donor parents in breeding programs.

The *O. barthii* species pairs of genetic distance ranged between 0.035 and 0.394 with an average of 0.269 for the 8885 SNPs across the 163 accessions (Figure 4b; Table 3; Table S4). The values are similar to those previously reported by our team (0.01–0.34) for 115 *O. barthii* accessions that were genotyped with 26,073 SNPs markers (Ndjiondjop et al., 2019). The 3130 accessions of the African cultivated rice, *O. glaberrima*, had pairs of genetic distance between 0.020 and 0.417 with an average of 0.220, using 7585 polymorphic SNPs markers (Figure 4c; Table 3; Table S4). Unsurprisingly, *O. barthii* exhibited higher diversity than *O. glaberrima*, with values of 0.269 versus 0.220 (Table 3). These observations confirmed results reported in previous studies (Gouda et al., 2020; Ndjiondjop et al., 2017, 2019). Approximately 26.2% of pairs of *O. glaberrima* accessions had genetic distance between 0.051 and 0.100 (Figure 4c; Table S4). The low genetic distance observed between *O. glaberrima* and *O. sativa* could enhance their crossability by improving genetic compatibility, facilitating effective recombination, enabling gene flow, and reducing barriers to hybridizations. This can significantly benefit breeding programs aimed at incorporating beneficial traits from one species into the other, ultimately leading to the development of new rice varieties with desirable traits (Choi et al., 2019; Meyer et al., 2016; Ndjiondjop et al., 1999, 2010; Sarla & Swamy, 2005; Sweeney & McCouch, 2007).

In *O. sativa*, the genetic distance for pairs of accessions belonging to *indica* and *japonica* subspecies was between 0.009–0.460 and 0.012–0.529, respectively. The *indica* subspecies had a higher average genetic distance, 0.231, compared to subspecies *japonica*, 0.180. In the *indica* subspecies, most of the pairs of accessions had genetic distance values that fell between 0.401–0500 (43.33%) and 0.201–0.300 (41.20%) (Figure 4d). However, in the *japonica* subspecies, most of the pairs of accessions had genetic distance values between 0.101 and 0.200 (61.40%) (Figure 4e). In the *O. sativa* accessions, 53.1% of *indica* and 75.6% of *japonica* pairs were moderately distant, while 46.7% of *indica* accessions and 12.3% of *japonica* accessions pairs were highly distant (Figure 4d,e). Overall, the 9013 *Oryza* sp showed high genetic variability

TABLE 3 Number of accessions per species/subspecies, number of polymorphic single nucleotide polymorphisms (SNPs) markers, and minimum, maximum, and mean of the pair of identity by state (IBS) based genetic distance across the 9013 accessions.

Species/subspecies	No. of accessions	No. of polymorphic SNPs markers	Minimum	Maximum	Mean
Oryza sativa spp. *indica*	4928	18,902	0.009	0.460	0.231
Oryza sativa spp. *japonica*	707	15,865	0.012	0.529	0.180
Oryza glaberrima	3,130	7585	0.020	0.417	0.220
Oryza barthii	163	8885	0.035	0.394	0.269
Oryza longistaminata	85	7220	0.002	0.448	0.200
Oryza sp.	9013	27,718	0.000	0.469	0.267

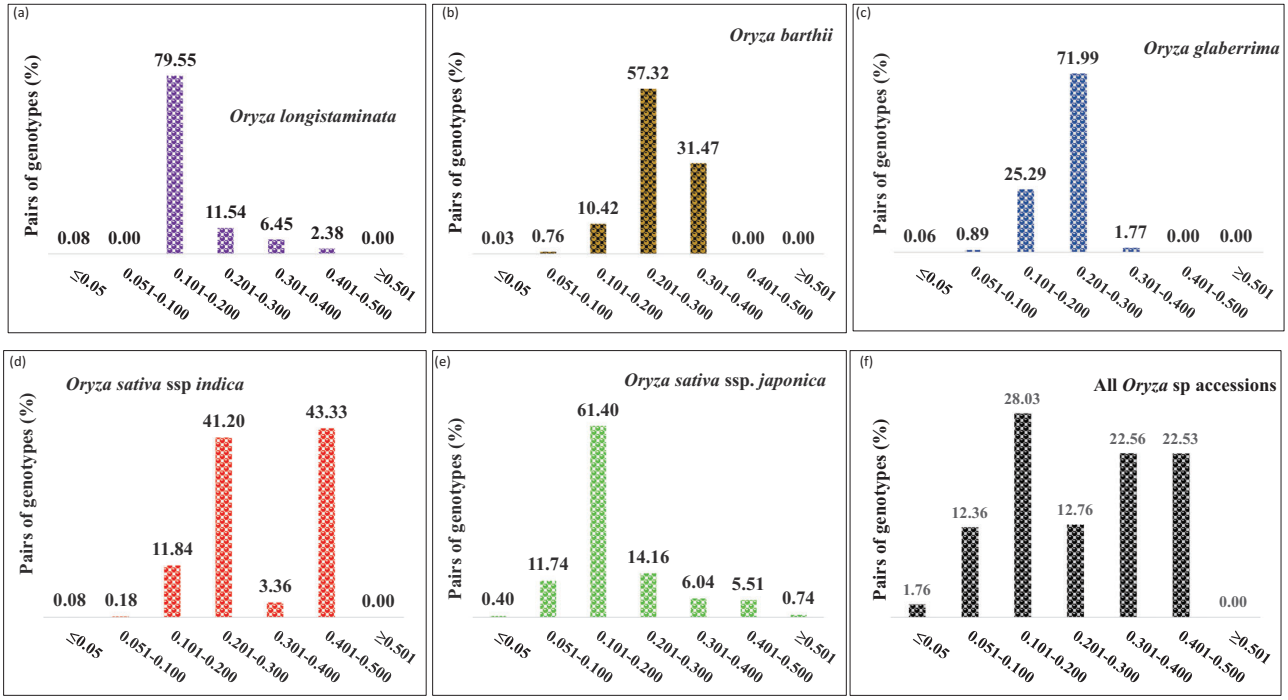

FIGURE 4 Frequency distribution of pairwise identity by state (IBS)-based genetic distance computed in (a) 85 *Oryza longistaminata*, (b) 163 *Oryza barthii*, (c) 3130 *Oryza glaberrima*, (d) 4928 *Oryza sativa* ssp *indica*, (e) 707 *Oryza sativa* ssp *japonica*, and (f) all 9013 accessions based on 7220; 8885; 7585; 18,902; 15,865; and 27,718 polymorphic single nucleotide polymorphisms (SNPs), respectively.

with 40.1% and 45.1% of pairs of accessions with moderate and high genetic distance, respectively (Figure 4f). The results demonstrate that the AfricaRice accessions are genetically diverse both within and across species, making them a valuable source for discovering novel alleles that can be incorporated into new rice varieties through breeding programs.

3.4 | Molecular diversity indices

Molecular diversity of accessions was estimated using several diversity estimators including the number of segregating sites (S), proportion of polymorphic sites (Ps), theta (Θ), nucleotide diversity (π), and gene diversity (D) (Table S5). The observed values of m, Ps, Θ, and π based on species indicated that the *indica* accessions, 54.67% of the overall accessions, were more diverse (Ps = 0.911 and π = 0.147) than the *japonica* accessions that represented 7.84% of the population (Ps = 0.774 and π = 0.094) (Table S5). *Oryza longistaminata* accessions were the least diverse compared to other species with Ps = 0.251 and π = 0.045. The *O. longistaminata* accessions had the smallest number of both wild and cultivated genotyped accessions, thus could explain the low values observed in the present study. These observations confirmed those previously observed by Ndjiondjop et al. (2022), who reported that genetic diversity estimators can be affected by sample size.

When the accessions were grouped by continent of origin, accessions from Africa were observed to be more diverse (Ps = 1.00 and π = 0.256) than accessions from Asia (Ps = 0.982 and π = 0.161); the Caribbean; Central, South, and North America (Ps = 0.962 and π = 0.160); and Oceania and the Middle East (Ps = 0.505 and π = 0.175). According to Ps and π, accessions that originated from Nigeria were the most diverse (Ps = 0.999 and π = 0.254), while those from Mozambique were the least diverse (Ps = 0.503 and π = 0.144) (Table S5). Nigeria also had the highest number of genotyped accessions and represented about 24.04% of the overall accessions in this study. These observations reaffirm that sample size influence genetic diversity indices but also confirms that it influences measures of geographic diversity. Thus, the high genetic diversity observed in Nigerian accessions is likely attributed to the large number of genotyped accessions. Tajima's *D* test is used to distinguish a randomly evolved DNA sequence ("neutrally") from a non-randomly evolved DNA sequence. This includes directional selection or balancing selection, demographic expansion or contraction, genetic hitchhiking, or introgression (Tajima, 1989). It measures the shift in the site frequency when an allele moves to fixation by comparing the observed diversity to the expected diversity if selection is absent (Cadzow et al., 2014; Lewontin & Krakauer, 1973; Nielsen, 2001; Schmidt & Pool, 2002). That is, it compares the number of segregating sites to the number of nucleotide differences between individuals. As a haplotype moves to fixation, it causes an initial reduction in variation around the allele and results in negative values of Tajima's *D* statistic, giving rice to new mutations to create an excess of rare variants as it reaches fixation (Braverman et al., 1995; Eckshtain-Levi et al., 2018; Schmidt & Pool, 2002). In the current study, positive Tajima's D values were observed for all agro-ecologies group except for floating ecology. Differences were observed in Tajima's *D* values between species, with negative values being observed for accessions belonging to *O. longistaminata* and *O. sativa indica* species/subspecies (Table S5). Negative Tajima's D values were also observed for accessions from Ghana, Togo, Sri Lanka, United States, China, Thailand, Bangladesh, Indonesia, Zambia, Malawi, Benin, and India. These negative *D* values suggest the presence of excess rare alleles between *O. longistaminata* and *O. sativa indica* species/subspecies, and countries of origin, which is indicative of a recent selective sweep event, purifying selection, population expansion after a less recent bottleneck, or linkage to a swept gene (Schmidt & Pool, 2002; L. Zhang, Ma, et al., 2021). These observations suggest that the AfricaRice germplasm is genetically diverse; however, there remains a need to increase genetic diversity in the collection. Acquiring additional rice accessions from underrepresented countries and species will help to expand the genetic diversity of the AfricaRice genebank.

3.5 | Genetic relationships and population structure

Assessment of genetic relationships and divergence of genetic resources is important in supporting management of collections and identifying potential parents for breeding programs (Pathaichindachote et al., 2019). This will minimize the use of closely related accessions as parents in breeding programs, reducing inbreeding depression and increasing genetic variation (Suvi et al., 2019). In the present study, polymorphic SNPs were used to assess genetic relatedness using three clustering approaches, PCA projection, distance-based neighbor-joining cluster, and population structure analysis based on Bayesian cluster analysis.

The first, second, and third PCs of PCA projection explained 48.9%, 9.0%, and 3.0% of the overall variation, respectively. Thus, the first three PCs explained 60.0% of the molecular variation across the 9013 accessions (Table S6). The three PCs of the accessions were graphed and displayed a clear structure of the population (Figure 5). Indeed, PC1 and PC2 alone can already provide valuable insights into the genetic diversity among the 9013 accessions genotyped using the DArT markers. PC1 captured most of the variance, thus revealing the most significant differences among the accessions, particularly genetic diversity between species and subspecies. Despite capturing a smaller proportion of the variance, PC2 offered additional insights into other aspects of variation, such as phenotypic differences or geographical variations. Moreover, each additional PC captured an additional portion of the residual variance in the data, allowing for a more comprehensive understanding of the variation structure (Greenacre et al., 2022). Based on these considerations, the 9013 accessions were projected onto three axes, as shown in Figure 5. NJ is an algorithm used to analyze the evolution of species according to genetic sequences and viewed as phylogenetic trees. The IBS-based genetic distance matrix was used to construct an NJ tree to understand the relationships between the 9013 accessions and how these relationships reflect their genetic status, agro-ecology, and geographical origin. The NJ tree analysis clustered accessions into six groups and was similar to the results of the PCA clustering (Figures 5 and 6). Group membership ranged from 60 accessions in NJ-Group 2 to 3233 accessions in NJ-Group 1 (Table S1).

The first group, NJ-Group 1, contained all the *O. glaberrima* and 63.2% of the *O. barthii* accessions (Figure 6; Figure S2). Approximately 96.6% of accessions in this group are traditional landraces and 3.1% are wild accessions (Figure S3). Traditional African rice varieties such as CG14, CG 20, and IG 10 clustered in this group. Irrigated and rainfed lowland ecologies accounted for 39.8% and 36.4% of the total accessions of NJ-Group 1, while 18.9% of the accessions were of upland ecology (Figure S4; Table S1). For NJ-Group 1, 93.8%

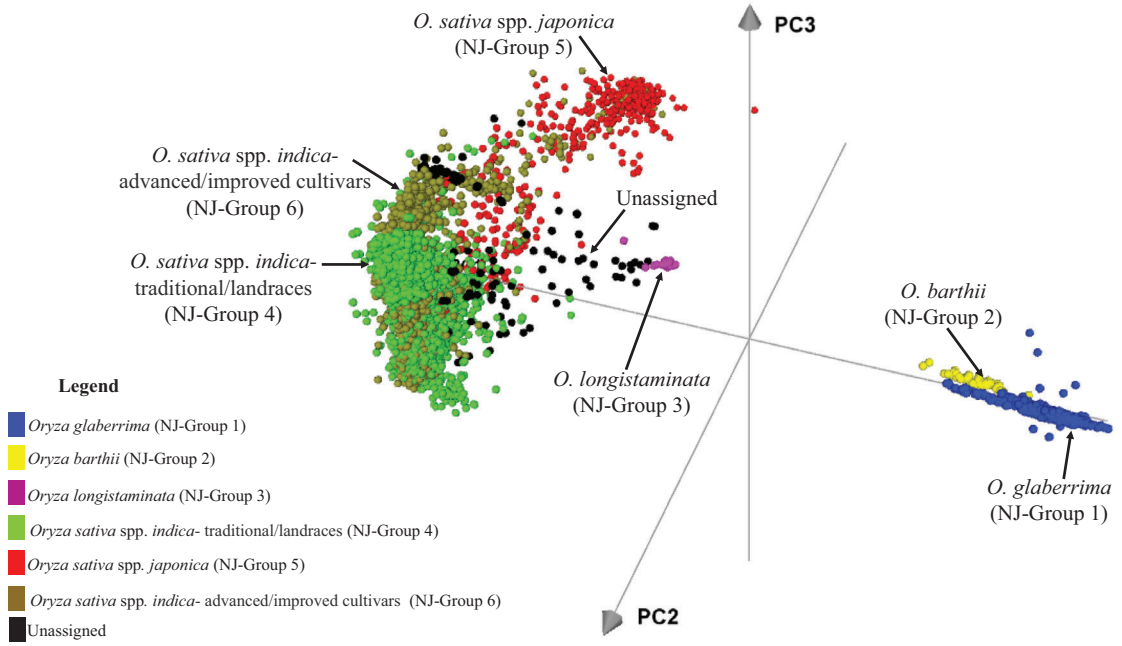

FIGURE 5 Principal component analysis of 9013 *Oryza* accessions using 27,718 polymorphic single nucleotide polymorphisms (SNPs). Accessions with the same color belong to the same group. Accessions are colored according to relationships between them. The principal component analysis (PCA) was colored using iTOL v4 online program (Letunic & Bork, 2019).

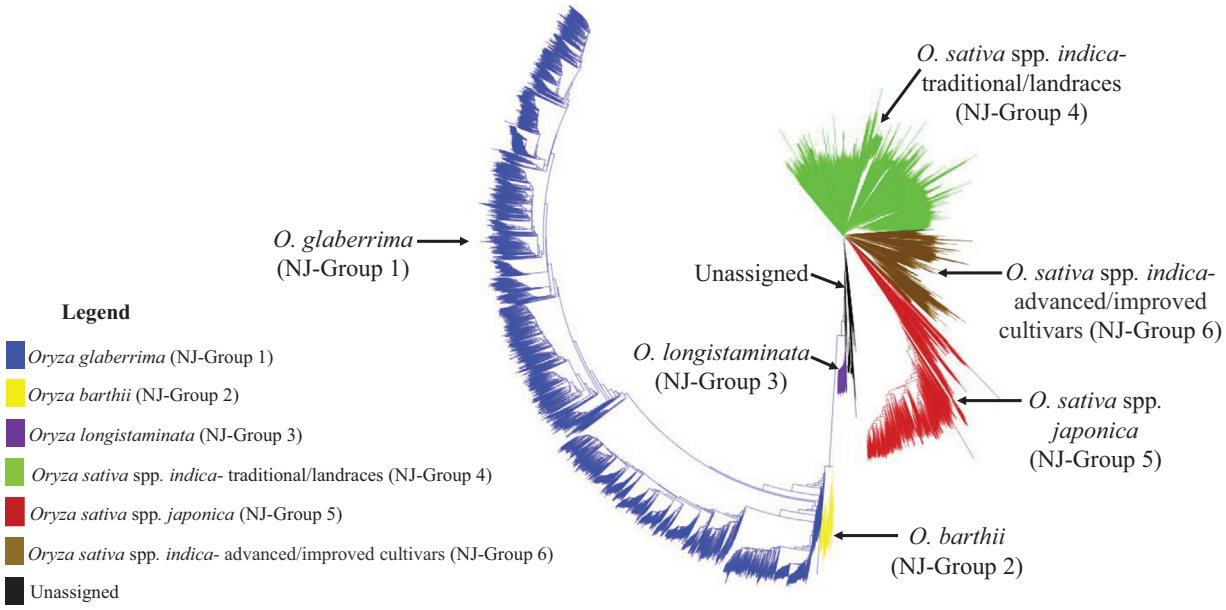

FIGURE 6 Neighbor-joining tree of 9013 *Oryza* accessions created with 27,718 polymorphic SNPs. Accessions with the same color belong to the same group. Accessions are colored according to relationships between them. The neighbor joining (NJ) tree was colored using iTOL v4 online program (Letunic & Bork, 2019). Details on each sample are in Table S1.

of the accessions originated from West Africa and 4.1% from Central Africa, which is not surprising since *O. barthii* and *O. glaberrima* accessions originated from West African countries and made up much of the group (Figure S6). Accessions in this group originated from 37 countries, with the majority originating from Nigeria (28.3%) followed by Liberia (24.1%) and then Mali (1.16%) (Figure S6). Of the 3130 *O. glaberrima* accessions in NJ-Group, 1535 were early maturing (90–105 days) and 1555 had intermediate maturity (105–150 days) (Table S1). These early and intermediate maturity accessions

can benefit farmers who may want to plant for a second season, thereby enhancing household food security and income generation.

The second group, NJ-Group 2, consisted of 60 African wild rice accessions, which are all *O. barthii* species (Figure S2). Out of these, 51 accessions are of upland ecology, while seven belong to the irrigated lowland ecology (Figure S4; Table S1). Most of the accessions in this group originated from Chad (35%), Mali (25%), and Cameroon (25%) (Figure S6). *Oryza barthii* and *O. glaberrima* are two closely related species that are native to Africa and are often clustered together in phylogenetic analyses (Gouda et al., 2021; Ndjiondjop et al., 2017; Wambugu et al., 2019). This is principally because available evidence suggests that *O. glaberrima* was domesticated from wild populations of *O. barthii*, hence these two species have closely related seed morphological traits such as shape, size, and color. However, we hypothesize that the 103 *O. barthii* accessions that clustered with the *O. glaberrima* accessions in NJ-Group 1 are intermediate or "weedy" accessions because of hybridization between *O. barthii* and *O. glaberrima*. These results agree with previous studies by Gouda et al. (2021) and Orjuela et al. (2014), who found that nine O. *barthii* accessions (WAB0009247, WAB0038205, WAB0038232, WAB0028947, WAB0038229, WAB0038219, WAB0038231, WAB0038217, and WAB0038228) that clustered in NJ-Group 1 may be intermediate O. *barthii*, also known as O. *barthii* weeds. *Oryza barthii* and *O. glaberrima* are inter-fertile and hybridize readily (Sow et al., 2014a), hence the possibility of taxonomic misclassification cannot be ruled out (Orjuela et al., 2014; Wambugu et al., 2019). In such cases, species specific diagnostic markers could offer a vital tool for resolving taxonomic identity (Ndjiondjop et al., 2018b). The re-sequencing of these accessions could also elucidate the connection between them.

NJ-Group 3 was exclusively composed of the 85 *O. longistaminata* accessions genotyped in this study (Figure S2).

NJ-Group 4 contained 3230 accessions, 99.4% of which are *O. sativa indica* subspecies (Figure S2). Traditional/landraces and improved cultivars had the highest representation in this group, with 58.9% and 41.1%, respectively (Figure S3). GIZA 178, a known *indica* from Africa (Ni et al., 2002), was also a member of NJ-Group 4. Eight accessions in NJ-Group 4 were previously reported to cluster with *indica* accessions in a study by Agrama et al. (2010). Many of the accessions are from rainfed lowland (52.0%), irrigated lowland (43.5%) and mangrove swamp (3.0%) agro-ecologies (Table S1; Figure S4). In this group, 1868 accessions originated from West Africa, 454 from South-eastern Asia, and 391 accessions from Central America (Figure S6). The accessions in this group originate from 64 countries, with the majority from Nigeria (18.6%), followed by the Philippines (11.9%) and Colombia (10.8%) (Table S1). Eighty-five of the 1403 accessions in this group were of early maturity (90–105 days) and 661 were of intermediate maturity (105–150 days) (Table S1).

NJ-Group 5 consisted of 686 accessions of the *japonica* subspecies and was 97.1% of the *japonica* accessions genotyped. The group also consisted of 135 *indica* accessions that were mostly landraces/traditional cultivars (63.2%) (Table S1; Figures S2 and S3). Three tropical *japonica* accessions, Moroberekan, IRAT 104, and WAB56-104, and one temperate *japonica*, GIZA 172, were previously reported to cluster with *japonica* accessions by Agrama et al. (2010), thus confirming this group as subspecies *japonica*. Similarly, several traditional upland *japonica* accessions, Danane, Kleminsin, OS 6 (Faro 11), and Iguape cateto were also members of this group. Most of the accessions in this group, ~83.3%, were from upland ecology, while the remainder were of lowland and mangrove swamp agro-ecologies (Figure S4). These accessions originate from 55 countries, with the largest percentage from Nigeria (22.8%), Cote d'Ivoire (15.0%), and Liberia (11.8%) (Table S1; Figure S6). Only 9.5% and 7.3% of the accessions in this group were from Asia and the Americas (Figure S5).

NJ-Group 6 contained 1450 accessions of *O. sativa indica* subspecies. This group was made up of traditional/landraces cultivars (47.8%), while the remaining 52.1% were advanced/improved rice cultivars. The traditional lowland *indica* accessions Cisadane, BW311-9, Seberang, and IR36 were members of this group. Most accessions in this group were of rainfed, irrigated, and mangrove ecologies, 47.6%, 30.5%, and 22%, respectively (Figure S3). Additionally, 308 of 457 (67%) mangrove ecologies accessions were part of this group. (Figure S4). The accessions in this group originate from 46 countries, with the majority from the African (84.4%) and Asian (13.0%) continents (Figure S5). The accessions in this group share some genetic relation due to their appearance in the group and the landraces may possess valuable alleles to develop improved varieties. However, further research is needed to help elucidate the genetic connections between these landraces and advanced cultivars (Table S1; Figures S2 and S3).

There were 135 accessions that were not assigned to a group, the "unassigned" group, which accounted for 1.4% of the total genotyped accessions. The accessions in this group may be admixed based on their position in the NJ tree (Figures 2, 5 and 6) (Garris et al., 2005; Huggins et al., 2019; Wang et al., 2018). Thus, we hypothesize that these accessions were derived from crosses between *Indica* and *Japonica* cultivars or hybridization between *O. barthii* and *O. glaberrima*.

Population structure analysis is essential for understanding the genetic diversity, distribution, and evolutionary dynamics of populations, which have significant implications for conservation, management, and breeding strategies (Kimwemwe et al., 2023; Schierenbeck, 2017). In the present study,

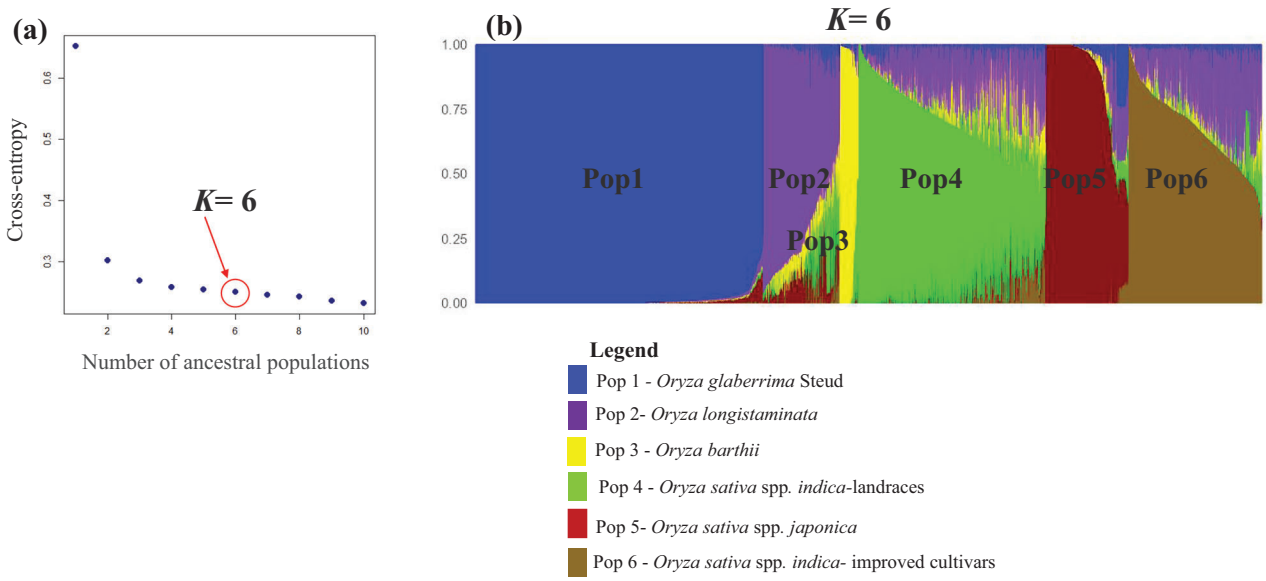

FIGURE 7 Population structure of 90,133 accessions constructed with 27,718 polymorphic single nucleotide polymorphisms (SNPs) markers. (a) A plot of cross-entropy versus the number of ancestral populations, *K*, where $K = 1–10$, to determine the number of populations using the R package LEA (Landscape and Ecological Association); (b) graphical display of the population structure of 9013 *Oryza* accessions at $K = 6$. Probability of membership >60% was assigned to the same group, while those with <60% probability in any group were assigned to a "mixed" group. (Details on each sample belonging in each group are in Table S1.

population structure analyses were also performed among the 9013 accessions using 27,718 SNPs. Before performing population structure analyses, a curve of cross-entropy versus number of ancestral populations was performed. The entropy criterion is a cross-validation technique that assesses the quality of fit of the model. The membership of each accession was run from $K = 1$ to $K = 10$ for the 9013 accessions. The optimal number of ancestral populations was determined to be six according to *K* ($K = 6$) (Figure 7a), hereafter referred to as Pop1 to Pop6 (Figure 7b). The genetic population structure at $K = 6$ was illustrated using a multidimensional scaling (MDS) plot in Figure S7a. An NJ phylogenetic tree was also constructed to represent the genetic distances among these genetic groups (Figure S7b). Of the 9013 *Oryza* sp accessions studied, 8592 (95.3%) of the total accessions were assigned into one of the six genetic populations, referred here as population (Pop). The remaining 421 accessions (4.7%) were assigned to a mixed group. The number of accessions in the six populations ranged from 52 to 3307 accessions. Pop1 was composed of 3130 accessions that were exclusively classified as *O. glaberrima*. Pop2 consisted of 85 accessions, all of which were classified as *O. longistaminata*. The 85 accessions in this population were the exact same accessions in NJ-Group 3 (Figure 6; Table S1; Figure S7). Pop3 consisted of 60 accessions, 52 of which were considered as *O. barthii* and the remaining eight as admixed. About 32.5% of the genotyped *O. barthii* clustered in this population. According to population structure analysis where $K = 6$, the remaining 111 genotyped *O. barthii* accessions were admixed. However, NJ tree analysis grouped 103 of the *O. barthii* accessions in NJ-Group 1 as *O. glaberrima* and the other eight in NJ-Group 2 as *O. barthii* (Figures 6 and 7; Table S1). The traditional/landraces *indica* accessions clustered in Pop4 and consisted of 3307 accessions (Table S1). Approximately 93.9% of accessions from Pop4 (3108 of 3307 accessions) also grouped in NJ-Group 4. Of the remaining accessions in Pop4, 94 clustered as *japonica* (NJ-Group 5) and 134 were unassigned according to NJ tree analysis (Table S1). A total of 663 of the 707 genotyped *japonica* accessions clustered to form Pop5, 661 of which were also members of NJ-Group 5. The last population, Pop6, contained 1355 accessions that were also members of NJ-Group 6, the improved-indica group (Table S1; Figure S7; Figure 7).

In summary, about 507 and 1856 *japonica* and *indica* accessions were classified as traditional/landraces, based on NJ tree and Bayesian cluster analysis, respectively (Table S1). These landraces/traditional rice accessions are valuable germplasm to Africa since they are genetically diverse and locally adapted and were developed by farmers over many years. These traditional varieties could possess unique alleles for resistance to pests and diseases, tolerance to environmental stresses like drought, and the ability to grow in low-input agricultural systems. A total of 421 of the 9013 accessions, 4.7%, were classified as admixtures indicating shared genetics between some of the *Oryza* species in this study. Approximately, 4.8%, 26.4%, and 68.8% of the admixed accessions were *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, and *O. barthii* subspecies/species, respectively (Table S1; Figure

S7). Similar cases of admixture have been reported in other genebanks such as the International Rice Research Institute (IRRI) genebank (Wang et al., 2018).

3.6 | Features of AfricaRice genebank accessions

In previous studies, population structure analysis suggested the existence of several subpopulation in rice (Agrama et al., 2009; Garris et al., 2005; Yan et al., 2010). In this study, PCA, phylogenic trees, and population structure analysis separated these accessions into six possible genetic groups. Although it is well documented that geographical distribution is an important factor that influences genetic diversity of species (Gutaker et al., 2020; Phillips et al., 2018; F. Zhang, Wang, et al., 2021), the PCA and NJ tree analysis of geographical regions or continents in this study did not show any clear clustering patterns (Figures S5 and S6). This is in contrast to findings of a study by Wang et al. (2018) where genomic analysis of 3010 rice accessions, majority of which were obtained from the IRRI, gave a clustering pattern that was largely associated with geographical origin. This observation may be due to the movement or introduction of rice into new regions for cultivation, thus leading to a mixture of genetic diversity across continents.

To understand patterns of genetic relationship among the 9013 accessions in NJ tree and PCA, multiple plots were constructed using five categorical variables (species/subspecies, group membership from continents, ecologies, regions of origin, and biological status). PCA and the NJ tree identified six clusters that largely correspond to the various species/subspecies, *O. barthii*, *O. glaberrima*, *O. sativa indica*, *O. sativa japonica*, and *O. longistaminata*, although there were notable overlaps (Figure S2). The observed clusters were consistent with the Bayesian cluster analysis for $K = 6$ (Figure S1). The *O. glaberrima* accessions clustered close to the *O. barthii* group, confirming observations of previous studies. Noteworthy was the high number of intermediate *O. barthii* accessions observed, 103 and 111 accessions of the 163 genotyped depending on the clustering method (Figures 6 and 7). This could be attributed to the sympatric distribution of these two species, thus providing opportunities for gene flow between them. In Mali, for example, the wild relative *O. barthii* tends to occur in fields where cultivated *O. glaberrima* rice is grown. Additionally, as previously stated, the ancestral progenitor of *O. glaberrima* is *O. barthii*, so there is an expectation of high relatedness between them. These observations and hypothesis agree with observations reported by earlier studies (Gouda et al., 2020; Ndjiondjop et al., 2017; Orjuela et al., 2014; Sow et al., 2014b).

The biological status, whether an accession was improved, traditional/landrace or wild, was assessed using multidimensional scaling and NJ tree. Three clusters with varying degrees of admixture were identified (Figure S3). This method was also applied to ecologies and place of origin/collection, but no distinct pattern was observed (Figures S4–S6). These analyses support the hypothesis that rice accessions moved freely across regions and habitats leading to genotype similarity.

3.7 | Genetic differentiation in AfricaRice genebank accessions

To understand the extent of genetic differentiation between the four rice species, genetic distances among pairs of accessions were analyzed (Table S7). In practice, Nei's G_{ST} test is analogous to F_{ST} test where 0 indicates identical populations. An F_{ST} of 0.01–0.05 indicates low differentiation, 0.05–0.15 indicates moderate differentiation, F_{ST} of >0.15 indicates a high level of differentiation (Govindaraju, 1989), and an F_{ST} value of 1 indicates complete isolation, suggesting that the populations have different fixed alleles (Pemberton, 2004).

Estimates of the coefficient of gene differentiation (Nei's G_{ST} test) determined that the highest divergence (0.656) occurred between the Asian rice, *O. sativa japonica*, and African rice, *O. glaberrima* (Table S7a). High divergence (0.651) was also observed between *O. glaberrima* and the wild species *O. longistaminata*. The lowest divergence (0.068) identified occurred between *O. glaberrima* and its wild progenitor *O. barthii*, indicating low genetic differentiation between the two species. *Japonica* accessions showed more divergence with other species compared to *indica* accessions (Table S7a). Interestingly, wild species and advanced/breeding lines also showed high divergence (0.323) (Table S7b). This could be a result of different evolutionary processes experienced by wild rice and advanced/breeding lines. Wild rice populations have been subjected to natural selection and have adapted to their specific environments over many generations, leading to high genetic diversity, whereas advanced/breeding lines have undergone artificial selection for specific traits, resulting in a narrower gene pool.

The genetic differentiation among the predicted six group was analyzed according to NJ tree using Nei's G_{ST} test (Table S7c). The pairwise estimates of the Nei's G_{ST} values between pairs of clusters ranged from 0.018 to 0.649, indicating that they expressed low to high genetic differentiation. The largest genetic differentiation (0.634) was detected between NJ-Group 1 (*O. glaberrima* group) and NJ-Group 3 (*O. longistaminata* group), while the lowest genetic differentiation (0.018) was detected between the two *indica* groups, NJ-Group 4 and NJ-Group 6. High genetic differentiation (0.332) was observed between indica and japonica groups, which is slightly similar to that observed in these same subspecies by Hour et al. (2020) (0.308) in 136 *O. sativa* germplasm collected from the National Plant Genetic

TABLE 4 Analysis of molecular variance (AMOVA) for single nucleotide polymorphism (SNP) variation among and within groups (populations) based on 9013 genotyped Oryza accessions with 27,718 polymorphic SNPs for the categories continents, regions, country of origin, ecologies germplasm types, species, and Bayesian population structure at $k = 6$.

Category	Source of variation	*df*	Sum of squares	Variance components	Percentage of variation	*PhiPT*
Continents[a]	Among populations	4	3,998,531	1519.980	18.25	–
	Within populations	9008	61,340,773	6809.588	81.75	–
	Total	9012	65,339,303	8329.568	100.00	0.182
Regions[a]	Among populations	10	5,700,410	1282.809	16.23	–
	Within populations	9002	59,638,893	6625.071	83.77	–
	Total	9012	65,339,303	7907.881	100.00	0.162
Country of origin[a]	Among populations	74	14,689,289	1747.200	23.56	–
	Within populations	8938	50,650,014	5666.817	76.44	–
	Total	9012	65,339,303	7414.018	100.00	0.235
Ecology[a]	Among populations	7	5,090,906	827.2517	11.01	–
	Within populations	9005	60,248,397	6690.5494	88.99	–
	Total	9012	65,339,303	7517.8011	100.00	0.110
Germplasm types[a]	Among populations	2	7,881,424	2016.687	24.03	–
	Within populations	9010	57,457,880	6377.123	75.97	–
	Total	9012	65,339,303	8393.810	100.00	0.240
Species[a]	Among populations	4	36,730,757	7098.903	69.10	–
	Within populations	9008	28,608,546	3175.904	30.90	–
	Total	9012	65,339,303	10,274.808	100.00	0.690
Groups based on cluster analysis	Among populations	6	37,483,773	5869.111	65.48	–
	Within populations	9006	27,855,531	3092.997	34.52	–
	Total		65,339,303	8962.108	100.00	0.654
Bayesian population structure at $k = 6$	Among populations	6	36,294,115	5633.258	63.59	–
	Within populations	9006	29,045,189	3225.093	36.41	–
	Total	9012	65,339,303	8858.351	100.00	0.635

Abbreviation: *df*, degree of freedom.

[a]Details regarding the composition of each of the eight groups used to run AMOVA analysis were compiled in Table S1.

Resource Center, Taiwan. When accession ecologies were considered, the largest genetic differentiation (0.565) was detected between accessions adapted to shallow/forest and mangrove swamp ecologies, while the lowest genetic differentiation (0.005) was between hydromorphic and irrigated lowland ecologies (Table S7d). Accessions that originated from West African countries showed moderate genetic differentiation from those from the other regions stated in the present study (Table S7e). Surprisingly, accessions with origin in the Caribbean showed low genetic differentiation with accessions from Oceania (0.004). The genetic differentiation (Nei's G_{ST} analogue of F_{ST}) between the six predicted populations based on population structure at $K = 6$ was calculated. The resulting pairwise estimates of Nei's G_{ST} values ranged from 0.016 to 0.688 (Table S7f). The largest genetic differentiation (0.688) was observed between Pop1 and Pop5, while the lowest genetic differentiation was observed between Pop4 and Pop6 (0.016). Interestingly, high genetic differentiation was observed between Pop1 and the other five populations as well as the admixed group. Our results highlight that the species/subspecies and accessions biological status are major factors influencing their genetic structure, with Pop1 and Pop5 showing a higher genetic differentiation (0.688). This finding contrasts with Kimwemwe et al. (2023), Salem and Sallam (2016), and Xu et al. (2016) studies, which reported that the source of the rice genotypes was the major factor influencing their genetic makeup.

The distribution of genetic diversity between and within groups (populations) was analyzed using an AMOVA on the 9013 rice accessions genotyped with 27,718 polymorphic SNPs (Table 4). The results showed that more genetic variation exists within populations than across populations when looking at continents, regions, country of origin, ecology, and germplasm type (75.9%–88.9%). However, genetic differentiation was greater across populations for species, based on cluster analysis and Bayesian clustering analysis (63.5%–69.1%). The maximum genetic diversity (69.1%) across populations was observed for species, while the

lowest diversity (11.0%) was observed for ecology. Additionally, the highest level of diversity within populations was observed for ecology (88.9%) and the lowest diversity was observed in species (30.9%). When the rice accessions were grouped based on their geographic origin, a large proportion of the genetic variation was within regions (83.7%) and a smaller proportion across regions (16.3%).

The genetic variation for cluster analysis and Bayesian analysis groups was similar for across and within accession groups. Within rice accessions, variation for continents and germplasm type accounted for 81.7% and 75.9% of the total genetic variation, respectively. The most important differentiation was observed for species followed by groups based on cluster analysis and on Bayesian population structure at $K = 6$, with AMOVA values of 0.690, 0.654, and 0.635, respectively (Table 4). These results suggest that 27,718 molecular markers improved the classification of AfricaRice accessions. The results also confirm observations reported in previous studies (Gouda et al., 2021; Ndjiondjop et al., 2017, 2018b). By effectively capturing genetic variation and revealing population structure, our results contribute to a deeper understanding of the relationships between the four *Oryza* species genotyped in the present study compared to what was previously known.

3.8 | Ascertainment bias in SNP genotyping for germplasm management

The procedure used to select SNPs for SNP genotyping may create ascertainment bias in the data (Clark et al., 2005; Nielsen, 2004; Semon et al., 2005). Ascertainment bias is described as a systematic deviation from a theoretical population genetics statistic that can arise through non-random sampling, biased SNP discovery, or unequal distribution of genetic diversity across populations (Clark et al., 2005; Lachance & Tishkoff, 2013; Nielsen, 2000, 2004). The small sample size used to discover SNPs tends to favor common alleles or high frequency SNPs, which inherently causes ascertainment bias unless the genome of every individual is sequenced (Gravel et al., 2011; Nielsen, 2004). An excess of rare alleles in a population is an indicator of population growth. However, the preference for high-frequency SNPs in genotyping may reduce the detection of rare alleles in a population, thus reducing evidence of population growth (Nielsen, 2000). The use of these SNP arrays can lead to overestimation of population genetic measures such as nucleotide diversity, linkage disequilibrium, fixation, and recombination, possibly leading to compromised inferences and erroneous conclusions (Lachance & Tishkoff, 2013; Moragues et al., 2010; Nielsen, 2004). Although SNP arrays such as DArT have implicit ascertainment bias, they can still be useful in germplasm management, characterization, and curation. As most of the accessions in the *Oryza* species populations of this study originate in Africa, any private alleles detected in each population may not be present outside of Africa due to the small sample size of accessions from other countries. While the private alleles associated with individual *Oryza* species populations may not be present in accessions outside of Africa, they can be used to fingerprint accessions in the AfricaRice collection to determine their degree of relatedness. The alleles can also be used to genotype African accessions in genebanks and breeding programs outside of Africa or track allele movement from African germplasm in modern cultivars. The SNPs may also be a valuable tool to identify selective sweeps in landraces of a species and genetic duplication. One approach that can be taken to reduce ascertainment bias specifically for germplasm management is the DNA pooled-sampling method (Arca et al., 2021). DNA from individuals in a species is bulked and genotyped together. This method can still result in ascertainment bias due to design of the SNP array and the subset of SNPs selected for the array. However, there are steps that can be taken to reduce ascertainment bias. A two-step model approach, using controlled pools of 1000 SNPs to calibrate a logistic regression, was employed to test polymorphism and allele frequency (Arca et al., 2021). This method was able to detect little or no ascertainment bias and reliably estimated genetic relatedness among populations with a large range of genetic divergence (Arca et al., 2021).

4 | CONCLUSION

The results of this study indicate that the 27,718 DArT SNP markers identified were able to separate the 9013 accessions into six groups, including the *indica* and *japonica* subspecies. DArTseq SNP markers have once again demonstrated their pivotal role in characterizing, assessing genetic diversity, and elucidating population structure within an extensive collection of rice accessions. These data are invaluable for initiatives aimed at crop enhancement and conservation. This study indicated that high genetic variation exists between the *O. longistaminata*, *O. barthii*, *O. glaberrima*, *O. sativa ssp. indica*, and *O. sativa ssp. Japonica* accessions. Findings of this study indicate the rich genetic landscape within these accessions, which can potentially be harnessed for crop improvement and breeding strategies. Researchers and breeders should recognize the importance of these diverse genetic resources in their efforts to improve rice varieties to enhance food security and sustainability in Africa and worldwide.

AUTHOR CONTRIBUTIONS

Arnaud Comlan Gouda: Data curation; formal analysis; methodology; writing—original draft. **Jean Rodrigue Sangare**: Formal analysis; writing—original draft. **Karlin Gnikoua**: Data curation; methodology. **Peterson Wambugu**: Writing—review and editing. **Trevis D. Huggins**:

Writing—review and editing. **Marie Noelle Ndjiondjop**: Conceptualization; methodology; validation; writing—review and editing.

ACKNOWLEDGMENTS

The present study was supported by a grant given to Marie Noelle Ndjiondjop (AfricaRice) BMZ funding of Genebank Uplift from Deutsche Gesellschaft für International.

CONFLICT OF INTEREST STATEMENT

On behalf of all authors, the corresponding author states that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

All relevant files are included within this article and its Supporting Information. The raw genotype data of all accessions will be deposited in a public database upon acceptance of the manuscript.

ORCID

Arnaud Comlan Gouda https://orcid.org/0000-0002-9152-4736

REFERENCES

Adeboye, K. A., Oyedeji, O. E., Alqudah, A. M., Börner, A., Oduwaye, O., Adebambo, O., & Daniel, I. O. (2020). Genetic structure and diversity of upland rice germplasm using diversity array technology (DArT)-based single nucleotide polymorphism (SNP) markers. *Plant Genetic Resources: Characterisation and Utilisation*, *18*(5), 343–350. https://doi.org/10.1017/S1479262120000362

Agrama, H. A., Yan, W., Jia, M., Fjellstrom, R., & Mcclung, A. M. (2010). Genetic structure associated with diversity and geographic distribution in the USDA rice world collection. *Natural Science*, *2*(4), 247–291. https://doi.org/10.4236/ns.2010.24036

Agrama, H. A., Yan, W., Lee, F., Fjellstrom, R., Chen, M., Jia, M., & Mcclung, A. (2009). Genetic assessment of a mini-core subset developed from the USDA Rice Genebank. *Crop Science*, *49*(4), 1336–1346. https://doi.org/10.2135/cropsci2008.06.0551

Arca, M., Mary-Huard, T., Gouesnard, B., Bérard, A., Bauland, C., Combes, V., Madur, D., Charcosset, A., & Nicolas, S. D. (2021). Deciphering the genetic diversity of landraces with high-throughput SNP genotyping of DNA bulks: Methodology and application to the maize 50k array. *Frontiers in Plant Science*, *11*, 568699. https://doi.org/10.3389/fpls.2020.568699

Arouna, A., Soullier, G., Mendez del Villar, P., & Demont, M. (2020). Policy options for mitigating impacts of COVID-19 on domestic rice value chains and food security in West Africa. *Global Food Security*, *26*, 100405. https://doi.org/10.1016/j.gfs.2020.100405

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Çiftçi, V., Sáenz De Miera, L. E., Aasim, M., Nadeem, M. A., Aktaş, H., Özkan, H., & Hatipoğlu, R. (2017). A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from central fertile crescent. *PLoS One*, *12*(1), e0167821. https://doi.org/10.1371/Journal.Pone.0167821

Belefant-Miller, H., Miller, G. H., & Moldenhauer, K. A. K. (2012). Utilizing the genetic diversity within rice cultivars. *Planta*, *235*(3), 641–647. https://doi.org/10.1007/s00425-011-1566-x

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, *140*(2), 783–796. https://doi.org/10.1093/genetics/140.2.783

Cadzow, M., Boocock, J., Nguyen, H. T., Wilcox, P., Merriman, T. R., & Black, M. A. (2014). A bioinformatics workflow for detecting signatures of selection in genomic data. *Frontiers in Genetics*, *5*, 293. https://doi.org/10.3389/fgene.2014.00293

Chang, L., Hu, M., Ning, J., He, W., Gao, J., Ndjiondjop, M. N., Fu, Y., Liu, F., Sun, H., Gu, P., Sun, C., & Zhu, Z. (2022). The genetic control of glabrous glume during African rice domestication. *Journal of Genetics and Genomics*, *49*(5), 427–436. https://doi.org/10.1016/j.jgg.2022.02.009

Chesnokov, Y., & Kosolapov, V. (2020). Morphological genetic markers in plants. *Russian Journal of Genetics*, *56*, 1406–1415. https://doi.org/10.1134/S1022795420120042

Choi, J. Y., Zaidem, M., Gutaker, R., Dorph, K., Singh, R. K., & Purugganan, M. D. (2019). The complex geography of domestication of the African rice *Oryza glaberrima*. *PLoS Genetics*, *15*(3), e1007414. https://doi.org/10.1371/journal.pgen.1007414

Choudhury, R., Singh, D., Singh, N., Kumar, A. K., & Srinivasan, S. (2014). Analysis of genetic diversity and population structure of rice germplasm from north-eastern region of India and development of a core germplasm set. *PLoS One*, *9*(11), 113094. https://doi.org/10.1371/journal.pone.0113094

Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H., & Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, *15*(11), 1496–1502. https://doi.org/10.1101/gr.4107905

Eckshtain-Levi, N., Weisberg, A. J., & Vinatzer, B. A. (2018). The population genetic test Tajima's D identifies genes encoding pathogen-associated molecular patterns and other virulence-related genes in *Ralstonia solanacearum*. *Molecular Plant Pathology*, *19*(9), 2187–2192. https://doi.org/10.1111/mpp.12688

Falcon, W. P., Naylor, R. L., & Shankar, N. D. (2022). Rethinking global food demand for 2050. *Population and Development Review*, *48*(4), 921–957. https://doi.org/10.1111/PADR.12508

FAO. (2014). *Genebank standards for plant genetic resources for food and agriculture* (rev. ed.). Commission on Genetic Resources for Food and Agriculture. https://www.fao.org/4/i3704e/i3704e.pdf

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*, 925–929. https://doi.org/10.1111/2041-210X.12382

Gao, L. (2004). Population structure and conservation genetics of wild rice *Oryza rufipogon* (Poaceae): A region-wide perspective from microsatellite variation. *Molecular Ecology*, *13*(5), 1009–1024. https://doi.org/10.1111/j.1365-294X.2004.02108.x

Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., & McCouch, S. (2005). Genetic structure and diversity in *Oryza sativa* L. *Genetics*, *169*(3), 1631–1638. https://doi.org/10.1534/genetics.104.035642

Gouda, A. C., Ndjiondjop, M. N., Djedatin, G. L., Warburton, M. L., Goungoulou, A., Kpeki, S. B., & Diaye, A. N. (2020). Comparisons of sampling methods for assessing intra- and inter-accession genetic diversity in three rice species using genotyping by sequencing. *Scien-*

tific Reports, *10*, Article 13995. https://doi.org/10.1038/s41598-020-70842-0

Gouda, A. C., Warburton, M. L., Djedatin, G. L., Kpeki, S. B., Wambugu, P. W., Gnikoua, K., & Ndjiondjop, M. N. (2021). Development and validation of diagnostic SNP markers for quality control genotyping in a collection of four rice (Oryza) species. *Scientific Reports*, *11*(1), Article 18617. https://doi.org/10.1038/s41598-021-97689-3

Govindaraju, D. R. (1989). Variation in gene flow levels among predominantly self-pollinated plants. *Journal of Evolutionary Biology*, *2*(3), 173–181. https://doi.org/10.1046/j.1420-9101.1989.2030173.x

Gravel, S., Henn, B. M., Gutenkunst, R. N., Indap, A. R., Marth, G. T., Clark, A. G., Yu, F., Gibbs, R. A., & Bustamante, C. D. (2011). Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 11983–11988. https://doi.org/10.1073/pnas.1019276108

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*(1), Article 100. https://doi.org/10.1038/S43586-022-00184-W

Gutaker, R. M., Groen, S. C., Bellis, E. S., Choi, J. Y., Pires, I. S., Bocinsky, R. K., Slayton, E. R., Wilkins, O., Castillo, C. C., Negrão, S., Oliveira, M. M., Fuller, D. Q., d'Alpoi Guedes, J. A., Lasky, J. R., & Purugganan, M. D. (2020). Genomic history and ecology of the geographic spread of rice. *Nature Plants*, *6*(5), 492–502. https://doi.org/10.1038/s41477-020-0659-6

Halewood, M., Jamora, N., Noriega, I. L., Anglin, N. L., Wenzl, P., Payne, T., Ndjiondjop, M. N., Guarino, L., Lava Kumar, P., Yazbek, M., Muchugi, A., Azevedo, V., Tchamba, M., Jones, C. S., Venuprasad, R., Roux, N., Rojas, E., & Lusty, C. (2020). Germplasm acquisition and distribution by CGIAR genebanks. *Plants*, *9*(10), 1296. https://doi.org/10.3390/plants9101296

Hour, A. L., Hsieh, W. H., Chang, S. H., Wu, Y. P., Chin, H. S., & Lin, Y. R. (2020). Genetic diversity of landraces and improved varieties of rice (*Oryza sativa* L.) in Taiwan. *Rice*, *13*(1), 82. https://doi.org/10.1186/s12284-020-00445-w

Huggins, T. D., Chen, M. H., Fjellstrom, R. G., Jackson, A. K., McClung, A. M., & Edwards, J. D. (2019). Association analysis of three diverse rice (Oryza sativa L.) germplasm collections for loci regulating grain quality traits. *The Plant Genome*, *12*(1), 170085. https://doi.org/10.3835/plantgenome2017.09.0085

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Kimwemwe, P. K., Bukomarhe, C. B., Mamati, E. G., Githiri, S. M., Civava, R. M., Mignouna, J., Kimani, W., & Fofana, M. (2023). 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*, *13*(7), 1906. https://doi.org/10.3390/agronomy13071906

Konishi, S., Izawa, T., Lin, S. Y., Ebana, K., Fukuta, Y., Sasaki, T., & Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. *Science*, *312*(5778), 1392–1396. https://doi.org/10.1126/science.1126410

Kumar, A., Kumar, S., Singh, K. B., Prasad, M., Thakur, J. K., & Kbm, S. (2020). Designing a mini-core collection effectively representing 3004 diverse rice accessions. *Plant Communications*, *1*, 100049. https://doi.org/10.1016/j.xplc.2020.100049

Kumar, J., & Agrawal, V. (2019). Assessment of genetic diversity, population structure and sex identification in dioecious crop, *Trichosanthes dioica* employing ISSR, SCoT and SRAP markers. *Heliyon*, *5*(3), e01346. https://doi.org/10.1016/j.heliyon.2019.e01346

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lachance, J., & Tishkoff, S. A. (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *BioEssays*, *35*(9), 780–786. https://doi.org/10.1002/bies.201300014

Lakew, T., Tanaka, K., & Ishikawa, R. (2021). Genetic diversity of African wild rice (*Oryza longistaminata* Chev. et Roehr) at the edge of its distribution. *Genetic Resources and Crop Evolution*, *68*(5), 1769–1784. https://doi.org/10.1007/S10722-020-01080-6

Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, *74*(1), 175–195. https://doi.org/10.1093/genetics/74.1.175

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Liu, X., Mo, L., Guo, X., Zhang, Q., Li, H., Liu, D., & Lu, H. (2021). How cysteine protease gene PtCP5 affects seed germination by mobilizing storage proteins in *Populus trichocarpa*. *International Journal of Molecular Sciences*, *22*(23), 12637. https://doi.org/10.3390/ijms222312637

Londo, J. P., Chiang, Y. C., Hung, K. H., Chiang, T. Y., & Schaal, B. A. (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(25), 9578–9583. https://doi.org/10.1073/pnas.0603152103

MacKay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: Challenges and prospects. *Nature Reviews Genetics*, *10*(8), 565–577. https://doi.org/10.1038/nrg2612

Maruyama, T., & Fuerst, P. A. (1985). Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics*, *111*(3), 675–689. https://doi.org/10.1093/genetics/111.3.675

Mba, C., & Ogbonnaya, F. C. (2022). Utilizing plant genetic resources to develop climate resilient crops. In O. O. Obembe, E. O. Ekundayo, A. S. Okoli, A. Gidado, C. O. Adetunji, A. B. Ibrahim, & B. E. Ubi (Eds.), *Agricultural biotechnology, biodiversity and bioresources conservation and utilization* (pp. 373–404). CRC Press. https://doi.org/10.1201/9781003178880-22

Mccouch, S. R., Cho, Y. G., Kang, M.-R., Eun, M. Y., Kuiper, M., Pot, J., Groenen, J. T. M., Mccouch, S. R., Kuiper, M., & Eun, M. Y. (1998). Communicated by G. Wenzel Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, *97*, 370–380. http://probe

McCouch, S. R., McNally, K. L., Wang, W., & Hamilton, R. S. (2012). Genomics of gene banks: A case study in rice. *American Journal of Botany*, *99*(2), 407–423. https://doi.org/10.3732/ajb.1100385

Melaku, G., Labroo, M., Liyu, H., Shilai, Z., Guangfu, H., Jing, Z., Tesfaye, K., Haileselassie, T., & Hu, F. (2019). Genetic diversity and differentiation of the African wild rice (*Oryza longistaminata* chev. et roehr) in Ethiopia. *Scientific African*, *6*, e00138. https://doi.org/10.1016/j.sciaf.2019.e00138

Meyer, R. S., Choi, J. Y., Sanches, M., Plessis, A., Flowers, J. M., Amas, J., Dorph, K., Barretto, A., Gross, B., Fuller, D. Q., Bimpong, I. K., Ndjiondjop, M. N., Hazzouri, K. M., Gregorio, G. B., & Purugganan, M. D. (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nature Genetics*, *48*(9), 1083–1088. https://doi.org/10.1038/ng.3633

Mohammadi, S. A., & Prasanna, B. M. (2003). Analysis of genetic diversity in crop plants—Salient statistical tools and considerations. *Crop Science*, *43*(4), 1235–1248. https://doi.org/10.2135/cropsci2003.1235

Mohidem, N. A., Hashim, N., Shamsudin, R., & Man, H. C. (2022). Rice for food security: Revisiting its production, diversity, rice milling process and nutrient content. *Agriculture*, *12*(6), 741. https://doi.org/10.3390/agriculture12060741

Mommer, L., Dumbrell, A. J., Wagemaker, C. A. M., & Ouborg, N. J. (2011). Belowground DNA-based techniques: Untangling the network of plant root interactions. *Plant and Soil*, *348*(1–2), 115–121. https://doi.org/10.1007/s11104-011-0962-0

Montcho, D., Sow, M., Kolade, O., Bocco, R., & Agbangla, C. (2021). Genetic diversity study of African rice (*Oryza glaberrima*) and its wild relatives using microsatellites markers. *Annual Research & Review in Biology*, *36*(5), 75–87. https://doi.org/10.9734/arrb/2021/v36i530378

Moragues, M., Comadran, J., Waugh, R., Milne, I., Flavell, A. J., & Russell, J. R. (2010). Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *Theoretical and Applied Genetics*, *120*(8), 1525–1534. https://doi.org/10.1007/s00122-010-1273-1

Nachimuthu, V. V., Raveendran, M., Duraialaguraja, S., Sivakami, R., Pandian, B. A., Ponniah, G., Gunasekaran, K., Swaminathan, M., Suji, K. K., & Sabariappan, R. (2015). Analysis of population structure and genetic diversity in rice germplasm using SSR markers: An initiative towards association mapping of agronomic traits in *Oryza sativa*. *Rice*, *8*(1), Article 30. https://doi.org/10.1186/s12284-015-0062-5

Ndjiondjop, M. N., Alachiotis, N., Pavlidis, P., Goungoulou, A., Kpeki, S. B., Zhao, D., & Semagn, K. (2019). Comparisons of molecular diversity indices, selective sweeps and population structure of African rice with its wild progenitor and Asian rice. *Theoretical and Applied Genetics*, *132*(4), 1145–1158. https://doi.org/10.1007/s00122-018-3268-2

Ndjiondjop, M. N., Albar, L., Fargette, D., Fauquet, C., & Ghesquière, A. (1999). The genetic basis of high resistance to rice yellow mottle virus (RYMV) in cultivars of two cultivated rice species. *Plant Disease*, *83*(10), 931–935. https://doi.org/10.1094/PDIS.1999.83.10.931

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. (2022). Genetic variation and population structure of *Oryza sativa* accessions in the AfricaRice collection and development of the AfricaRice *O. sativa* core collection. *Crop Science*, *63*(2), 724–739. https://doi.org/10.1002/CSC2.20898

Ndjiondjop, M. N., Manneh, B., Cissoko, M., Drame, N. K., Kakai, R. G., Bocco, R., Baimey, H., & Wopereis, M. (2010). Drought resistance in an interspecific backcross population of rice (*Oryza* spp.) derived from the cross WAB56-104 (*O. sativa*)×CG14 (*O. glaberrima*). *Plant Science*, *179*(4), 364–373. https://doi.org/10.1016/j.plantsci.2010.06.006

Ndjiondjop, M. N., Semagn, K., Gouda, A. C., Kpeki, S. B., Tia, D. D., Sow, M., Goungoulou, A., Sie, M., Perrier, X., Ghesquiere, A., & Warburton, M. L. (2017). Genetic variation and population structure of *Oryza glaberrima* and development of a mini-core collection using DArTseq. *Frontiers in Plant Science*, *8*, 1748. https://doi.org/10.3389/fpls.2017.01748

Ndjiondjop, M. N., Semagn, K., Sow, M., Manneh, B., Gouda, A. C., Kpeki, S. B., Pegalepo, E., Wambugu, P., Sié, M., & Warburton, M. L. (2018a). Assessment of genetic variation and population structure of diverse rice genotypes adapted to lowland and upland ecologies in Africa using SNPs. *Frontiers in Plant Science*, *9*, 446. https://doi.org/10.3389/fpls.2018.00446

Ndjiondjop, N. M., Kassa, S., Jianwei, Z., Comlan, G. A., Bienvenu, K. S., Goungoulou, A., Peterson, W., Nani, D. K., Bimpong, I. K., & Dule, Z. (2018b). Development of species diagnostic SNP markers for quality control genotyping in four rice (*Oryza* L.) species. *Molecular Breeding*, *38*(11), Article 131. https://doi.org/10.1007/s11032-018-0885-z

Ni, J., Colowit, P. M., & Mackill, D. J. (2002). Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop Science*, *42*(2), 601–607. https://doi.org/10.2135/cropsci2002.6010

Nielsen, R. (2000). Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics*, *154*(2), 931–942. https://doi.org/10.1093/genetics/154.2.931

Nielsen, R. (2001). Statistical tests of selective neutrality in the age of genomics. *Heredity*, *86*(6), 641–647. https://doi.org/10.1046/j.1365-2540.2001.00895.x

Nielsen, R. (2004). Population genetic analysis of ascertained SNP data. *Human Genomics*, *1*(3), Article 218. https://doi.org/10.1186/1479-7364-1-3-218

OECD-FAO. (2021). *OECD-FAO agricultural outlook 2021–2030*. OECD Publishing.

Orjuela, J., Sabot, F., Chéron, S., Vigouroux, Y., Adam, H., Chrestin, H., Sanni, K., Lorieux, M., & Ghesquière, A. (2014). An extensive analysis of the African rice genetic diversity through a global genotyping. *Theoretical and Applied Genetics*, *127*(10), 2211–2223. https://doi.org/10.1007/s00122-014-2374-z

Pathaichindachote, W., Panyawut, N., Sikaewtung, K., Patarapuwadol, S., & Muangprom, A. (2019). Genetic diversity and allelic frequency of selected Thai and exotic rice germplasm using SSR markers. *Rice Science*, *26*(6), 393–403. https://doi.org/10.1016/j.rsci.2018.11.002

Pemberton, J. (2004). Introduction to conservation genetics. In R. Frankham, J. D. Ballou, & D. A. Briscoe (Eds.). Cambridge University Press. *Genetical Research*, *83*(3). https://doi.org/10.1017/s0016672304216913

Phillips, J. D., Gillis, D. J., & Hanner, R. H. (2018). Introduction incomplete estimates of genetic diversity within species: Implications for DNA barcoding. *Ecology and Evolution*, *9*(5), 2996–3010. https://doi.org/10.1002/ece3.4757

Phung, N. T. P., Mai, C. D., Mournet, P., Frouin, J., Droc, G., Ta, N. K., Jouannic, S., Lê, L. T., Do, V. N., Gantet, P., & Courtois, B. (2014). Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. *BMC Plant Biology*, *14*(1), Article 371. https://doi.org/10.1186/s12870-014-0371-7

Quiones-Prez, C. Z., Senz-Romero, C., & Wehenkel, C. (2014). Genetic diversity and conservation of *Picea chihuahuana* Martnez: A review.

African Journal of Biotechnology, *13*(28), 2786–2795. https://doi.org/10.5897/ajb2014.13645

Ramirez-Villegas, J., Khoury, C. K., Achicanoy, H. A., Diaz, M. V., Mendez, A. C., Sosa, C. C., Kehel, Z., Guarino, L., Abberton, M., Aunario, J., Al Awar, B., Alarcon, J. C., Amri, A., Anglin, N. L., Azevedo, V., Aziz, K., Capilit, G. L., Chavez, O., Chebotarov, D., … Zavala, C. (2022). State of ex situ conservation of landrace groups of 25 major crops. *Nature Plants*, *8*(5), 491–499. https://doi.org/10.1038/s41477-022-01144-8

Razak, S. A., Ezzah, H., Azman, N., Kamaruzaman, R., Saidon, S. A., Fairuz, M., Yusof, M., Ismail, S. N., Azwan Jaafar, M., & Abdullah, N. (2020). Genetic diversity of released Malaysian rice varieties based on single nucleotide polymorphism markers. *Original Paper Czech Journal of Genetics and Plant Breeding*, *56*(2), 62–70. https://doi.org/10.17221/58/2019-CJGPB

Reif, J. C., Zhang, P., Dreisigacker, S., Warburton, M. L., Van Ginkel, M., Hoisington, D., Bohn, M., & Melchinger, A. E. (2005). Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics*, *110*(5), 859–864. https://doi.org/10.1007/s00122-004-1881-8

Reig-Valiente, J. L., Viruel, J., Sales, E., Marqués, L., Terol, J., Gut, M., Derdak, S., Talón, M., & Domingo, C. (2016). Genetic diversity and population structure of rice varieties cultivated in temperate regions. *Rice*, *9*(1), 58. https://doi.org/10.1186/s12284-016-0130-5

Salem, K. F. M., & Sallam, A. (2016). Analysis of population structure and genetic diversity of Egyptian and exotic rice (*Oryza sativa* L.) genotypes. *Comptes Rendus—Biologies*, *339*(1), 1–9. https://doi.org/10.1016/j.crvi.2015.11.003

Sansaloni, C., Franco, J., Santos, B., Percival-Alwyn, L., Singh, S., Petroli, C., Campos, J., Dreher, K., Payne, T., Marshall, D., Kilian, B., Milne, I., Raubach, S., Shaw, P., Stephen, G., Carling, J., Pierre, C. S., Burgueño, J., Crosa, J., … Pixley, K. (2020). Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. *Nature Communications*, *11*(1), 4572. https://doi.org/10.1038/s41467-020-18404-w

Sarla, N., & Swamy, B. P. M. (2005). *Oryza glaberrima*: A source for the improvement of *Oryza sativa*. *Current Science*, *89*(6), 955–963.

Schierenbeck, K. A. (2017). Population-level genetic variation and climate change in a biodiversity hotspot. *Annals of Botany*, *119*(2), 215–228. https://doi.org/10.1093/aob/mcw214

Schmidt, D., & Pool, J. (2002). *The effect of population history on the distribution of the Tajima's D statistic*. http://www.deenaschmidt.com/TajimasD.pdf

Sekiya, N., Oizumi, N., Kessy, T. T., Fimbo, K. M. J., Tomitaka, M., Katsura, K., & Araki, H. (2020). Importance of market-oriented research for rice production in Tanzania. A review. *Agronomy for Sustainable Development*, *40*(1), Article 7. https://doi.org/10.1007/s13593-020-0611-1

Semon, M., Nielsen, R., Jones, M. P., & Mccouch, S. R. (2005). The population structure of African cultivated rice Oryza glaberrima (Steud.): Evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. *Genetics*, *169*(3), 1639–1647. https://doi.org/10.1534/genetics.104.033175

Singh, N., Choudhury, D. R., Singh, A. K., Kumar, S., Srinivasan, K., Tyagi, R. K., Singh, N. K., & Singh, R. (2013). Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS One*, *8*(12), e84136. https://doi.org/10.1371/journal.pone.0084136

Song, Z., Li, B., Chen, J., & Lu, B. R. (2005). Genetic diversity and conservation of common wild rice (*Oryza rufipogon*) in China. *Plant Species Biology*, *20*(2), 83–92. https://doi.org/10.1111/j.1442-1984.2005.00128.x

Song, Z. P., Lu, B. R., Zhu, Y. G., & Chen, J. K. (2003). Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. *New Phytologist*, *157*(3), 657–665. https://doi.org/10.1046/j.1469-8137.2003.00699.x

Soullier, G., Demont, M., Arouna, A., Lançon, F., & Mendez del Villar, P. (2020). The state of rice value chain upgrading in West Africa. *Global Food Security*, *25*, 100365. https://doi.org/10.1016/j.gfs.2020.100365

Sow, M., Ndjiondjop, M. N., Sido, A., Mariac, C., Laing, M., & Bezançon, G. (2014a). Genetic diversity, population structure and differentiation of rice species from Niger and their potential for rice genetic resources conservation and enhancement. *Genetic Resources and Crop Evolution*, *61*(1), 199–213. https://doi.org/10.1007/s10722-013-0026-9

Sow, M., Sido, A., Laing, M., & Ndjiondjop, M. N. (2014b). Agro-morphological variability of rice species collected from Niger. *Plant Genetic Resources: Characterisation and Utilisation*, *12*(1), 22–34. https://doi.org/10.1017/S1479262113000221

Sun, D., Robbins, K., Morales, N., Shu, Q., & Cen, H. (2022). Advances in optical phenotyping of cereal crops. *Trends in Plant Science*, *27*(2), 191–208. https://doi.org/10.1016/J.TPLANTS.2021.07.015

Suvi, W. T., Shimelis, H., Laing, M., Mathew, I., & Shayanowako, A. I. T. (2019). Assessment of the genetic diversity and population structure of rice genotypes using SSR markers. *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, *70*(1), 76–86. https://doi.org/10.1080/09064710.2019.1670859

Sweeney, M., & McCouch, S. (2007). The complex history of the domestication of rice. *Annals of Botany*, *100*(5), 951–957. https://doi.org/10.1093/aob/mcm128

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Vanlalsanga, & Singh, Y. T. (2019). Genetic diversity and population structure in upland rice (*Oryza sativa* L.) of Mizoram, North East India as revealed by morphological, biochemical and molecular markers. *Biochemical Genetics*, *57*(3), 421–442. https://doi.org/10.1007/s10528-018-09901-1

Veltman, M. A., Flowers, J. M., van Andel, T. R., & Eric Schranz, M. (2019). Origins and geographic diversification of African rice (*Oryza glaberrima*). *PLoS One*, *14*(3), 1–28. https://doi.org/10.1371/journal.pone.0203508

Vilayheuang, K., Borrayo, E., Kawase, M., & Watanabe, K. N. (2020). Development of Laos Khao Kai Noi rice landrace (*Oryza sativa* L.) core collection as a model for rice genetic resources management in the Laos National Genebank. *IOP Conference Series: Earth and Environmental Science*, *482*(1), 012039. https://doi.org/10.1088/1755-1315/482/1/012039

Wambugu, P. W., Ndjiondjop, M. N., & Henry, R. (2019). Advances in molecular genetics and genomics of African rice (*Oryza glaberrima* steud). *Plants*, *8*(10), 376. https://doi.org/10.3390/plants8100376

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R. R., Zhang, F., Mansueto, L., Copetti, D., Sanciangco, M., Palis, K. C., Xu, J., Sun, C., Fu, B., Zhang, H., Gao, Y., … Leung, H. (2018). Genomic variation in 3,010 diverse acces-

sions of Asian cultivated rice. *Nature*, *557*(7703), 43–49. https://doi.org/10.1038/s41586-018-0063-9

Weise, S., Lohwasser, U., & Oppermann, M. (2020). Document or lose it—On the importance of information management for genetic resources conservation in Genebanks. *Plants*, *9*(8), 1050. https://doi.org/10.3390/PLANTS9081050

Whitfield, S., Chapman, S., Mahop, M. T., Deva, C., Masamba, K., & Mwamahonje, A. (2021). Exploring assumptions in crop breeding for climate resilience: Opportunities and principles for integrating climate model projections. *Climatic Change*, *164*(3–4), 1–18. https://doi.org/10.1007/S10584-021-02997-2/FIGURES/5

Xu, Q., Yuan, X., Wang, S., Feng, Y., Yu, H., Wang, Y., Yang, Y., Wei, X., & Li, X. (2016). The genetic diversity and structure of indica rice in China as detected by single nucleotide polymorphism analysis. *BMC Genetics*, *17*(1), 53. https://doi.org/10.1186/s12863-016-0361-x

Yan, W. G., Agrama, H., Jia, M., Fjellstrom, R., & McClung, A. (2010). Geographic description of genetic diversity and relationships in the USDA rice world collection. *Crop Science*, *50*(6), 2406–2417. https://doi.org/10.2135/cropsci2010.02.0096

Zachariah Peery, M., Kirby, R., Reid, B. N., Stoelting, R., Doucet-Bëer, E., Robinson, S., Vásquez-Carrillo, C., Pauli, J. N., & Palsboll, P. J. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology*, *21*(14), 3403–3418. https://doi.org/10.1111/j.1365-294X.2012.05635.x

Zhang, F., Wang, C., Li, M., Cui, Y., Shi, Y., Wu, Z., Hu, Z., Wang, W., Xu, J., & Li, Z. (2021). The landscape of gene–CDS–haplotype diversity in rice: Properties, population organization, footprints of domestication and breeding, and implications for genetic improvement. *Molecular Plant*, *14*(5), 787–804. https://doi.org/10.1016/J.MOLP.2021.02.003

Zhang, L., Ma, X., Zhang, X., Xu, Y., Ibrahim, A. K., Yao, J., Huang, H., Chen, S., Liao, Z., Zhang, Q., Niyitanga, S., Yu, J., Liu, Y., Xu, X., Wang, J., Tao, A., Xu, J., Chen, S., Yang, X., … Zhang, L. (2021). Reference genomes of the two cultivated jute species. *Plant Biotechnology Journal*, *19*(11), 2235–2248. https://doi.org/10.1111/PBI.13652

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gouda, A. C., Sangare, J. R., Gnikoua, K., Wambugu, P., Huggins, T. D., & Ndjiondjop, M. N. (2024). Genetic variation and population structure of the rice accessions maintained in the AfricaRice genebank using DArTseq. *Crop Science*, 1–20. https://doi.org/10.1002/csc2.21395