agronomy

MDPI

Article

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)

Paul Kitenge Kimwemwe [1,2,3,4,*,†], Chance Bahati Bukomarhe [1,2,3,†], Edward George Mamati [1], Stephen Mwangi Githiri [1], René Mushizi Civava [3,5], Jacob Mignouna [2], Wilson Kimani [6,*] and Mamadou Fofana [2]

[1] Department of Horticulture and Food Security, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi P.O. Box 62000-00200, Kenya; bukomarhe.chance@students.jkuat.ac.ke (C.B.B.); egmamati@agr.jkuat.ac.ke (E.G.M.); githirim@agr.jkuat.ac.ke (S.M.G.)
[2] Olusegun O. Research Campus, International Institute of Tropical Agriculture (IITA), Bukavu P.O. Box 1222, Democratic Republic of the Congo; j.mignouna@cgiar.org (J.M.); ma.fofana@cgiar.org (M.F.)
[3] Institut National pour l'Etude et la Recherche Agronomiques (INERA), Kinshasa P.O. Box 2037, Democratic Republic of the Congo; renecivava@uea.ac.cd
[4] Faculty of Agriculture and Environmental Sciences, Université de Kalemie (UNIKAL), Kalemie P.O. Box 570, Democratic Republic of the Congo
[5] Faculty of Agriculture and Environmental Sciences, Université Evangélique en Afrique (UEA), Bukavu P.O. Box 3323, Democratic Republic of the Congo
[6] International Livestock Research Institute (ILRI), Nairobi P.O. Box 30709-00100, Kenya
* Correspondence: kimwemwe.paul@students.jkuat.ac.ke (P.K.K.); w.kimani@cgiar.org (W.K.); Tel.: +243-824484040 (P.K.K.)
† These authors contributed equally to this work.

Abstract: Understanding the genetic diversity and population structure of rice is crucial for breeding programs, conservation efforts, and the development of sustainable agricultural practices. This study aimed to assess the genetic diversity and population structure of 94 rice (*Oryza sativa* L.) genotypes from the Democratic Republic of Congo using a set of 8389 high-quality DArTseq-based single nucleotide polymorphism (SNP) markers. The average polymorphic information content (PIC) of the markers was 0.25. About 42.4% of the SNPs had a PIC value between 0.25 and 0.5, which were moderately informative. The ADMIXTURE program was used for structure analysis, which revealed five sub-populations (K = 5), with admixtures. In principal component analysis (PCA), the first three principal components accounted for 36.3% of the total variation. Analysis of molecular variance revealed significant variation between sub-populations (36.09%) and within genotypes (34.04%). The low overall number of migrants (Nm = 0.23) and high fixation index (F_{st} = 0.52) indicated limited gene flow and significant differentiation between the sub-populations. Observed heterozygosity (H_o = 0.08) was lower than expected heterozygosity (H_e = 0.14) because of the high inbreeding (F_{is} = 0.52) nature of rice. A high average Euclidean genetic distance (0.87) revealed the existence of genetic diversity among the 94 genotypes. The significant genetic diversity among the evaluated rice genotypes can be further explored to obtain potentially desirable genes for rice improvement.

Keywords: population structure; genetic diversity; rice germplasm; DArTseq; single nucleotide polymorphism

check for updates

Citation: Kimwemwe, P.K.; Bukomarhe, C.B.; Mamati, E.G.; Githiri, S.M.; Civava, R.M.; Mignouna, J.; Kimani, W.; Fofana, M. Population Structure and Genetic Diversity of Rice (*Oryza sativa* L.) Germplasm from the Democratic Republic of Congo (DRC) Using DArTseq- Derived Single Nucleotide Polymorphism (SNP). *Agronomy* **2023**, *13*, 1906. https://doi.org/10.3390/agronomy13071906

Academic Editors: Evangelia V. Avramidou, Aliki Kapazoglou and Ermioni Malliarou

Received: 5 June 2023
Revised: 26 June 2023
Accepted: 27 June 2023
Published: 19 July 2023

1. Introduction

Rice (*Oryza sativa* L.) is an important cereal crop that serves as a staple food for over 50% of the global population. Rice cultivation, processing, and distribution support

millions of livelihoods and contribute to national and international trade, making rice a vital component of economic growth and food security worldwide [1]. Rice provides over 20% of the calories consumed worldwide. In the Democratic Republic of Congo (DRC), rice is the second most-consumed cereal after maize, providing food and income for many households [2]. Rice is cultivated in a wide range of agro-climatic conditions worldwide [3].

The cultivated rice, *O. sativa*, was derived from its wild counterparts, *O. nivara* or *O. rufipogon*, and is classified into two distinct varietal groups, ssp. indica and ssp. japonica [4]. During the process of domestication, selective sweeps led to a significant reduction in the genetic diversity of *O. sativa* [5]. Several studies have focused on examining the genetic structure of rice cultivars at a local scale, typically within a specific country. These localized investigations offer a comprehensive understanding of rice genetic diversity within a given region, shedding light on the intricate relationship between genetic diversity and human cultivation practices [6–11]. Furthermore, these studies provide valuable insights for developing conservation strategies specific to the region. Due to the combined effects of natural and artificial selection in diverse environments, rice exhibits considerable morphological, physiological, and genetic diversity, with more than 120,000 distinct rice varieties identified [4,6].

For centuries, farmers in DRC have practiced traditional methods of rice cultivation by selecting and preserving seeds from the best-performing plants. Over time, these processes have resulted in the development of landraces and locally adapted rice varieties with distinct traits that are well-suited to the region's diverse agro-ecological conditions [12]. To address the changing demands for rice, the national rice breeding program has actively selected and introduced new varieties aimed at enhancing rice productivity while preserving the genetic diversity and cultural significance of rice in the region. With its potential, four million hectares of irrigable lowland, DRC could play a significant role in ensuring food security by contributing to the increase in rice production and availability in the Sub-Saharan Africa region.

According to various forecasts, the world population is expected to grow significantly and reach 9.7 billion by 2050 [13]. Among the nine countries that are projected to have over 50% population growth between now and 2050, the majority are located in tropical regions [3]. These countries include India, Nigeria, the Democratic Republic of Congo, Ethiopia, the United Republic of Tanzania, Indonesia, and Egypt [13]. In these countries, rice is the fastest-growing and preferred food commodity, driven by high population growth, rapid urbanization, and changes in eating habits [14,15]. However, rice productivity in the region is low, and crop yields are often impacted by various biotic and abiotic stresses, including pests, diseases, and adverse environmental factors [2,16].

Utilizing plants' genetic diversity, breeders develop new and improved crop varieties with desirable traits, such as nutritional and grain quality, resistance to pests and diseases, tolerance to flood and drought, and improved yield, in order to tackle worldwide issues related to food security, sustainability, and adaptability to climate change [17]. Understanding the genetic diversity and population structure of rice germplasm available to breeders is very critical for rice improvement. According to Khan et al. [18], greater genetic variability enhances the probability of identifying superior genotypes within a population. As pointed out by Novoselović et al. [19], a limited genetic base is a significant challenge that renders plants more susceptible to biotic and abiotic stress conditions.

The morphological markers have been effectively used for diversity studies especially for traits such as height, yield, maturity, and grain color, as well as resistance to insects and diseases [20,21]. Nevertheless, the exclusive reliance on morphological markers was shown to be unreliable because they are limited in number, susceptible to environmental factors, and influenced by the plant growth stages [22,23]. Therefore, molecular markers have become an indispensable tool in genetic research [24], particularly in assessing genetic diversity. Molecular markers allow precise and rapid varietal identification, germplasm characterization, collection, and management. Earlier molecular markers such as random amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR), and ampli-

fied fragment length polymorphism (AFLP) have been frequently used for fingerprinting and characterization of varieties and germplasm accessions of different crop species [24]. Despite the potential benefits of these molecular markers in plant breeding, their usefulness is limited due to their reliance on prior sequence information, which can be costly and time-consuming to obtain [24]. Genotyping-by-sequencing (GBS) techniques such as diversity array technology sequencing (DArTseq) is a low-cost and rapid genotyping method that enables the screening of hundreds of highly polymorphic markers without previous sequence information for the detection of loci [23,25]. DArTseq further reduces genome representation and provides abundant markers [26], thus improving the rate of genotype calling and the ability to sequence more samples at a lower cost. DArTseq produces both dominant (SilicoDArT) and co-dominant (SNP) markers that have successfully been applied for genetic diversity and population structure analysis in several crops such as cassava [27], taro [28], rice [29], rye [30], sorghum [31], and wheat [19]. The use of DArTseq markers enables a comprehensive analysis of genetic diversity, complete genome profiling, and high-density mapping of complex traits, all of which are crucial for marker-based breeding [24].

This study aimed to evaluate the genetic diversity and population structure of rice germplasm accessions in the eastern DRC, utilizing DArTseq-derived SNP markers. It provides a better understanding of the level of genetic variation of rice germplasm resources for both the future investigations into superior genes and the identification of parental lines for rice improvement programs.

2. Materials and Methods

2.1. Plant Materials

Ninety-four rice genotypes (Table 1), available at the rice breeding program under the Institut National pour l'Etude et la Recherche Agronomiques (INERA) in the Eastern DRC, were utilized in this study. This collection was comprised of genotypes from different research centers, including AfricaRice (43), International Rice Research Institute (IRRI)-Burundi (25), IRRI-Kenya (9), INERA (12), and local landraces (5).

2.2. Sample Preparation and DNA Extraction

Twenty seeds from each of the 94 genotypes were pre-germinated through water-soaking and incubating at 28 °C in a growth chamber for 48 h. The sprouted grains were then sown on seedling nursery trays and raised in a greenhouse at Jomo Kenyatta University of Agriculture and Technology (JKUAT) for two weeks. Leaf samples were sent to SEQART AFRICA at the International Livestock Research Institute (ILRI), Nairobi, for genotyping. Total genomic DNA was extracted using the NucleoMag® plant genomic DNA extraction kit (Macherey-Nagel™, Dueren, Germany). The quality and quantity of DNA were assessed using 0.8% agarose gel electrophoresis and a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific™, Waltham, MA, USA), respectively.

2.3. Genotypic Data

Genotyping was done using DArTseq method, which relies on genome complexity reduction using enzymes [26,32]. The extracted genomic DNA was digested with two restriction enzymes (*PstI* and *MseI*). The resulting fragments were then ligated to adapters that contained unique barcodes for each sample [25]. The ligated fragments were then amplified by PCR to generate a library of DNA fragments for sequencing. The library fragments were then sequenced using single read sequencing runs of 77 cycles by the Illumina Hiseq2500. DArTseq markers scoring was achieved using DArTsoft14, an in-house marker scoring pipeline. The markers of the single nucleotide polymorphisms (SNPs) were scored.

2.4. Marker Filtering and Quality Analysis

The generated SNP markers were aligned to the rice reference genome, *Oryza sativa* V7.0 [33], to determine their positions along the 12 rice chromosomes. The criteria for data filtration were as follows: non-informative monomorphic markers were removed, markers with a call rate >95% and minor allele frequency >5% were retained. The VCFtools V0.1.13 [34] software was used for marker filtration. To analyze the characteristics and distribution of the markers along the 12 rice chromosomes, parameters such as polymorphic information content (PIC), reproducibility, and call rate were determined using the dartR package in R [35]. Additionally, the same package was utilized to calculate the proportion of mutation types, including transversion (Tv) and transition (Ts), responsible for the observed polymorphism.

Table 1. Rice genotypes used in the study.

Entry No.	Genotype Name	Source/Program	Entry No.	Genotype Name	Source/Program
1	Komboka	IRRI-Burundi	48	D20-ARS-3-2	AfricaRice
2	IR64	IRRI-Burundi	49	IR96279-33-3-1-2	IRRI-Burundi
3	IBEI6	AfricaRice	50	ARS134-B-1-1-5-B	AfricaRice
4	GIZA128	IRRI-Burundi	51	Orylux7-1	AfricaRice
5	Nipponbare	IRRI-Burundi	52	WAHX14N-926	AfricaRice
6	Jasmine	IRRI-Burundi	53	MR254	AfricaRice
7	NL59	AfricaRice	54	Golmy	AfricaRice
8	FKR	AfricaRice	55	ARS848-15-3-2-4	AfricaRice
9	08FAN10	IRRI-Burundi	56	IR93348:32-B-15-3-B-B-B-1	IRRI-Burundi
10	WAB2066-TGR2	AfricaRice	57	ARS168-3-B-1-B	AfricaRice
11	WAB2066-TGR3	AfricaRice	58	IR88638	IRRI-Burundi
12	IR99084-B-B-13	INERA-DRC	59	ARICA12	AfricaRice
13	IR127229	IRRI-Burundi	60	ARICA3	AfricaRice
14	IR106172-78:1-B-B	INERA-DRC	61	IR64-sub-1	IRRI-Burundi
15	ARS848-15-3-2-3	AfricaRice	62	HHZSAL6	AfricaRice
16	IR106364-B-B-CNUS	INERA-DRC	63	ARS755-3-3-1-B	AfricaRice
17	ARS844-24-10-2-B	AfricaRice	64	ARS134-B-1-1-5	AfricaRice
18	ARS168-1-B-3-B	AfricaRice	65	IR990-48-B-B-12	IRRI-Burundi
19	ARS851-1-3	AfricaRice	66	IR64-biofortified	IRRI-Burundi
20	IR87638-10-2-2-4	INERA-DRC	67	IR107015-37	IRRI-Burundi
21	IR98419-B-B-11	INERA-DRC	68	ARS79-5-11-11	AfricaRice
22	IR97071-24-1-1-1	INERA-DRC	69	V18/RRS126-48-1-13-2	AfricaRice
23	ARS803-4-5-4-3	AfricaRice	70	Orylux11	AfricaRice
24	IR93856-23-1-1-1	INERA-DRC	71	ARS134-B-B-B	AfricaRice
25	ARS790-5-11-1-1	AfricaRice	72	Magoti	Local landrace
26	IR17015-6-5-3-B1	INERA-DRC	73	Runingu	Local landrace
27	IR106359-B-18-5	INERA-DRC	74	ARS169-2-B-3-B	AfricaRice
28	IR95624-B-138-3	INERA-DRC	75	ARS134-B-1-1-4	AfricaRice
29	IR13A461	IRRI-Burundi	76	IR82574/643-1-2	IRRI-Burundi
30	Mugwiza	IRRI-Burundi	77	Orylux5	AfricaRice
31	Vuninzara	IRRI-Burundi	78	SAHEL210	AfricaRice
32	IR97045-24-1-1-1	IRRI-Burundi	79	IR841	IRRI-Burundi
33	Kigoma	Local landrace	80	ARS39-145/EP-3	AfricaRice
34	Makasane	IRRI-Burundi	81	ARS101-4-B-1-1-B	AfricaRice
35	Rukaramu	Local landrace	82	ARS101-4-B-1-3	AfricaRice
36	Musesekara	IRRI-Burundi	83	NERICA-L-19-Sab-1	AfricaRice
37	Yasho-Yasho	Local landrace	84	ARS756-1-1-3-B-2-2	AfricaRice
38	Kasozi	IRRI-Burundi	85	ARS563-425-1-B-2-3	AfricaRice
39	IR7525	IRRI-Burundi	86	ARICA4	IRRI-Kenya
40	Orylux7	AfricaRice	87	ARICA17	IRRI-Kenya
41	ART29	INERA-DRC	88	Basmati370	IRRI-Kenya
42	Sipi	INERA-DRC	89	IRAT109	IRRI-Kenya
43	CRS36	IRRI-Burundi	90	NERICA1	IRRI-Kenya
44	ARICA2	AfricaRice	91	NERICA2	IRRI-Kenya
45	NL19	AfricaRice	92	NERICA10	IRRI-Kenya
46	NL14	AfricaRice	93	NERICA12	IRRI-Kenya
47	NL17	AfricaRice	94	PAN84	IRRI-Kenya

2.5. Population Structure and Genetic Diversity Analysis

Population structure was analyzed to gain insight into the ancestry of the 94 rice genotypes. The filtered SNP markers were utilized for structure analysis using ADMIXTURE V1.3.0 [36] which uses a model-based maximum likelihood estimation. The optimal number of sub-populations was determined by evaluating the cross-validation errors based on K values (K = 1–10). As recommended by Alexander and Lange [36], the K value associated with the lowest cross-validation error was considered the ideal number of sub-populations. The population structure was then visualized using R software, Version 4.2.3. Additionally, principal component analysis (PCA) was conducted using the adegenet package in

R [37] to investigate genetic relatedness patterns among the sub-populations. The resulting PCA scores were exported and used to generate a 3D PCA plot using the SRPLOT online platform (http://www.bioinformatics.com.cn/srplot, accessed on 24 April 2023).

The results obtained from ADMIXTURE were used to calculate genetic diversity indices, including observed heterozygosity (H_o), expected heterozygosity (H_e), and inbreeding coefficient (F_{is}) for the sub-populations. These calculations were performed using the adegenet package in R [37]. Analysis of molecular variance (AMOVA), as described by Nei [38], was used to partition the total variance into among and within populations variance using the poppr package of R [39]. The gene flow (Nm) level among the sub-populations was estimated using the formula $Nm = (1 - F_{st})/4F_{st}$, as suggested by Wright [40].

To assess the genetic differentiation among the sub-populations identified in the structure analysis, Nei's pairwise fixation indices (F_{st}) [41] were generated using the hierfstat package in R [42].

2.6. Phylogenetic Analysis

A neighbor-joining (NJ) phylogenetic tree was constructed using the ape package in R [43] to visualize the genetic differentiation among the sub-populations. The relationships among individuals were analyzed by generating a pairwise genetic distance matrix using the Euclidean distance method implemented in R. The resulting phylogenetic tree was created using the hclust function in R and exported in Newick format using the ape package for annotation in the interactive tree of life (iTOL) version 6.5.2 (https://itol.embl.de/, accessed on 19 April 2023) [44].

3. Results

3.1. Characterization of DArTseq-Derived SNP Markers

A total of 31,366 SNP markers were generated from the 94 rice genotypes, out of which 27,831 markers (88.7%) were mapped to the reference genome. After filtering, 8389 informative SNP markers were retained for structure and diversity analyses. The distribution of markers on the 12 rice chromosomes, along with their characteristics after filtering, are presented in Figure 1. The number of SNP markers ranged from 475 (chromosome 10) to 1014 (chromosome 1), with an average of 699 markers per chromosome (Figure 1a).

The 8389 informative SNP markers had an average polymorphic information content (PIC) value of 0.25, ranging between 0 and 0.5 (Figure 1b). The reproducibility of SNP markers varied from 87% to 100%, with a mean of 98%. Approximately 88% of SNP markers exhibited $\geq$95% reproducibility (Figure 1c). The call rate ranged from 96% to 100%, with an average of 97% (Figure 1d).

The SNP mutation types are summarized in Table 2, showing the frequency of transitions (Ts; i.e., A/G, T/C substitutions) and transversions (Tv; i.e., A/T, A/C, T/G or C/G substitutions).

The proportion of polymorphisms due to different transitions ranged from 31.06% (A/G) to 31.54% (T/C). The proportion of polymorphisms due to transversions ranged from 7.5% (C/G) to 10.2% (A/T). Overall, among the SNP variations, transitions (62.6%) were more frequent than transversions (37.4%), with a Ts/Tv ratio of 1.67.

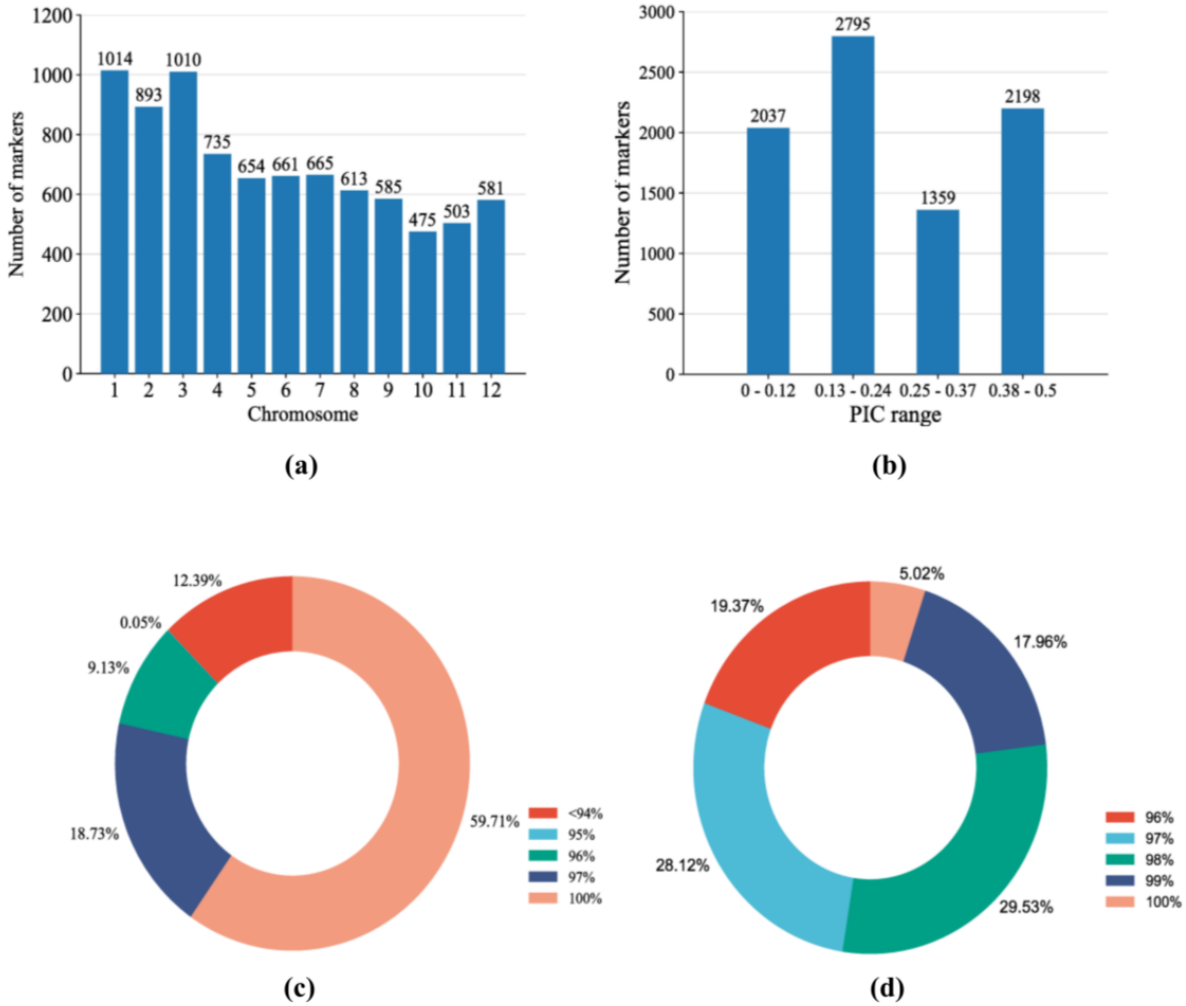

Figure 1. SNP markers' characteristics: (**a**) distribution of SNP markers on the 12 rice chromosomes, (**b**) PIC range values of the SNP markers, (**c**) reproducibility of the SNP markers, and (**d**) call rate of the SNP markers.

Table 2. Proportion of SNP transitions and transversions mutation types across the genomes of the 94 rice genotypes.

	Transitions (Ts)		**Transversions (Tv)**				**Ts/Tv**
Substitution	A/G	T/C	A/T	A/C	T/G	C/G	
Number of alleles	2606	2646	859	829	816	633	
Frequency (%)	31.06	31.54	10.2	9.88	9.7	7.5	1.67
Total	5252 (62.6%)		3137 (37.4%)				

3.2. Population Structure Analysis

Population structure based on a filtered set of 8389 SNP DArTseq markers gave five distinct sub-populations across the 94 genotypes. The sub-populations (referred to as Pop1, Pop2, Pop3, Pop4, and Pop5) (Figure 2) were identified based on the K value corresponding to the lowest cross-validation error in the ADMIXTURE [36]. Considering K = 5, Pop1 was comprised of four rice genotypes out of which two were local landraces (Magoti, Runingu), and two (Jasmine and IR127229) were obtained from IRRI-Burundi. Pop2 was made up of 24 genotypes including 13 from AfricaRice, six from IRRI-Burundi and five from INERA-DRC. Pop3 was a group of 21 rice genotypes among which 10 were from AfricaRice, five from IRRI-Burundi, five from INERA-DRC and one from IRRI-Kenya. Pop4 (n = 37) was formed by genotypes from all the five sources though mainly comprised of

AfricaRice genotypes (19) and IRRI-Burundi genotypes (12). Additionally, three genotypes were the local landraces, two from INERA-DRC and one from IRRI-Kenya. Pop5 was composed of eight genotypes out of which seven were IRRI-Kenya genotypes and one AfricaRice genotype.

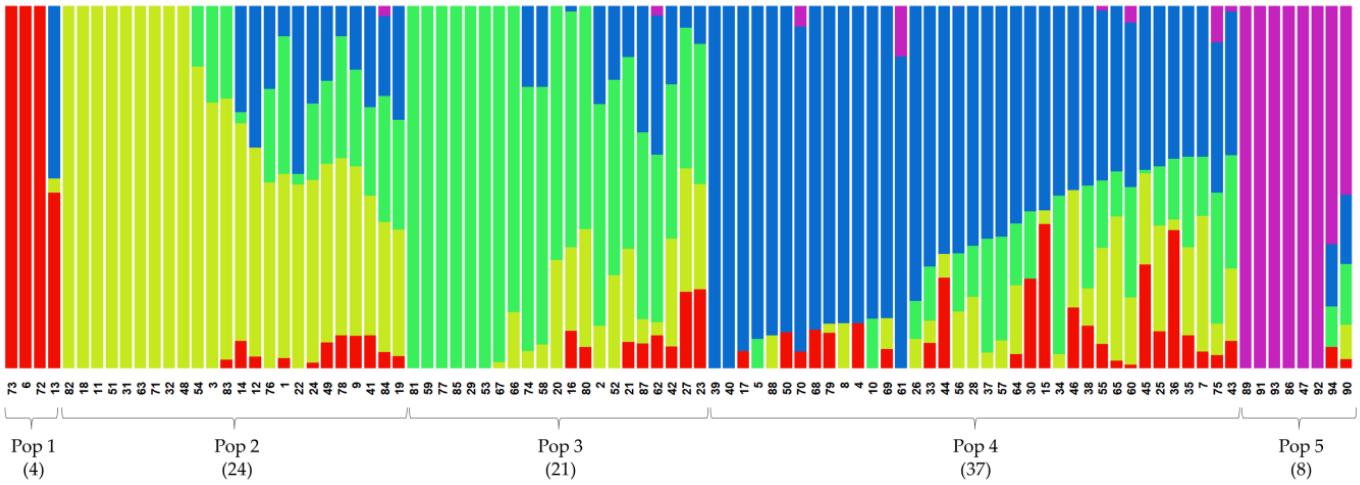

Figure 2. Population structure of 94 rice genotypes based on variation in SNP DArTseq markers with K = 5. Each genotype is represented by a vertical bar that is segmented into K colors, indicating the likelihood of membership to each cluster.

PCA was used to further explore genetic relationships among the 94 rice genotypes (Figure 3). Based on the SNP markers, the first three axes of the PCA explained 36.3% of the total genetic variation, resulting in the grouping of the 94 genotypes into three clusters.

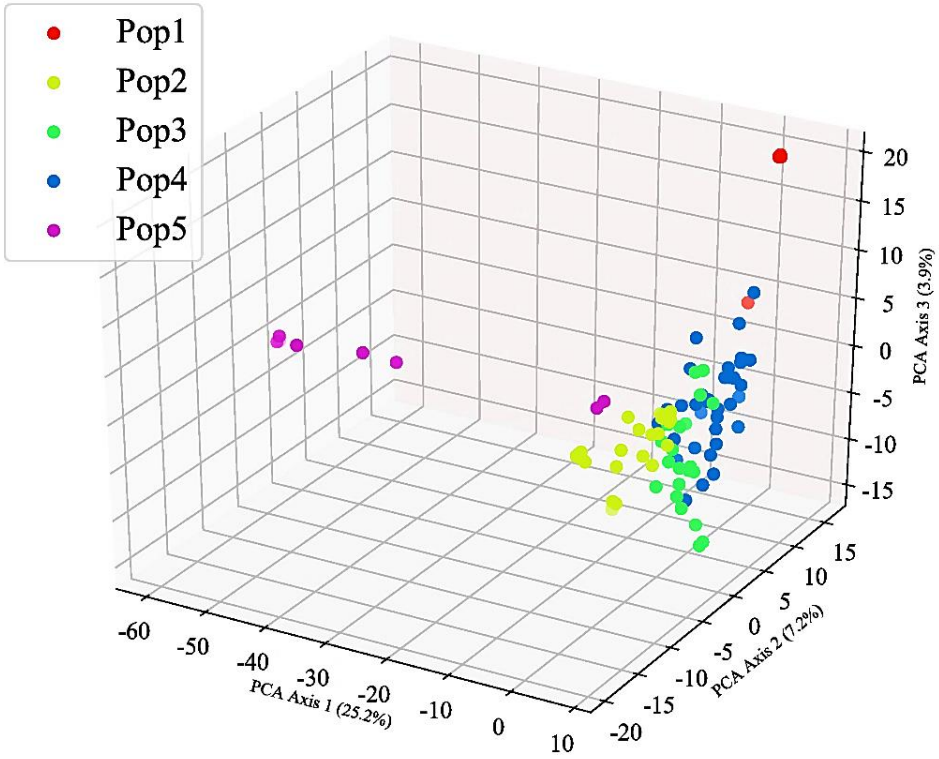

Figure 3. PCA illustrating relatedness and distribution of the 94 rice genotypes based on 8389 SNP markers. Each color corresponds to a specific sub-population from the ADMIXTURE results.

The first cluster consisted of two landraces and one genotype from IRRI-Burundi; the second cluster gathered together most of the genotypes from AfricaRice, IRRI-Burundi and INERA-DRC, some from IRRI-Kenya and landraces. The third cluster was mainly composed of genotypes from IRRI-Kenya.

3.3. Genetic Diversity and Phylogenetic Analysis

The analysis of molecular variance (AMOVA) conducted on the 94 rice genotypes revealed highly significant genetic differences ($p < 0.001$) between sub-populations and within genotypes. However, no significant difference ($p > 0.001$) was observed between genotypes within sub-populations, as shown in Table 3. Among the total genetic variations observed in the 94 rice genotypes, 36.09% was attributed to genetic differentiation between the sub-populations, 34.04% to genetic differentiation within the genotypes, and the remaining proportion (29.87%) was due to genetic differences between genotypes within the sub-populations. Additionally, the overall fixation index (F_{st}) and number of migrants (Nm) among the sub-populations were 0.52 and 0.23, respectively (Table 3).

Table 3. Analysis of molecular variance among 94 genotypes based on the SNPs.

Source	DF	MS	Estimated Variance	Proportion of Variation (%)	F_{st}	Nm
Between sub-populations	4	32,575.18	892.93	36.09 (<0.001)	0.52	0.23
Between genotypes within sub-population	89	2319.99	738.89	29.87 (>0.001)		
Within genotypes	94	842.20	842.2	34.04 (<0.001)		
Total	187	2224.31	2474.03	100		

DF: degree of freedom; MS: mean of squares; F_{st}: fixation index; Nm: number of migrants.

The percentage of polymorphic loci per population (PPL) ranged from 21.28% (Pop1) to 85.64% (Pop4), with an average of 54.23% (Table 4). The mean values for the expected heterozygosity (H_e), observed heterozygosity (H_o), unbiased expected heterozygosity (uH_e), and inbreeding coefficient (F_{is}) in the overall population were 0.14, 0.08, 0.15, and 0.52, respectively. The gene diversity values, calculated as expected heterozygosity (H_e), varied from 0.08 (Pop 1) to 0.21 (Pop 4). The observed heterozygosity (H_o) ranged from 0.02 (Pop5) to 0.13 (Pop4), while the inbreeding coefficient (F_{is}) varied from 0.31 (Pop2) to 0.75 (Pop5).

Table 4. Genetic diversity indices for the five rice sub-populations based on 8389 SNP markers.

Sub-Population	No. of Sample	PPL	H_e	H_o	uH_e	F_{is}
Pop1	4	21.28	0.08	0.05	0.09	0.52
Pop2	24	69.14	0.15	0.11	0.16	0.31
Pop3	21	65.36	0.18	0.07	0.19	0.62
Pop4	37	85.64	0.21	0.13	0.22	0.38
Pop5	8	29.73	0.09	0.02	0.10	0.75
Average		54.23	0.14	0.08	0.15	0.52

PPL: percentage of polymorphic loci; H_o: observed heterozygosity; H_e: expected heterozygosity; uH_e: unbiased expected heterozygosity; F_{is}: inbreeding coefficient.

The population pairwise fixation indices, presented in Table 5, estimate genetic differentiation among populations due to genetic structure. Pop5 showed greater genetic distance from Pop1, Pop2, Pop3, and Pop4 with F_{st} values of 0.83, 0.78, 0.75, and 0.73, respectively. The minimum genetic distances were observed between Pop4 and Pop3 ($F_{st} = 0.07$), followed by Pop3 and Pop2 ($F_{st} = 0.11$), and between Pop4 and Pop2 ($F_{st} = 0.13$).

Table 5. Population's pairwise genetic differentiation index (F_{st}).

	Pop1	Pop2	Pop3	Pop4	Pop5
Pop1	0				
Pop2	0.43	0			
Pop3	0.33	0.11	0		
Pop4	0.26	0.13	0.07	0	
Pop5	0.83	0.78	0.75	0.73	0

The neighbor-joining tree (Figure 4) illustrates genetic relatedness among the five sub-populations. The analysis resulted in the formation of three distinct groups. The first group consisted of Pop2, Pop3, and Pop4. The second group comprised only Pop1, while the third group was composed of Pop5.

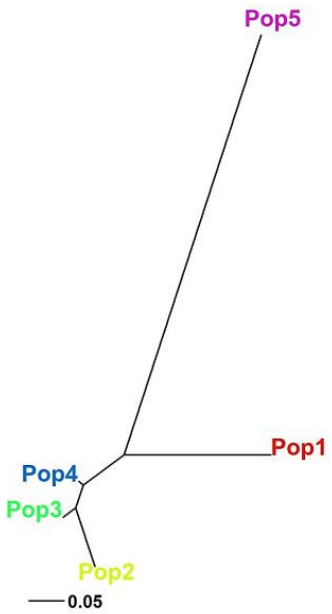

Figure 4. Neighbor-joining tree showing the genetic differentiation among the five sub-populations as revealed by ADMIXTURE analysis on 94 rice genotypes. The relationship between these sub-populations is depicted using F_{st} values as a basis of comparison.

The Euclidean genetic distance among the 94 rice genotypes, based on SNP markers, ranged from 0.00 to 1.60, with an average of 0.87 (Table S1). The lowest genetic distances were observed in Pop5 between NERICA2 and NERICA10 (0.13), in Pop1 between Magoti and Runingu (0.15), Magoti and Jasmine (0.19), and Jasmine and Runingu (0.2). In Pop2, the lowest genetic distance was observed between ARS755-3-3-1-B and ARS168-1-B-3-B (0.18). The highest genetic distance was exhibited between NERICA-L-19-sab-1 (from AfricaRice, Pop2) and NERICA12 (from IRRI-Kenya, Pop5), between NERICA-L-19-sab-1 (from AfricaRice, Pop2) and ARICA4 (from IRRI-Kenya, Pop5), and between WAHX14N-926 (from AfricaRice, Pop3) and ARICA4 (from IRRI-Kenya, Pop5). The resulting genetic distance matrix was used to construct a neighbor-joining tree that classified the 94 rice genotypes into two distinct clusters (Figure 5).

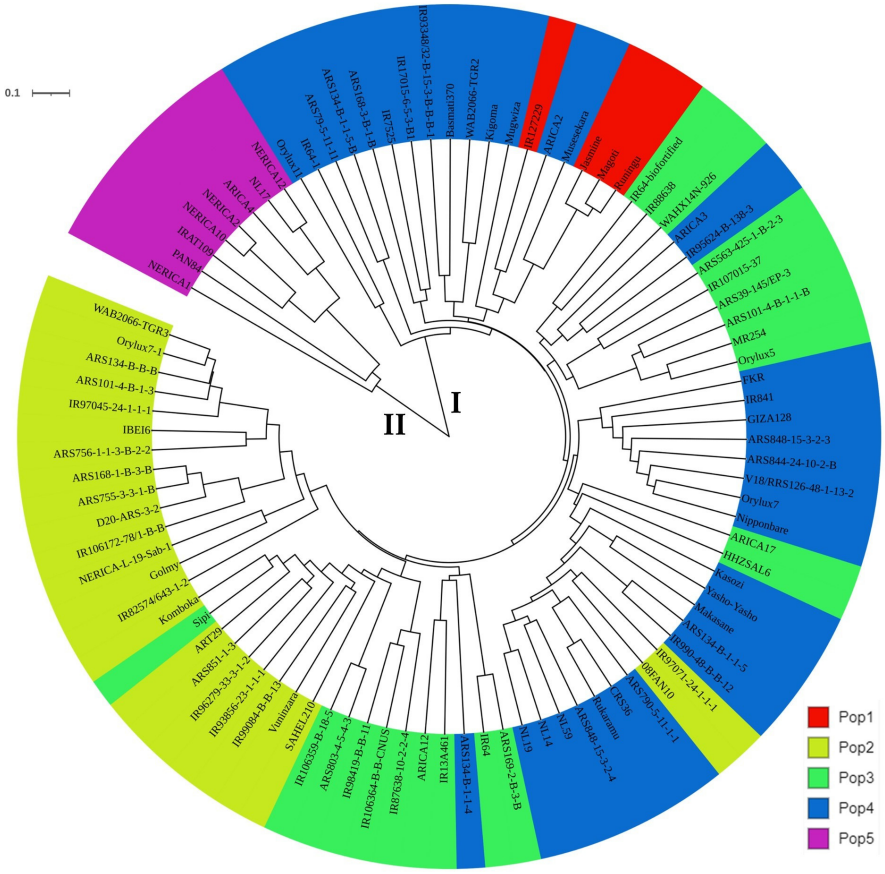

Figure 5. Phylogenetic tree showing clustering pattern among 94 rice genotypes based on SNP markers. The colors are based on the five sub-populations from ADMIXTURE results. I: cluster 1 and II: cluster 2.

4. Discussion

We studied the population structure and genetic diversity of rice genotypes maintained at the Institut National pour l'Etude et la Recherche Agronomiques (INERA) in the Democratic Republic of Congo (DRC) using SNP markers derived from DArTseq technology. In this study, a total of 8389 SNP markers passed quality control analyses designed to remove non-informative markers, as well as the markers with a minor allele frequency of less than 5% and a call rate lower than 95%. This ensured that the selected markers were suitable for estimating the genetic diversity and structure of the 94 rice genotypes under investigation. It was observed that the number of SNPs generated using DArTseq technology in rice was higher compared to previous studies by Adeboye et al. [45] and Thant et al. [46], although lower than the findings reported by Ndjiondjop et al. [47]. Analysis of SNP markers distribution across the rice chromosomes revealed an average of 699 SNP markers per chromosome, indicating a wide distribution of markers. The abundance of polymorphic markers on chromosomes is commonly associated with the level of genetic diversity [45,48].

The PIC values for a set of genetic markers are a useful tool for evaluating the informativeness of these markers in population diversity studies [49]. According to Botstein et al. [49], markers with PIC values greater than 0.5 are highly informative, those with PIC values between 0.5 and 0.25 are considered moderately informative, while markers with PIC values less than 0.25 remain to some extent informative. In the current study, 42.4% of SNPs were moderately informative, which demonstrated the usefulness of these markers for genetic diversity analysis in rice. According to Eltaher et al. [48], due to their bi-allelic nature, SNP markers have limited informative value, resulting in low to moderate

PIC values that are restricted to a maximum of 0.5. However, the average SNP PIC value of 0.25 obtained in this study was slightly higher than that found in previous studies in rice [9,45,46], cassava [50], and sorghum [51]. Our study reports a higher frequency of SNP transitions as compared to that of SNP transversions. It is commonly observed that transitions are more frequent than transversions in true SNPs [52–54]. In fact, the ratio of transition to transversion frequencies is often used as a measure of evolutionary distance between species or individuals [52]. Moreover, it happens that in a set of three available SNPs, two of them are transitions while the third is a transversion [52,55]. Similar findings have also been reported on rice [45], maize [56,57], and *Camelina sativa* [55]. According to Guo et al. [58], transversions would have a larger impact in disrupting transcription factor binding, leading to significant alterations in gene expression.

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 [59]. In the present study, 94 rice genotypes were assessed and classified into five sub-populations that displayed significant divergence among them, with varying degrees of diversity observed within each sub-population. It was observed that the composition of genotypes within each sub-population exhibited a dependence on their respective collection sources. The result highlights the source of the rice genotypes as a major factor influencing their genetic makeup, with Pop1 and Pop5 having a higher percentage of genotypes from specific sources (local landraces and IRRI-Kenya, respectively). However, Pop2, Pop3, and Pop4, predominantly made up of AfricaRice, IRRI-Burundi, and INERA-DRC genotypes, exhibited a common genetic background pointing to the fact that breeding activities have led to genetic similarities among these groups [60]. Similarities between 22 diverse rice collections from different sources were also reported by Salem [8]. Principal component analysis (PCA) was used to confirm the patterns of admixture among populations [61,62]. The analysis of rice genotypes using the PCA plot revealed that the 94 genotypes could be grouped into three clusters. The distribution of the genotypes based on PCA was similar to that of the structure analysis using ADMIXTURE. These findings were consistent with previous reports [30,47]. We noted that the first cluster comprised three genotypes, of which two were local landraces and one was from IRRI-Burundi. The second cluster was composed of genotypes from all five sources, and the third cluster was predominantly formed by IRRI-Kenya genotypes. This result revealed the presence of common alleles among the genotypes within each cluster, which could be attributed to breeding activities such as selection for specific traits and hybridization [60].

AMOVA is a commonly used statistical method for evaluating the extent to which different levels of population structure contribute to genetic variation patterns [63]. The results of our study revealed that genetic diversity among the 94 rice genotypes was largely determined by the differentiation between sub-populations and within genotypes, which may be due to low genetic exchange and limited gene flow. According to Wright [40], a number of a migrants (Nm) value lower than 1 indicates limited gene flow among subpopulations. In this study, the observed Nm value was 0.23, which suggests limited genetic exchange and significant differentiation (F_{st} = 0.52) observed among sub-populations, consistent with the AMOVA. This revealed that there is a potential for identifying and selecting diverse genotypes from different sub-populations for breeding programs [64]. The genetic differences between genotypes within sub-populations contributed the least to the total genetic variation, indicating genetic similarity between the genotypes within sub-populations, which could be due to breeding practices, such as selection for specific traits or the use of parent lines with similar genetic backgrounds. This can be useful for establishing breeding populations with specific traits of interest [65]. Our study findings corroborate those of previous research on rice genetic diversity and population structure, which also reported a significant differentiation between sub-populations and within genotypes using SSR markers [66] among genotypes from AfricaRice, IRRI and Tanzania.

According to Eltaher et al. [48] and Luo et al. [55], genetic indices serve as indicators of genetic diversity. In our study, the mean H_o value (<0.1) aligns with previous research on rice [46] (H_o = 0.03), [30] (H = 0.0975), [45] (H_o < 1), but is slightly lower than findings by Suvi et al. [66] (H_o = 0.17) using SSR markers. The lower H_o compared to H_e was expected due to rice's self-pollinating nature [67], resulting in a relatively higher degree of inbreeding within the population, as also supported by Ndjiondjop et al. [47]. Examining gene diversity based on PPL, H_o, H_e, and F_{is} values, the genotypes within sub-populations (Pop2, Pop3, and Pop4) displayed diversity, suggesting possible occurrences of selection and hybridization among the accessions within these sub-populations. This genetic exchange contributes to increased diversity within the populations [68]. In contrast, Pop5 and Pop1 exhibited lower diversity, likely influenced by factors such as rice's inherent high inbreeding nature and specific selective breeding criteria employed by breeders for Pop5, along with strict selection practices by farmers for Pop1.

The F_{st} was further used to quantify sub-population differentiation resulting from genetic structure. According to Wright [69], an F_{st} value of 0.25 or higher is considered significant in differentiating sub-populations, while values in the range of 0.15–0.25 indicate moderate differentiation. In contrast, differentiation is considered insignificant if the F_{st} value is 0.05 or less. Significant genetic differentiation was observed between all pairs of sub-populations, except for sub-populations 3 and 4, 2 and 3, 2 and 4. The lack of differentiation among sub-populations comprising genotypes from AfricaRice, IRRI-Burundi, and INERA-DRC may be due to the exchange of genetic materials, the effect of maintenance strategies, and the selection criteria applied by breeders across the involved institutes [48,55].

In this study, the average Euclidean genetic distance, a measure of genetic variation between pairs of genotypes, was found to be 0.87. This result is consistent with previous reports on elite rice genotypes from Chile [70] and Ugandan rice genotypes [30], which reported genetic distances of 0.87 and 0.86, respectively. The phylogenetic tree illustrates the distances between genotypes or groups, indicating their degree of relationship, with closely related groups positioned close to each other [71]. The NJ tree classified the 94 genotypes into two major groups, revealing a shared gene pool within each cluster. Cluster I comprised genotypes from AfricaRice, IRRI-Burundi, INERA-DRC, and the local landraces, while cluster II mainly consisted of IRRI-Kenya genotypes. Low genetic distances observed in pairs of genotypes such as Magoti and Runingu, Magoti and Jasmine, Jasmine and Runingu, ARS755-3-3-1-B and ARS168-1-B-3-B, NERICA2 and NERICA10, suggest that these pairs were potentially collected separately but share a close genetic background. The high genetic distance observed between genotypes of different sub-populations may be attributed to the distinct genetic makeup of the IRRI-Kenya rice and local landrace collections, indicating differences from other sources, which aligns with the AMOVA results.

5. Conclusions

The DArTseq platform was utilized to generate high-density markers across the rice genome, enabling the assessment of population structure and genetic diversity among 94 rice germplasm accessions in the eastern DRC. A total of 8389 SNP markers, exhibiting high polymorphism and moderate informativeness, were employed. The structure analysis revealed that the rice panel consisted of five sub-populations, including admixtures. Phylogenetic and principal component analysis further categorized the sub-populations into two and three clusters, respectively, primarily based on their sources. Notably, significant genetic variation was observed both among populations and within genotypes, useful for rice improvement. The observed limited diversity among genotypes within populations can be advantageous for establishing breeding populations focused on specific traits of interest. Moreover, the introduction of novel alleles from divergent sources and the utilization of advanced genomic techniques can contribute to the creation of a more diverse population. This study provides valuable insights into structure and genetic diversity, which can be harnessed for rice breeding programs, including exploring the genetic basis of desired traits and the identification of potential parental lines for development of new rice varieties.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy13071906/s1, Table S1: Euclidean genetic distances among the 94 rice genotypes based on 8389 DArTseq SNP markers.

Author Contributions: Each author contributed substantially to this work. Conceptualization, E.G.M., W.K., P.K.K., S.M.G., C.B.B., R.M.C. and M.F.; methodology, E.G.M., W.K., P.K.K., C.B.B., S.M.G., R.M.C. and M.F.; software, P.K.K., W.K. and C.B.B.; validation, E.G.M., M.F., S.M.G., R.M.C. and J.M.; formal analysis, P.K.K., W.K. and C.B.B.; investigation, P.K.K. and C.B.B.; resources, P.K.K. and C.B.B.; data curation, P.K.K. and C.B.B.; writing—original draft preparation, P.K.K. and C.B.B.; visualization, P.K.K., W.K. and C.B.B.; writing—review and editing, P.K.K., C.B.B., E.G.M., W.K., S.M.G., J.M., R.M.C. and M.F.; supervision, E.G.M., S.M.G., R.M.C., J.M. and M.F.; project administration, J.M. and M.F.; funding acquisition, J.M. and M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the International Institute of Tropical Agriculture (IITA) and the Institut National pour l'Etude et la Recherche Agronomiques [National Institute for Agronomic Research and Study] (INERA-DR Congo), through the AFR RI-Regional Great Lakes Integrated Agriculture Development Project (PICAGL), ZR-MINAGRI-261053-CS-CDS, grant number P143307, funded by the World Bank Group.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank the training manager of IITA under the PICAGL project (2019–2022) for administrative support. We are also grateful to the INERA staff, especially the rice breeding program for their help in providing the collection of rice genotypes.

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

References

1. Muthayya, S.; Sugimoto, J.D.; Montgomery, S.; Maberly, G.F. An Overview of Global Rice Production, Supply, Trade, and Consumption: Global Rice Production, Consumption, and Trade. *Ann. N. Y. Acad. Sci.* **2014**, *1324*, 7–14. [CrossRef] [PubMed]
2. Food and Agriculture Organization of the United Nations; World Food Programme; International Fund for Agricultural Development. *Food Loss Analysis: Causes and Solutions—Case Studies on Maize and Rice in the Democratic Republic of Congo*; International Fund for Agricultural Development: Rome, Italy, 2019; p. 188.
3. Campos, H.; Caligari, P.D. *Genetic Improvement of Tropical Crops*; Springer: Berlin/Heidelberg, Germany, 2017.
4. Hour, A.; Hsieh, W.; Chang, S.; Wu, Y.; Chin, H.; Lin, Y. Genetic Diversity of Landraces and Improved Varieties of Rice (*Oryza sativa* L.) in Taiwan. *Rice* **2020**, *13*, 82. [CrossRef]
5. Wang, L.; Hao, L.; Li, X.; Hu, S.; Ge, S.; Yu, J. SNP Deserts of Asian Cultivated Rice: Genomic Regions under Domestication. *J. Evol. Biol.* **2009**, *22*, 751–761. [CrossRef]
6. Roy, S.; Marndi, B.C.; Mawkhlieng, B.; Banerjee, A.; Yadav, R.M.; Misra, A.K.; Bansal, K.C. Genetic Diversity and Structure in Hill Rice (*Oryza sativa* L.) Landraces from the North-Eastern Himalayas of India. *BMC Genet.* **2016**, *17*, 107. [CrossRef]
7. Mbanjo, E.G.N.; Jones, H.; Caguiat, X.G.I.; Carandang, S.; Ignacio, J.C.; Ferrer, M.C.; Boyd, L.A.; Kretzschmar, T. Exploring the Genetic Diversity within Traditional Philippine Pigmented Rice. *Rice* **2019**, *12*, 27. [CrossRef] [PubMed]
8. Salem, K.F.M.; Sallam, A. Analysis of Population Structure and Genetic Diversity of Egyptian and Exotic Rice (*Oryza sativa* L.) Genotypes. *Comptes Rendus Biol.* **2016**, *339*, 1–9. [CrossRef] [PubMed]
9. Singh, N.; Choudhury, D.R.; Singh, A.K.; Kumar, S.; Srinivasan, K.; Tyagi, R.K.; Singh, N.K.; Singh, R. Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. *PLoS ONE* **2013**, *8*, e84136. [CrossRef]
10. Liu, W.; Shahid, M.Q.; Bai, L.; Lu, Z.; Chen, Y.; Jiang, L.; Diao, M.; Liu, X.; Lu, Y. Evaluation of Genetic Diversity and Development of a Core Collection of Wild Rice (*Oryza rufipogon* Griff.) Populations in China. *PLoS ONE* **2015**, *10*, e0145990. [CrossRef]
11. Aesomnuk, W.; Ruengphayak, S.; Ruanjaichon, V.; Sreewongchai, T.; Malumpong, C.; Vanavichit, A.; Toojinda, T.; Wanchana, S.; Arikit, S. Estimation of the Genetic Diversity and Population Structure of Thailand's Rice Landraces Using SNP Markers. *Agronomy* **2021**, *11*, 995. [CrossRef]
12. Joshi, B.K.; Ghimire, K.H.; Neupane, S.P.; Gauchan, D.; Mengistu, D.K. Approaches and Advantages of Increased Crop Genetic Diversity in the Fields. *Diversity* **2023**, *15*, 603. [CrossRef]
13. Gu, D.; Andreev, K.; Dupre, M.E. Major Trends in Population Growth Around the World. *China CDC Wkly.* **2021**, *3*, 604–613. [CrossRef]
14. Seck, P.A.; Touré, A.A.; Coulibaly, J.Y.; Diagne, A.; Wopereis, M.C.S. Africa's Rice Economy before and after the 2008 Rice Crisis. In *Realizing Africa's Rice Promise*; Wopereis, M.C.S., Johnson, D.E., Ahmadi, N., Tollens, E., Jalloh, A., Eds.; CABI: Wallingford, UK, 2013; pp. 24–34. [CrossRef]

15. Balasubramanian, V.; Sie, M.; Hijmans, R.J.; Otsuka, K. Increasing Rice Production in Sub-Saharan Africa: Challenges and Opportunities. In *Advances in Agronomy*; Elsevier: Amsterdam, The Netherlands, 2007; Volume 94, pp. 55–133. [CrossRef]
16. Iqbal, T.; Hussain, I.; Ahmad, N.; Nauman, M.; Ali, M.; Saeed, S.; Zia, M.; Ali, F. Genetic variability, correlation and cluster analysis in elite lines of rice. *J. Sci. Agric.* **2018**, *2*, 85. [CrossRef]
17. Swarup, S.; Cargill, E.J.; Crosby, K.; Flagel, L.; Kniskern, J.; Glenn, K.C. Genetic Diversity Is Indispensable for Plant Breeding to Improve Crops. *Crop Sci.* **2021**, *61*, 839–852. [CrossRef]
18. Khan, M.H.; Dar, Z.A.; Dar, S.A. Breeding Strategies for Improving Rice Yield—A Review. *Agric. Sci.* **2015**, *06*, 467–478. [CrossRef]
19. Novoselović, D.; Bentley, A.R.; Šimek, R.; Dvojković, K.; Sorrells, M.E.; Gosman, N.; Horsnell, R.; Drezner, G.; Šatović, Z. Characterizing Croatian Wheat Germplasm Diversity and Structure in a European Context by DArT Markers. *Front. Plant Sci.* **2016**, *7*, 184. [CrossRef]
20. Franco, J.; Crossa, J.; Ribaut, J.M.; Bertran, J.; Warburton, M.L.; Khairallah, M. A Method for Combining Molecular Markers and Phenotypic Attributes for Classifying Plant Genotypes. *Theor. Appl. Genet.* **2001**, *103*, 944–952. [CrossRef]
21. Ullah, F.; Gao, Y.; Sari, İ.; Jiao, R.-F.; Saqib, S.; Gao, X.-F. Macro-Morphological and Ecological Variation in Rosa Sericea Complex. *Agronomy* **2022**, *12*, 1078. [CrossRef]
22. Govindaraj, M.; Vetriventhan, M.; Srinivasan, M. Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives. *Genet. Res. Int.* **2015**, *2015*, 431487. [CrossRef] [PubMed]
23. Nadeem, M.A.; Nawaz, M.A.; Shahid, M.Q.; Doğan, Y.; Comertpay, G.; Yıldız, M.; Hatipoğlu, R.; Ahmad, F.; Alsaleh, A.; Labhane, N.; et al. DNA Molecular Markers in Plant Breeding: Current Status and Recent Advancements in Genomic Selection and Genome Editing. *Biotechnol. Biotechnol. Equip.* **2018**, *32*, 261–285. [CrossRef]
24. Deres, D.; Feyissa, T. Concepts and Applications of Diversity Array Technology (DArT) Markers for Crop Improvement. *J. Crop Improv.* **2022**, 1–21. [CrossRef]
25. Wenzl, P.; Carling, J.; Kudrna, D.; Jaccoud, D.; Huttner, E.; Kleinhofs, A.; Kilian, A. Diversity Arrays Technology (DArT) for Whole-Genome Profiling of Barley. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9915–9920. [CrossRef]
26. Kilian, A.; Wenzl, P.; Huttner, E.; Carling, J.; Xia, L.; Blois, H.; Caig, V.; Heller-Uszynska, K.; Jaccoud, D.; Hopper, C.; et al. Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In *Data Production and Analysis in Population Genomics*; Pompanon, F., Bonin, A., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2012; Volume 888, pp. 67–89. [CrossRef]
27. Adu, B.G.; Akromah, R.; Amoah, S.; Nyadanu, D.; Yeboah, A.; Aboagye, L.M.; Amoah, R.A.; Owusu, E.G. High-Density DArT-Based SilicoDArT and SNP Markers for Genetic Diversity and Population Structure Studies in Cassava (*Manihot esculenta* Crantz). *PLoS ONE* **2021**, *16*, e0255290. [CrossRef]
28. Fufa, T.W.; Abtew, W.G.; Amadi, C.O.; Oselebe, H.O. DArTSeq SNP-Based Genetic Diversity and Population Structure Studies among Taro [(*Colocasia esculenta* (L.) Schott] Accessions Sourced from Nigeria and Vanuatu. *PLoS ONE* **2022**, *17*, e0269302. [CrossRef] [PubMed]
29. Mogga, M.; Sibiya, J.; Shimelis, H.; Lamo, J.; Yao, N. Diversity Analysis and Genome-Wide Association Studies of Grain Shape and Eating Quality Traits in Rice (*Oryza sativa* L.) Using DArT Markers. *PLoS ONE* **2018**, *13*, e0198012. [CrossRef]
30. Niedziela, A.; Bednarek, P.T. Population Structure and Genetic Diversity of a Germplasm for Hybrid Breeding in Rye (*Secale cereale* L.) Using High-Density DArTseq-Based SilicoDArT and SNP Markers. *J. Appl. Genet.* **2023**, *64*, 217–229. [CrossRef]
31. Mudaki, P.; Wamalwa, L.N.; Muui, C.W.; Nzuve, F.; Muasya, R.M.; Nguluu, S.; Kimani, W. Genetic Diversity and Population Structure of Sorghum (*Sorghum bicolor* (L.) Moench) Landraces Using DArTseq-Derived Single-Nucleotide Polymorphism (SNP) Markers. *J. Mol. Evol.* **2023**, *91*, 552–561. [CrossRef] [PubMed]
32. Sansaloni, C.; Petroli, C.; Jaccoud, D.; Carling, J.; Detering, F.; Grattapaglia, D.; Kilian, A. Diversity Arrays Technology (DArT) and next-Generation Sequencing Combined: Genome-Wide, High Throughput, Highly Informative Genotyping for Molecular Breeding of Eucalyptus. *BMC Proc.* **2011**, *5*, P54. [CrossRef]
33. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S.; et al. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [CrossRef]
34. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The Variant Call Format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
35. Gruber, B.; Unmack, P.J.; Berry, O.F.; Georges, A. DartR: An R Package to Facilitate Analysis of SNP Data Generated from Reduced Representation Genome Sequencing. *Mol. Ecol. Resour.* **2018**, *18*, 691–699. [CrossRef]
36. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE Algorithm for Individual Ancestry Estimation. *BMC Bioinform.* **2011**, *12*, 246. [CrossRef]
37. Jombart, T.; Ahmed, I. Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data. *Bioinformatics* **2011**, *27*, 3070–3071. [CrossRef] [PubMed]
38. Nei, M. Genetic Distance between Populations. *Am. Nat.* **1972**, *106*, 283–292. [CrossRef]
39. Kamvar, Z.N.; Tabima, J.F.; Grünwald, N.J. Poppr: An R Package for Genetic Analysis of Populations with Clonal, Partially Clonal, and/or Sexual Reproduction. *PeerJ* **2014**, *2*, e281. [CrossRef]
40. Wright, S. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution* **1965**, *19*, 395–420. [CrossRef]

41. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, NY, USA, 1987. [CrossRef]
42. Goudet, J. Hierfstat, a Package for r to Compute and Test Hierarchical F-Statistics. *Mol. Ecol. Notes* **2005**, *5*, 184–186. [CrossRef]
43. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R Language. *Bioinformatics* **2004**, *20*, 289–290. [CrossRef] [PubMed]
44. Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]
45. Adeboye, K.A.; Oyedeji, O.E.; Alqudah, A.M.; Börner, A.; Oduwaye, O.; Adebambo, O.; Daniel, I.O. Genetic Structure and Diversity of Upland Rice Germplasm Using Diversity Array Technology (DArT)-Based Single Nucleotide Polymorphism (SNP) Markers. *Plant Genet. Resour.* **2020**, *18*, 343–350. [CrossRef]
46. Thant, A.A.; Zaw, H.; Kalousova, M.; Singh, R.K.; Lojka, B. Genetic Diversity and Population Structure of Myanmar Rice (*Oryza sativa* L.) Varieties Using DArTseq-Based SNP and SilicoDArT Markers. *Plants* **2021**, *10*, 2564. [CrossRef]
47. Ndjiondjop, M.N.; Semagn, K.; Zhang, J.; Gouda, A.C.; Kpeki, S.B.; Goungoulou, A.; Wambugu, P.; Dramé, K.N.; Bimpong, I.K.; Zhao, D. Development of Species Diagnostic SNP Markers for Quality Control Genotyping in Four Rice (*Oryza* L.) Species. *Mol. Breed.* **2018**, *38*, 131. [CrossRef] [PubMed]
48. Eltaher, S.; Sallam, A.; Belamkar, V.; Emara, H.A.; Nower, A.A.; Salem, K.F.M.; Poland, J.; Baenziger, P.S. Genetic Diversity and Population Structure of F3:6 Nebraska Winter Wheat Genotypes Using Genotyping-By-Sequencing. *Front. Genet.* **2018**, *9*, 76. [CrossRef]
49. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331.
50. Kawuki, R.S.; Herselman, L.; Labuschagne, M.T.; Nzuki, I.; Ralimanana, I.; Bidiaka, M.; Kanyange, M.C.; Gashaka, G.; Masumba, E.; Mkamilo, G.; et al. Genetic Diversity of Cassava (*Manihot Esculenta* Crantz) Landraces and Cultivars from Southern, Eastern and Central Africa. *Plant Genet. Resour.* **2013**, *11*, 170–181. [CrossRef]
51. Afolayan, G.; Deshpande, S.P.; Aladele, S.E.; Kolawole, A.O.; Angarawai, I.; Nwosu, D.J.; Michael, C.; Blay, E.T.; Danquah, E.Y. Genetic Diversity Assessment of Sorghum (*Sorghum bicolor* (L.) Moench) Accessions Using Single Nucleotide Polymorphism Markers. *Plant Genet. Resour.* **2019**, *17*, 412–420. [CrossRef]
52. Batley, J.; Edwards, D. SNP Applications in Plants. In *Association Mapping in Plants*; Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E., De Silva, H.N., Eds.; Springer: New York, NY, USA, 2007; pp. 95–102. [CrossRef]
53. Lai, K.; Duran, C.; Berkman, P.J.; Lorenc, M.T.; Stiller, J.; Manoli, S.; Hayden, M.J.; Forrest, K.L.; Fleury, D.; Baumann, U.; et al. Single Nucleotide Polymorphism Discovery from Wheat Next-Generation Sequence Data: Wheat AutoSNPdb. *Plant Biotechnol. J.* **2012**, *10*, 743–749. [CrossRef]
54. Boussaha, M.; Guyomard, R.; Cabau, C.; Esquerré, D.; Quillet, E. Development and Characterisation of an Expressed Sequence Tags (EST)-Derived Single Nucleotide Polymorphisms (SNPs) Resource in Rainbow Trout. *BMC Genom.* **2012**, *13*, 238. [CrossRef]
55. Luo, Z.; Brock, J.; Dyer, J.M.; Kutchan, T.; Schachtman, D.; Augustin, M.; Ge, Y.; Fahlgren, N.; Abdel-Haleem, H. Genetic Diversity and Population Structure of a Camelina Sativa Spring Panel. *Front. Plant Sci.* **2019**, *10*, 184. [CrossRef] [PubMed]
56. Kumar, B.; Rakshit, S.; Kumar, S.; Singh, B.K.; Lahkar, C.; Jha, A.K.; Kumar, K.; Kumar, P.; Choudhary, M.; Singh, S.B.; et al. Genetic Diversity, Population Structure and Linkage Disequilibrium Analyses in Tropical Maize Using Genotyping by Sequencing. *Plants* **2022**, *11*, 799. [CrossRef]
57. Morton, B.R.; Bi, I.V.; McMullen, M.D.; Gaut, B.S. Variation in Mutation Dynamics Across the Maize Genome as a Function of Regional and Flanking Base Composition. *Genetics* **2006**, *172*, 569–577. [CrossRef]
58. Guo, C.; McDowell, I.C.; Nodzenski, M.; Scholtens, D.M.; Allen, A.S.; Lowe, W.L.; Reddy, T.E. Transversions Have Larger Regulatory Effects than Transitions. *BMC Genom.* **2017**, *18*, 394. [CrossRef] [PubMed]
59. Schierenbeck, K.A. Population-Level Genetic Variation and Climate Change in a Biodiversity Hotspot. *Ann. Bot.* **2017**, *119*, 215–228. [CrossRef] [PubMed]
60. Xu, Q.; Yuan, X.; Wang, S.; Feng, Y.; Yu, H.; Wang, Y.; Yang, Y.; Wei, X.; Li, X. The Genetic Diversity and Structure of Indica Rice in China as Detected by Single Nucleotide Polymorphism Analysis. *BMC Genet.* **2016**, *17*, 53. [CrossRef]
61. Ma, J.; Amos, C.I. Principal Components Analysis of Population Admixture. *PLoS ONE* **2012**, *7*, e40115. [CrossRef]
62. Liu, C.-C.; Shringarpure, S.; Lange, K.; Novembre, J. Exploring Population Structure with Admixture Models and Principal Component Analysis. *Methods Mol. Biol.* **2020**, *2090*, 67–86. [CrossRef] [PubMed]
63. Fitzpatrick, B.M. Power and Sample Size for Nested Analysis of Molecular Variance. *Mol. Ecol.* **2009**, *18*, 3961–3966. [CrossRef]
64. Salgotra, R.K.; Chauhan, B.S. Genetic Diversity, Conservation, and Utilization of Plant Genetic Resources. *Genes* **2023**, *14*, 174. [CrossRef]
65. Li, G.; Zhang, J.; Yang, C.; Song, Y.; Zheng, C.; Wang, S.; Liu, Z.; Ding, Y. Optimal Yield-Related Attributes of Irrigated Rice for High Yield Potential Based on Path Analysis and Stability Analysis. *Crop J.* **2014**, *2*, 235–243. [CrossRef]
66. Suvi, W.T.; Shimelis, H.; Laing, M.; Mathew, I.; Shayanowako, A.I.T. Variation among Tanzania Rice Germplasm Collections Based on Agronomic Traits and Resistance to Rice Yellow Mottle Virus. *Agronomy* **2021**, *11*, 391. [CrossRef]
67. Sleper, D.A.; Poehlman, J.M. *Breeding Field Crops*, 5th ed.; Blackwell Pub: Ames, IA, USA, 2006.
68. Goulet, B.E.; Roda, F.; Hopkins, R. Hybridization in Plants: Old Ideas, New Techniques. *Plant Physiol.* **2017**, *173*, 65–78. [CrossRef]
69. Wright, S. *Evolution and the Genetics of Populations: A Treatise*; University of Chicago Press: Chicago, IL, USA, 1968.

70. Becerra, V.; Paredes, M.; Gutiérrez, E.; Rojo, C. Genetic Diversity, Identification, and Certification of Chilean Rice Varieties Using Molecular Markers. *Chil. J. Agric. Res.* **2015**, *75*, 267–274. [CrossRef]
71. Abaza, S. What Is and Why Do We Have to Know the Phylogenetic Tree? *Parasitol. United J.* **2020**, *13*, 68–71. [CrossRef]