CrossMark

Analysis of population structure and genetic diversity reveals gene flow and geographic patterns in cultivated rice (*O. sativa* and *O. glaberrima*) in West Africa

Octaviano Igor Yelome · Kris Audenaert · Sofie Landschoot · Alexandre Dansi · Wouter Vanhove · Drissa Silue · Patrick Van Damme · Geert Haesaert

Abstract To fully exploit the diversity in African rice germplasm and to broaden the gene pool reliable information on the population genetic diversity and phenotypic characteristics is a prerequisite. In this paper, the population structure and genetic diversity of 42 cultivated African rice (*Oryza* spp.) accessions originating from West Africa (Benin, Mali and Nigeria, Liberia etc.) were investigated using 20 simple sequence repeats (SSR) and 77 amplified fragment length polymorphisms (AFLP). Additionally, field trials were set up to gain insight into phenotypic characteristics that differentiate the genetic populations among rice accessions. The analysis revealed considerably high polymorphisms for SSR markers (PIC mean = 0.78) in the germplasm studied. A significant association was found between AFLP markers and geographic origin of rice accessions (R = 0.72). Germplasm structure showed that *Oryza sativa* accessions were not totally isolated from *Oryza glaberrima* accessions. The results allowed identification of five *O. glaberrima* accessions which grouped together with *O. sativa* accessions, sharing common alleles of 18 loci out of the 20 SSR markers analyzed. Population structure analysis revealed existence of a gene flow between *O. sativa* and *O. glaberrima* rice accessions which can be used to combine several interesting traits in breeding programs. Further studies are needed to clarify the contributions of this gene flow to valuable traits such

O. I. Yelome · K. Audenaert · S. Landschoot (✉) ·
W. Vanhove · P. Van Damme · G. Haesaert
Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
e-mail: igoryelome@gmail.com

S. Landschoot
e-mail: sofie.landschoot@ugent.be

A. Dansi
Faculté des Sciences et Techniques de Dassa, Laboratoire de Biotechnologie, Ressources Génétiques et Amélioration des Espèces Animales et Végétales (BIORAVE), Université d'Abomey, Abomey-Calavi, Benin

D. Silue
AfricaRice Center, 01, B.P.O. 2031, Cotonou, Benin

Present Address:
D. Silue
AfricaRice Center, 01, B.P.O. 2551, Bouake 01, Côte d'Ivoire

P. Van Damme
Faculty of Tropical AgriSciences, Czech University of Life Sciences Prague, Kamycka 129, 165 21 Prague 6 – Suchdol, Czech Republic

as abiotic and biotic stresses including disease resistance.

Keywords Amplified fragment length polymorphism · Genetic diversity · Gene flow · Rice · Simple sequence repeats

Introduction

Rice (*Oryza* spp.) is one of the world's most important self-pollinating diploid food crops that provides a significant part of staple food for millions of people (Chang 1976; Kubo and Purevdoj 2004). The genus *Oryza* is composed of two domesticated (*Oryza sativa* L. and *Oryza glaberrima* Steud.) and 22 wild species representing ten rice genome types (Vaughan 1994; Ge et al. 2001; Ammiraju et al. 2010; USDA-ARS 2013). Six out of the ten known genome types are present on the African continent and show that Africa has a very rich gene pool adapted to the African growing conditions (Sanchez et al. 2013; Wambugu et al. 2013). Both domesticated species (*O. glaberrima* and *O. sativa*) are grown in West Africa (Nuijten et al. 2009) with *O. glaberrima* being native from Africa (Portères 1956; Angladette 1966). Many *O. glaberrima* and *O. sativa* cultivars have been collected (over 20,000 rice accessions) and are conserved in the AfricaRice Genebank, a Consultative Group for International Agricultural Research (CGIAR) (Sié et al. 2012). However, because of the lack of knowledge on their genetic characteristics, they are still limitedly used for breeding and sustainable production (Roy Choudhury et al. 2014). Particularly *O. glaberrima* is underutilized since farmers replace it with *O. sativa* because of the higher grain yield of the latter species (Linares 2002; National Research Council 1996). Teeken et al. (2012) reported that the majority of *O. glaberrima* cultivars are entirely the farmers' outcome from maintenance of diversity since only limited research has been conducted to improve this species. In contrast, more attention was paid to *O. sativa* species. Given the known *O. glaberrima* potential to adapt to a wide range of adverse environments (Pham 1992; Jones et al. 1997; Futakuchi and Sié 2009; Sié et al. 2010; Mokuwa et al. 2013), strategies for a better valorization of its cultivars must be encouraged. Recently, we have initiated the

screening of a number of rice accessions (345 *O. glaberrima* and five *O. sativa* accessions) for their resistance to blast in upland conditions in Benin (Yelome et al. 2018). The relationship between genetic diversity based on AFLP markers and disease tolerance was evidenced. The present work was undertaken to further examine genetic structure diversity and relationships among the African rice germplasm. There is limited knowledge on these African rice germplasm resources as few studies describing the diversity based on both molecular (of genetic) and phenotypic characterization have been published. The characterization of genetic diversity can serve as a basis to formulate strategies to expand the diversity among cultivars and to broaden the genepool by making thought-out crosses (Nisar et al. 2008, Salem and Sallam 2016). Furthermore, both the relationship between the genetic structure and geographic distribution and the level of gene flow among rice accessions will be examined in the present study for efficient germplasm management. Knowledge on the genetic structure is very important for breeding and conservation. It can help by guiding the choice of genetically distant genotypes to incorporate a higher variation into segregating populations (Sié et al. 2012). Gene flow investigation may also serve to introduce potentially adaptive alleles to populations, and increase genetic diversity, which natural selection can act upon to provide an evolutionary response (North et al. 2011; Sexton et al. 2011). Molecular SSRs and AFLPs markers (Vos et al. 1995; Vieira et al. 2016) have been widely used to screen, characterize and evaluate genetic diversity in rice. A number of key distinctive traits will also be investigated to provide a better understanding of genetic diversity structure, which will facilitate an efficient management and valorization of assembled African rice germplasm.

The objectives of this study are to (1) assess the genetic diversity and phylogenetic relationship of a selection of African rice accessions originating from Benin, Mali, Nigeria, Ivory Coast, Liberia and Guinea; (2) examine the relationship between genetic diversity and geographic origin of rice accessions; (3) determine the genetic structure and the level of gene flow in the germplasm. This information is important in searching strategies for the selection of high-yielding accessions with resistance to blast disease.

Materials and methods

The germplasm of *Oryza* spp. included in this study is a subset of 42 rice accessions (five *O. sativa* accessions and 37 *O. glaberrima* accessions) originating from six West African countries (Suppl. Table 1). This representative subset has been derived from the entire African collection of 350 rice accessions based on geographical origins, pairwise genetic distance (revealed by 77 AFLP markers) and differential reactions of the accessions to blast disease. The average Polymorphic Information Content (PIC) was $24 \pm 4\%$ and Jaccard's genetic dissimilarity coefficients ranged from 0.03 to 0.98 for the collection of 350 accessions. (Yelome et al. 2018). Among the selected germplasm, we have identified several field blast resistance patterns: 26 highly resistant accessions; 9 moderately resistant, 3 moderately susceptible and 4 susceptible.

DNA extraction

Genomic DNA extraction from each accession was done according to the CetylTrimethyl Ammonium Bromide (CTAB) method described by Saghai-Maroof et al. (1984). Five seeds from each of the 42 rice samples were grown in a growth chamber at 27 °C. Approximately 200 mg of leaf samples were harvested on 21 days-old seedlings and ground in liquid nitrogen in 2 mL of CTAB using a pestle and mortar. Then, 500 μL of each of the obtained suspensions were transferred to centrifuge tubes and incubated in a water bath for 30 min at 60 °C. An equal volume of chloroform:isoamyl alcohol (24:1, v/v) was added to each tube. The suspension was gently mixed by inverting the tubes. Tubes were centrifuged at 12,000 rpm for 10 min and the supernatant recovered and mixed with an equal volume of ice-cold isopropanol. DNA was recovered as a pellet by centrifugation at 12,000 rpm for 5 min, washed with 100 μL of 70% ethanol, dried under vacuum and dissolved in 30 μL of TE buffer. Two microliter of RNase (10 g/mL) were added to the DNA and incubated for 1 h. DNA concentrations were measured using the QuantusTM Fluorometer (Promega).

AFLP analysis method

AFLP polymorphic markers used in this study were selected on the basis of the polymorphism they had revealed in 350 rice accessions in our previous work (Yelome et al. 2018). Genomic DNA was completely digested with 5 U of each MseI and EcoRI restriction enzymes and ligated to 5 pmol EcoRI and 25 pmol MseI adapters. Pre-amplification of ligation products was done with specific EcoRI and MseI primers without additive nucleotides and over 20 cycles of 94 °C during 30 s, 56 °C during 60 s, and 72 °C during 60 s were run in a thermal cycler. Concentration of the amplified DNA was checked on a 1.5% agarose gel. Next, DNA was diluted 25 times in TE buffer. For selective amplification, three primer combinations with three selective bases at the end (EcoRI + ACC/MseI + CTC,   EcoRI + ACC/MseI + CTA, EcoRI + ACC/MseI + CAT) were used to reveal a maximum number of polymorphic markers in the rice samples. Based on the entire collection of 350 accession, the primer combination EcorI + ACC/Mse + CAT appeared to be most informative with 79% polymorphic bands, followed by EcorI + ACC/Mse + CTC and Ecor I + ACC/Mse + CTA, with 69% and 66% polymorphic bands, respectively (Yelome et al. 2018). The EcoRI primer was end-labeled with NED fluorescent dye. Amplified products were mixed with GeneScan-500 ROX size standard (Applied Biosystems Inc., USA) and deionized formamide. Finally, they were loaded on an Applied Biosystems electrophoresis instrument 3130 for electrophoretic size separation with performance optimized polymer (POP 4).

SSR analysis method

A total of 20 polymorphic SSR markers distributing across the 12 chromosomes in rice was used for genotyping rice accessions (Suppl. Table 2). These markers were selected from a panel of 30 standard SSR markers developed by the Generation Challenge Program for rice diversity analysis (http://gramene.org/markers/microsat/50_ssr.html) and from other scientific works (Temnykh et al. 2000; McCouch et al. 2002; Dramé et al. 2011) according to their level of polymorphism in rice. The forward primers were labeled at the 5′ end of the oligonucleotide either with the fluorescent dyes 6-FAM[TM] (Blue) or 5-TET[TM]

(Green) and supplied by Applied Biosystems Inc. (USA). We tested them first for amplification and polymorphism on a subset of 11 DNA samples. Then, individual polymerase chain reaction (PCR) amplifications with each marker were carried out in a total volume of 10 μL containing 2 μL of genomic DNA (5 ng), 0.125 μL of each of the forward and reverse primers (diluted to a 5 μM concentration), 0.4 μL of the dNTPs mix (100 mM), 2 μL PCR buffer (Promega), 0.6 μL of Mgcl$_2$ and 0.0625 μL of Taq DNA polymerase (5u/μL, Promega). PCR runs were performed according to the following program: initial denaturation at 95 °C during 2 min, followed by 30 cycles of 30 s at 95 °C, the annealing temperature during 1 min, elongation 72 °C during 1 min and a final elongation step at 72 °C during 5 min. After the PCR reaction, PCR products were migrated on agarose gel (2%) to check the amplification. Of the PCR product, a master mix of 10 μL was prepared into optical 96-well MicroAmp plates using non-overlapping markers of two different dyes (FAM and TET) with reference to the allele sizes published on Gramene website. The master mix was composed of 1 μL of each of the fluorescent dyes (FAM: Blue and TET: Green), 7.7 μL of deionized formamide and 0.3 μL GeneScan-500 ROX size standard (Applied Biosystems Inc., USA). Products were denatured by heating during 3 min at 90° C and then loaded on an Applied Biosystems electrophoresis instrument 3130 for electrophoretic size separation with performance optimized polymer (POP 4).

Phenotypic description

Morphological characterization (phenotypic description of rice accessions) was conducted at AfricaRice's experimental site in Cotonou, Benin in upland ecology. The experimental layout was an augmented randomized complete block design with three reference controls (WAB0006881 (*O. sativa*), WAB0006871 (*O. glaberrima*) and WAB0007880 (*O. glaberrima*)) (Suppl. Table 1). Such a design does not require to repeat the accessions and is convenient for accessions with limited seed as it was the case in this study. A total of 9 blocks with 5 plots consisting of 42 rice accessions appearing exactly with the three reference controls replicated 9 times (once in each block of the design). Each single rice accession was sown in a plot of 5 m^2 with 20 cm spacing within rows

and lines. Thinning of seedlings was done at 14 days after sowing to leave one plant 20 cm apart. Morphological data were collected using 17 morphological traits described in the Standard Evaluation System for Rice (http://www.knowledgebank.irri.org/ses/SES.htm). NPK (15-15-15) was applied at planting (200 kg/ha) and two urea applications performed at top-dressed stage (35 kg/ha) and panicle initiation stage (65 kg/ha), respectively.

Statistical analysis

AFLP/SSR electropherograms were analyzed by GeneMapper Software Version 4.0. Genotype marker results were automatically compiled in a standard Excel format. Statistical parameters defining the diversity, the polymorphism Information Content (PIC) (Botstein et al. 1980), the Shannon's diversity index, allele number, frequency of major alleles, and heterozygosis were calculated using GenAlEx V6.502. The population structure was investigated by using the Bayesian model-based software program in STRUCTURE V2.3.4 (Pritchard et al. 2000). The number of resumed populations (K) was set from 1 to 10 with the 50,000 lengths of burnin period and 200,000 MCMC (Markov Chain Monte Carlo) repeats after burnin and 10 independent runs for each K number. The python script structureHarvester.py v0.6.92 (Earl and vonHoldt 2012) was used to summarize STRUCTURE output. This script generates ΔK values using the method described by Evanno et al. (2005). The optimal number of populations was determined by plotting the K number against the calculated ΔK (log probability for the rate of change of the data) and the sharp peak from STRUCTURE runs was considered. The run with maximum likelihood was then used to assign each accession into populations. Accessions with affiliation probabilities (inferred ancestry) ≥ 80% were assigned to a distinct population, and those < 80% were treated as admixture, i.e. these accessions seem to have a mixed ancestry from identified parents that belong to different gene pools or geographical origins. Using DARwin V6.0.8, genetic dissimilarity coefficients (Jaccard's distance) were calculated to determine the genetic relationship between accessions and dendrograms were constructed according to the unweighted pair-group mean arithmetic method (UPGMA). An analysis of molecular variance (AMOVA) was used to

estimate Fst (fixation) values and the partitioning of the molecular variance of rice accessions. Mantel test was used to investigate the correlation of genetic distances with geographical and phenotypical distances (Mantel 1967). and Principal coordinate analysis (PCo) was performed to show the associations between genetic populations.

Results

Genetic variability for AFLP and SSR markers among subset of 42 rice accessions

A total of 1232 fragments were identified in the subset of 42 rice accessions by 77 polymorphic AFLP markers with an average of 25.67 loci for each of the three primer combinations used (Table 1). The Shannon's information index (*I*) and the Polymorphism Information Content (PIC) averaged 0.46 and 0.35 ranging between 0.42–0.49 and 0.33–0.37, respectively. Jaccard's distance coefficients varied from 0.15 to 0.96 across all the tested samples.

Twenty SSR markers were used to generate 223 different alleles in total rice germplasm with an average of 11.15 alleles per locus (Table 2). Nearly half (49.32%) of the detected alleles was rare with a frequency < 0.05. Major allele frequency ranged from 0.16 to 0.56 averaging at 0.35. Three loci (RM162, RM287, RM316) revealed abundant alleles (frequency ≥ 50%) across all the tested samples. The highest number of different alleles was observed with marker RM5851 (19 alleles) followed by RM219 (18 alleles) and RM1236 (17 alleles), whereas a lower number (5 alleles) was found at RM125, RM162 and RM171 loci. The Shannon's information index (*I*) varied from 1.01 (RM162) to 2.68 (RM5851) with an average value of 1.88. PIC values ranged from 0.58 (RM162) to 0.91 (RM5851 and RM219) with an average of 0.78. Markers RM55, RM162 and RM1075 revealed more than 40% heterozygous individuals, whereas RM125 was homozygous. Jaccard's distance coefficients varied from 0.1 to 1 among the subset germplasm.

Population structure and genetic variation among subset of 42 rice accessions

The Bayesian based analysis of population structure using SSR markers showed that the log likelihood at K = 3 was optimal to group the subset of 42 rice accessions into three genetically distinct populations (Suppl. Fig. 1). Estimated affiliation probability values (inferred ancestry) ≥ 80% at K = 3 indicated that 9, 6 and 26 rice accessions were classified into population 1 (red color), population 2 (green color) and population 3 (blue color), respectively with more than 96% ancestry (Fig. 1). Accession WAB0006684 (*O. sativa*) from Benin shared ancestry with rice accessions in population 1 (77%) and population 2 (23%). Population 1 contained all remaining *O. sativa* accessions included in this study. The cluster analysis based on the UPGMA method using the 20 SSR markers also partitioned the total germplasm into three genetic groups (Fig. 2a). Accessions in each group corresponded to the three populations identified by STRUCTURE except for the admixture WAB0006684 (*O. sativa*). AMOVA results (Table 3) revealed higher levels of variation between rice accessions (80%) than within accessions (18%), whereas only 2% of variation were due to difference between the three populations. The overall Fixation (Fst) value was equal to 0.018 across populations. Fst values were low between all pairwise populations, but significant (*P* < 0.05) for genetic population structure

Table 1 Polymorphism levels detected by number of loci and bands per primer combination, Shannon's information index (I) and Polymorphism Information Content (PIC) for the AFLP markers

| Primer combinations	Number of loci detected	Number of total bands	I	PIC
ACC + CAT	31.00	478.00	0.46	0.37
ACC + CTC	24.00	341.00	0.42	0.33
ACC + CTA	22.00	413.00	0.49	0.35
Mean	25.67	410.67	0.46	0.35
Total	77.00	1232.00		

Table 2 Polymorphism levels detected by major allele frequency, number of different alleles (Na), number of effective alleles (Ne), Shannon's information index (I), observed heterozygosity (Ho) and Polymorphism Information Content (PIC) for 20 SSR markers on 11 chromosomes (Chr.)

Locus	Chr.	Frequency	Missing (%)	Na	Ne	I	Ho	PIC	Allele size (bp)
RM19	12	0.40	0.00	10.00	4.48	1.84	0.10	0.78	204–246
RM55	3	0.17	2.38	15.00	9.01	2.37	0.46	0.89	217–238
RM162	6	0.50	2.38	5.00	2.36	1.01	0.59	0.58	191–209
RM171	10	0.41	2.38	5.00	3.70	1.45	0.05	0.73	327–344
RM219	9	0.16	2.38	18.00	10.92	2.59	0.29	0.91	190–230
RM234	7	0.21	0.00	10.00	7.19	2.10	0.07	0.88	130–159
RM237	1	0.31	4.76	8.00	5.46	1.87	0.10	0.82	122–131
RM287	11	0.50	0.00	9.00	3.25	1.57	0.10	0.69	93–111
RM316	9	0.56	2.38	6.00	2.65	1.23	0.02	0.63	174–206
RM1075	2	0.23	0.00	13.00	7.52	2.23	0.45	0.87	181–210
RM1235	8	0.43	2.38	11.00	3.36	1.66	0.05	0.70	100–119
RM1236	10	0.43	2.38	17.00	4.41	2.01	0.29	0.77	119–198
RM3317	4	0.42	0.00	10.00	4.33	1.81	0.17	0.79	126–145
RM3483	12	0.48	0.00	11.00	3.51	1.64	0.12	0.72	144–207
RM3740	1	0.33	0.00	12.00	5.34	2.00	0.02	0.81	114–146
RM5851	12	0.17	0.00	19.00	11.68	2.68	0.05	0.91	204–270
RM125	7	0.34	2.38	5.00	3.81	1.42	0.00	0.74	118–122
RM133	6	0.37	0.00	11.00	5.04	1.94	0.36	0.81	216–226
RM223	8	0.21	0.00	15.00	7.98	2.35	0.26	0.88	137–160
RM280	4	0.43	0.00	13.00	4.27	1.91	0.02	0.77	147–180
Mean		0.35	1.19	11.15	5.51	1.88	0.18	0.78	
Min		0.16	0.00	5.00	2.36	1.01	0.00	0.58	
Max		0.56	4.76	19.00	11.68	2.68	0.59	0.91	

Fig. 1 Population structure of 42 rice accessions at k = 3 using 20 SSR markers
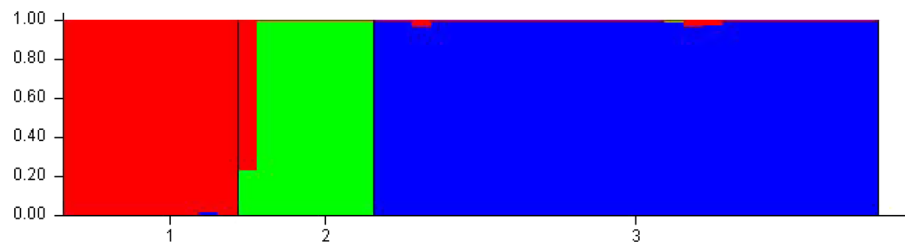
differentiation. A comparable level of population genetic differentiation (Fst) was observed between population 1 and population 2 (0.032) and between population 2 and population 3 (0.032), whereas a much lower Fst (0.016) was found between population 1 and population 3. The average number of different alleles and private alleles per locus was higher in population 3 (6.60 and 3.95) than in population 1 (5.65 and 3.25) and population 2 (2.50 and 0.80). The number of common alleles observed in 50% of individuals in the three populations was null. The Shannon's Information Index (I) and expected heterozygosity (He) were lower in population 2 (0.58 and 0.32) and population 3 (1.34 and 0.63) than in population 1 (1.43 and 0.68) (Table 4). Minimum Nei genetic distance of 1.5 was found between population 1 and 3, whereas population 2 and 3 were the most divergent populations (2.12).

(a)

(b)
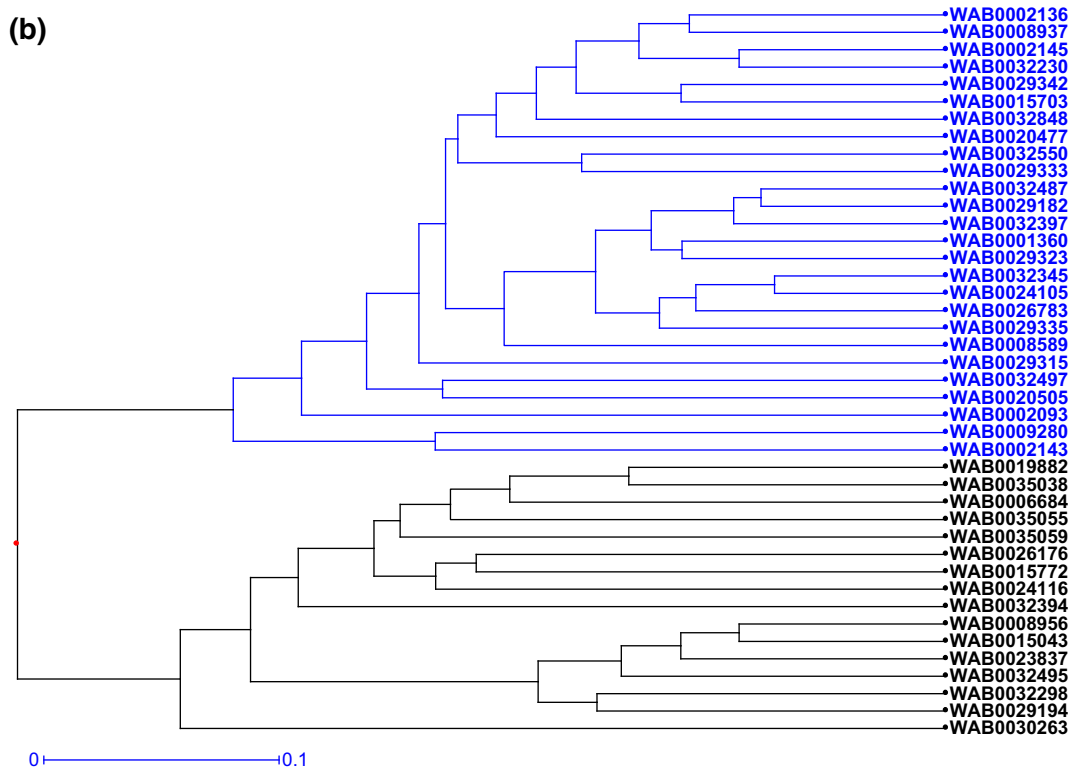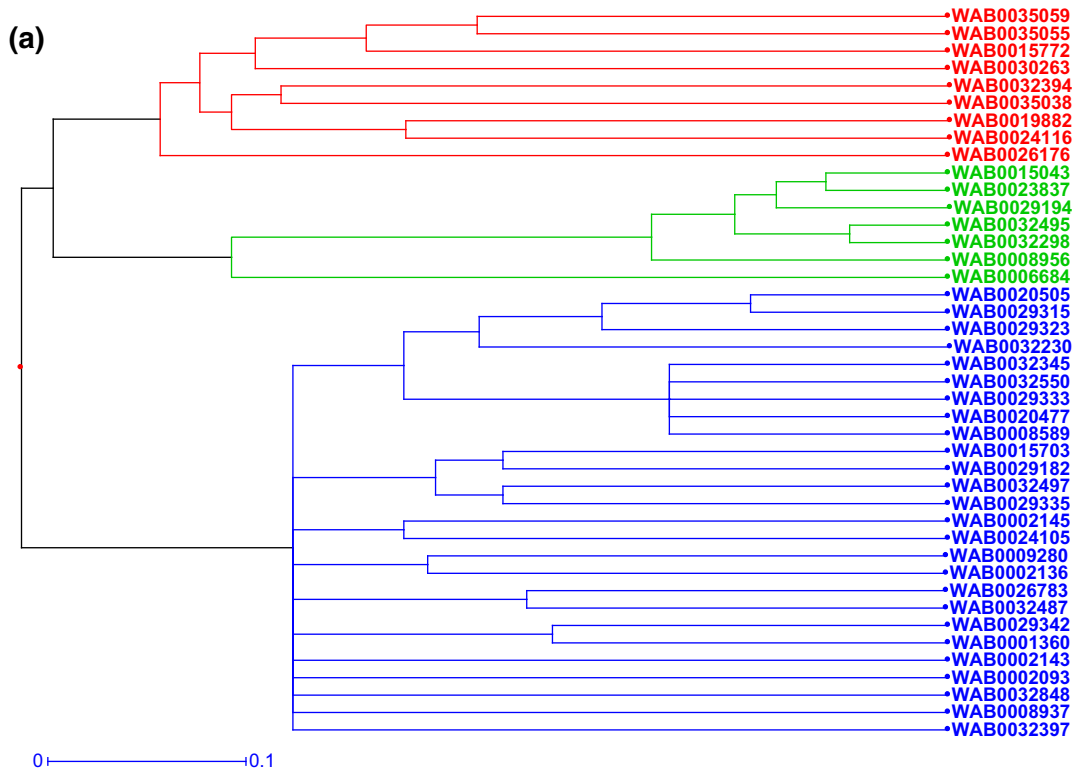
◀ **Fig. 2** UPGMA dendrogram of the subset of 42 rice accessions based on Jaccard's distance from **a** SSR markers and **b** AFLP markers; colors indicates identified clusters

Based on AFLP data, STRUCTURE analysis revealed that no populations could be distinguished across the subset of 42 rice accessions. UPGMA analysis showed two genetic clusters, cluster 1 (blue color) and cluster 2 (black color) comprised of 26 and 16 rice accessions, respectively (Fig. 2b). Cluster 1 corresponded to Pop 3, whereas cluster 2 corresponded to both Pop 1 and Pop 2. Mantel-test results (Suppl. Fig. 1) showed a significant correlation between genetic distances based on AFLP markers and genetic distances based on SSR markers ($R = 0.76$, $P = 0.01$). AMOVA results (Table 3) pointed to significant variations between clusters (52%) and within clusters (48%). The Nei genetic distance between cluster 1 and cluster 2 was estimated at 0.29. Both clusters showed an equal value of Shannon's Information Index (I) (0.33) and expected heterozygosity (He) (0.22), whereas the number of

total bands detected in cluster 1 (772) was much higher than in cluster 2 (460) (Table 4).

Geographic pattern and genetic relationship among subset of 42 rice accessions

Mantel-test revealed no correlation between genetic distances estimated by pairwise Fst using SSR markers and geographic distances for Benin, Liberia, Mali and Nigeria ($R = 0.13$, $P = 0.37$). Samples from Guinea and Ivory Coast, containing one sample each, were not included in the analysis (Fig. 3). Jaccard's genetic dissimilarity coefficients ranged from 0.48 to 9 in population 1. The highest similarity was found between Beninese accessions WAB0035055 and WAB0035059, whereas a large divergence was found between WAB0026176 (Mali) and WAB0035038 (Benin). In population 2, accessions WAB0032495 (Liberia) and WAB0032298 (Mali) shared 10% of genetic similarity, whereas WAB0023837 (Mali) and WAB0008956 (Liberia) were more divergent (40% dissimilarity). In population 3 he highest genetic similarity (20%) was found between Nigerian accessions WAB0029315 and

Table 3 AMOVA analysis for genetic groups differentiation by AFLP and SSR markers in the subset of 42 rice accessions

Source of variation	df	SS	MS	Estimated variance	Percentage of total variance	Probability (P)
AFLP markers						
Among clusters	1	203.48	203.48	9.81	52	< 0.001
Within clusters	40	366.38	9.16	9.16	48	
Total	41	569.86		18.97	100	
SSR marker						
Among Pops	2	43.13	21.56	0.18	2	< 0.001
Among Ind.	38	672.21	17.69	7.96	80	
Within Ind.	41	73.00	1.78	1.78	18	
Total	81	788.34		9.91	100	

Table 4 Polymorphism levels estimated among accessions groups identified by AFLP and SSR markers

Genetic groups	Number of samples	Total of alleles	I	He
AFLP markers				
Cluster 1	26 (*glaberrima*)	772	0.33	0.22
Cluster 2	16 (5 *sativa*, 11 *glaberrima*)	460	0.33	0.22
SSR markers				
Pop 1	9 (5 *glaberrima*, 4 *sativa*)	5.65	1.43	0.68
Pop 2	6 (*glaberrima*)	2.50	0.58	0.32
Pop 3	26 (*glaberrima*)	6.60	1.34	0.63

Fig. 3 Relationship between geographic distance and genetic distance
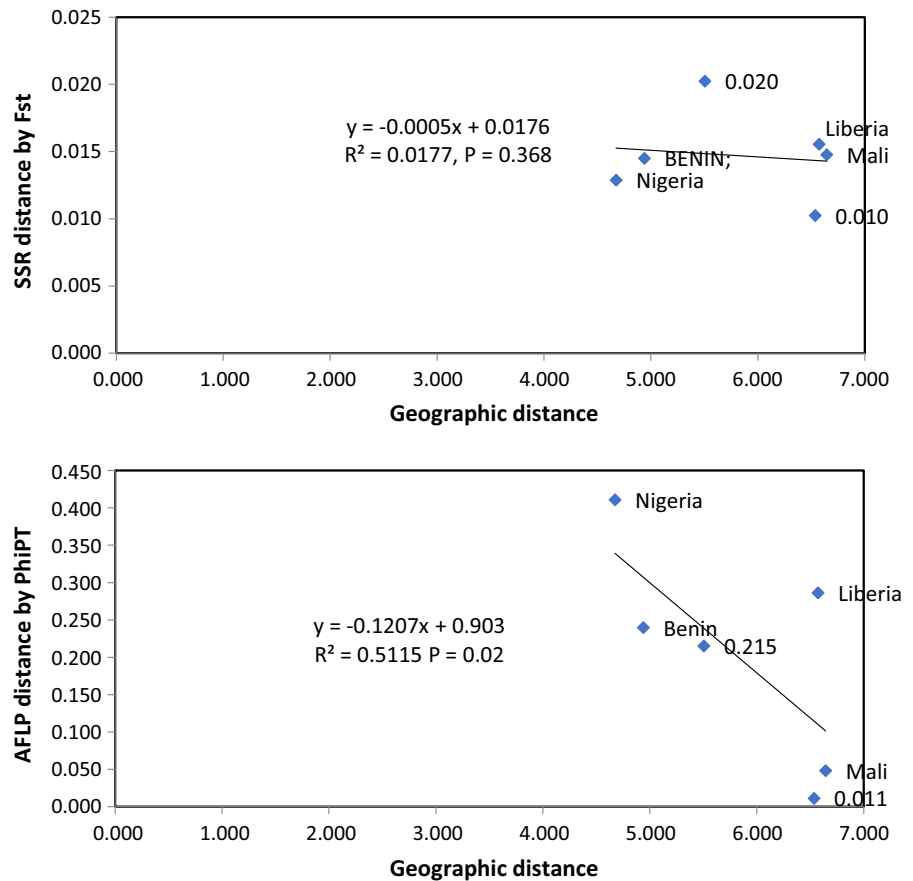
WAB0020505, whereas WAB0020477 (Nigeria) and WAB0032848 (Mali) were the most divergent (88% of dissimilarity) rice accessions.

Based on AFLPs, geographic distances were significantly associated with genetic distances (R = 0.72, P = 0.02). The graphics showed a continuous and linear decrease of genetic similarity when geographic distance increased (Fig. 3). High genetic distance was found between Benin and Nigeria (0.41), whereas Mali was closer to Benin (0.01). Cluster 1 was predominated by accessions from Mali and Nigeria, whereas all accessions from Benin were classified in cluster 2 with 2 accessions out of the three from Liberia. High genetic similarity of 15% was found between Malian accessions (WAB0032345 and WAB0024105) in cluster 1, whereas Nigerian accessions WAB0029333 and WAB0009280 were divergent at 80% (20% of genetic similarity). The most similar (82%) accessions in cluster 2 were WAB0015043 (Ivory Coast) and WAB0008956 (Liberia), whereas WAB0030263

(Nigeria) and WAB0023837 (Mali) were most divergent (80% dissimilarity).

Phenotypic discrimination and gene flow between *O. sativa* and *O. glaberrima* rice accessions

Mantel's statistical test revealed a significant association of AFLP markers with the 17 morphological traits evaluated (R = 0.60, P = 0.01). A number of characteristics differentiated cluster 1 from cluster 2. Accessions in cluster 1 were discriminated by the presence of anthocyanin coloration on basal leaf sheath, opened culm habit and spreading attitude of panicle branches whereas accessions in cluster 2 showed relatively opened culm habit and semi-compact panicle type. Secondary branching was particularly sparse in cluster 1, whereas cluster 2 had a dense panicle type. Cluster 1 showed truncate ligule shape and erect panicle form, whereas cluster 2 had cleft ligule shape and drooping panicle. Based on these

morphological distinctive traits, accessions in cluster 1 were found very similar to the two *O. glaberrima* (WAB0006871 and WAB0007880) reference controls included, whereas cluster 2 resembled *O. sativa* (WAB0006881) control. UPGMA tree using 77 AFP markers classified WAB0006881 (*O. sativa* control) in cluster 2, whereas the two *O. glaberrima* (WAB0006871 and WAB0007880) controls were grouped together in cluster 1 (Suppl. Fig. 3).

Based on SSR markers, allelic pattern, was investigated between the 5 *O. sativa* rice accessions and 37 *O. glaberrima* accessions involved in the three identified populations. On average, in both population 1 and 2, *O. glaberrima* accessions showed a similar pattern of diversity indexes with *O. sativa* accessions, whereas *O. glaberrima* accessions in population 3 performed a specific different pattern (Fig. 4). Principal coordinates analysis (PCoA) based on Nei's genetic distance (Suppl. Fig. 4) showed that *O. glaberrima* accessions in population 1 were closer related to *O. sativa* accessions and more isolated from the other accessions. The lowest fixation index (Fst = 0.001, $P = 0.48$) was observed between *O. glaberrima* accessions in population 1 and *O. sativa* accessions. Fst estimates for each of the 20 SSR markers analyzed were non-significant at 18 loci ($P > 0.05$) between *O. glaberrima* accessions in population 1 and *O. sativa* accessions.

Discussion

Mean PIC value (0.35) revealed by AFLPs in the currently evaluated germplasm is higher than that observed (0.24) previously among the entire germplasm of 350 accessions (Yelome et al. 2018),

indicating higher genetic diversity levels among studied rice collection of 42 rice accessions. Similar results have been reported by Roy Choudhury et al. (2014). These authors used SNP markers to assess genetic diversity and population structure of 6984 rice accessions by selecting a core set of 701 accessions. Agrama et al. (2009) showed that the genetic diversity level increases with elimination of genetically similar accessions during core set development. UPGMA cluster analysis using AFLPs identified two genetically different clusters, which were influenced by geographic distribution of assembled rice accessions. The highest similarity was recorded between Malian rice accessions (cluster 1) and between rice accessions originating from Ivory Coast and Liberia (cluster 2). Historically, it has been demonstrated that *O. glaberrima*, which is derived from the wild ancestor *O. barthii* A. Chev. (syn. O. breviligulata A.Chev. & Roehr.), was first domesticated in the Niger river delta (Mali), and subsequently evolved to two secondary centers: (1) through the Guinea highlands to Niger, crossing along the border of Benin and diffusing to Nigeria; and (2) through the Guinea forest (between Sierra Leone and the western Ivory Coast) to Liberia (Portéres 1962; Second 1982; Sarla and Swamy 2005; Wang et al. 2014). The domestication history of *O. sativa* is more recent, although recent research findings suggest that both species were domesticated independently (Wang et al. 2014). Genetic similarity decreased when geographic distance increased showing a clinal variation of AFLPs markers over the geographical origin of current germplasm studied (Diniz-Filho et al. 2013).

Although AFLPs generated the largest number of total alleles (1232 bands), higher ranges of Shannon's diversity index (1.01–2.68) and PIC (0.58–0.91) were
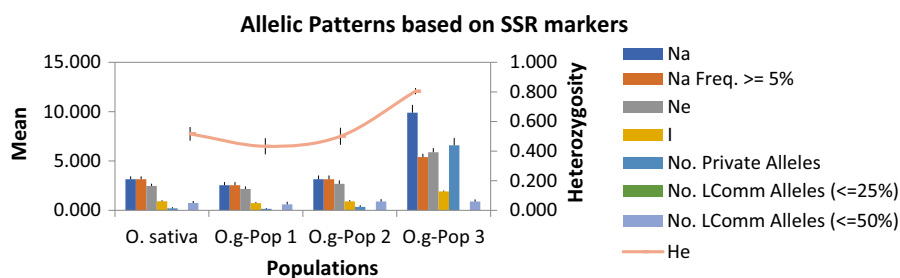
Fig. 4 Mean allelic patterned between *O. sativa* and *O. glaberrima* accessions groups for the number of different alleles (Na), the number of different alleles (Na) with a Frequency ≥ 5%, number of effective alleles (Ne), number of

alleles unique to a single population, number of locally common alleles (Freq. ≥ 5%) found in 25% or fewer Populations and Shannon's Information Index

estimated by SSR markers indicating more effectiveness in detecting genetic variation in the selected rice accessions. Comparing the two marker systems, microsatellites have revealed the highest polymorphism levels in soybean (Powell et al. 1996), lentil (Idrissi et al. 2015) and rice germplasm (Tarang and Gashti 2016). Vieira et al. (2016) explained that there are some advantages of using SSR markers to detect polymorphism because they are of multi-allelic and codominant in nature. SSR markers have the potential to detect high degrees of polymorphism compared other marker systems (Wu and Tanksley 1993; Saghai-Maroof et al. 1996; Powell et al. 1996; Morgante et al. 2002). AFLP-based methods are not always straightforward, as potential individual bands may actually be composed of multiple fragments (Shan et al. 1999). In a comparative study for estimating the amount of genetic diversity, when using a small set of molecular markers, the background analysis may sometimes overestimate the true genetic diversity levels and, therefore, the use of whole genome sequencing approaches would be a better choice for assessing rice cultivars under the current study. But, the evaluation of genetic diversity subjected to the whole genome sequencing is time-consuming and more expensive. In the present study, mean PIC value (0.78) is comparable to that reported by Ming et al. (2010) who have studied genetic diversity of 32 *O. sativa* and 4 *O. glaberrima* rice accessions. They reported PIC average of 0.716 using 54 SSR markers. Higher polymorphism levels were revealed in the present study than in previous studies on African rice (Semon et al. 2005; Dramé et al. 2011; Chen et al. 2017). Possible reasons could be that the germplasm used is a subset collection and the markers chosen were particularly informative. The major reason could be attributed to the large number of different alleles and heterozygosity detected (Kalinowski 2002). The most informative SSR markers were RM219 and RM5851 (PIC value of 0.91), which produced 18 and 19 alleles, respectively. These results are partially supported by Dramé et al. (2011) who observed maximum PIC value of 0.90 at RM219 locus with 17 alleles detected. But, the average number of different alleles per locus (11.15) detected in this study was higher than the numbers reported by Dramé et al. (2011) and Chen et al. (2017) (8.4 and 6, respectively). Out of the 20 tested SSR markers, only RM125 was found homozygous, whereas Dramé et al. (2011) have

recorded the greatest proportion of heterozygous individuals (0.86) at the same locus, RM125. Marker RM125 might thus be very specific to the accessions studied in our work.

The presence of three different populations was revealed by SSR markers, whereas there was no geographic structuring across the three populations. Average Fixation (Fst) value (0.018) was less than 0.05, but significant, suggesting little genetic distinctiveness between these populations (Balloux and Moulin 2002). Germplasm structure showed that *O. sativa* accessions were not totally isolated from *O. glaberrima* accessions. Four out of the five *O. sativa* accessions were grouped together with five *O. glaberrima* accessions in population 1. The *O. sativa* accession WAB0006684 shared ancestry with population 1 and population 2. No morphological difference in ligule shape, panicle form and secondary branching was found between *O. glaberrima* accessions in population 1 and all *O. sativa* accessions (five *O. sativa* accessions from Benin and control WAB0006881) studied. These morphological traits were reported in taxonomic identification of *O. sativa* species (Bezançon and Diallo 2006; Linares 2002). The low range of fixation indexes (Fst) across the three populations suggests a gene flow between both cultivated species. Fst value (0.001) was particularly lower between *O. glaberrima* accessions present in population 1 and all *O. sativa* accessions than those in population 2 and population 3. According to Slatkin (1985) gene flow strongly influences the spatial scale over which genetic differentiation will be observed. Higher amount of gene flow occurred between *O. glaberrima* accessions in population 1 and *O. sativa* accessions since Fst and gene flow are inversely related (Wright 1951). The movement of pollen is usually from *O. sativa* to *O. glaberrima* (Sano 1989). The occurrence of gene flow in the germplasm studied could be attributed to the evolutionary history of these populations, out-crossing between rice accessions from different species or effects of spontaneous mutations (Ko et al. 1994; Jusu 1999; Nuijten et al. 2009). Spontaneous interspecific hybridization is not uncommon in the rice gene pool (Second 1982; National Research Council 1996; Barry et al. 2007; Nuijten et al. 2009). The observed gene flow may have implications for rice breeding by introgression of *O. glaberrima* accessions (population 1) as male parents to increase genetic diversity and introduce new

potentially adaptive alleles into population 2 and population 3 (North et al. 2011; Sexton et al. 2011). Should these introgressed *O. glaberrima* still be considered as belonging to this species or as intra-specific accessions? What is the contribution of this geneflow into valuable traits such as biotic and abiotic stresses such as disease resistance? Further studies are needed to answer these questions.

In general, the present study showed that the subset of 42 rice accessions (*O. glaberrima* and *O. sativa*) can be classified according to the country of origin by AFLP markers. Highest polymorphism levels were detected by SSR markers. Fst values indicate a low proportion of genetic variation between the three populations identified by SSR markers. Both AFLPs and SSR markers reveal the presence of gene flow between *O. glaberrima* and *O. sativa* rice accessions that can be used to expand genetic diversity for rice improvement.

Authors' contributions   OIY carried out the field works, the genotyping, data analysis and drafted the manuscript. KA participated in project design, genotyping, data analysis and revised the manuscript. SL participated in statistical analysis of data and revised the manuscript. DS, AD, WV, PVD, and GH participated in project design, data analysis and revised the manuscript.

Compliance with ethical standards

Conflict of interest   The authors declare that they have no conflict of interest.

References

Agrama H, Yan WG, Lee F, Fjellstrom R et al (2009) Genetic assessment of a mini-core subset developed from the USDA Rice Genebank. Crop Sci 49:1336–1346

Ammiraju JS, Fan C, Yu Y, Song X, Cranston KA, Pontaroli AC, Lu F, Sanyal A, Jiang N, Rambo T et al (2010) Spatio-temporal patterns of genome evolution in allotetraploid species of the genus *Oryza*. Plant J. 63:430–442

Angladette A (1966) Le riz. Techniques agricoles et production tropicale. Maison neuve et La rose Ed., Paris (Ve), pp 609–669

Balloux F, Moulin NL (2002) The estimation of population differentiation with microsatellite markers. Mol Ecol 11:155–165

Barry MB, Pham JL, Noyer AJL, Billot AC, Courtois AB, Ahmad AN (2007) Genetic diversity of the two cultivated rice species (*O. sativa* and *O. glaberrima*) in maritime Guinea. Evidence for interspecifc recombination. Euphytica 154:127–137

Bezançon G, Diallo S (2006) *Oryza glaberrima* Steud. Record from Protabase. In: Brink M, Belay G (eds) PROTA (Plant Resources of Tropical Africa), Wageningen. http://database.prota.org/search.htm

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Chang TT (1976) The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica 25:425–441

Chen C, He W, Nassirou TY, Nsabiyumva A, Dong X, Adedze YMN, Jin D (2017) Molecular characterization and genetic diversity of different genotypes of *Oryza sativa* and *Oryza glaberrima*. Electron J Biotechnol 30:48–57

Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, Telles MPC, Rangel TF, Bini LM (2013) Mantel test in population genetics. Genet Mol Biol 36:475–485

Dramé KN, Sanchez I, Gregorio G, Ndjiondjop MN (2011) Suitability of a selected set of simple sequence repeats (SSR) markers for multiplexing and rapid molecular characterization of African rice (*Oryza glaberrima* Steud). Afr J Biotechnol 10:6675–6685

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Futakuchi K, Sié M (2009) Better exploitation of African Rice (*Oryza glaberrima* Steud.) in varietal development for Resource-Poor Farmers in West and Central Africa. Agric J 4:96–102

Ge S, Sang T, Lu BR, Hong DY (2001) Phylogeny of the genus *Oryza* as revealed by molecular approaches. In: Khush GS, Brar DS, Hardy B (eds) Rice genetics IV proceedings of the fourth international rice genetics symposium. IRRI, Los Banos, The Philippines, pp 89–105

Idrissi O, Udupa SM, Houasli C, De Keyser E, Van Damme P, De Riek J (2015) Genetic diversity analysis of Moroccan lentil (Lens culinaris Medik.) landraces using simple sequence repeat and amplified fragment length polymorphisms reveals functional adaptation towards agro-environmental origins. Plant Breed 134:322–332

Jones MP, Dingkuhn M, Aluko GK, Semon M (1997) Inter-specific *Oryza sativa* x *O. glaberrima* Steud. Progenies in upland rice improvement. Euphytica 92:237–246

Jusu MS (1999) Management of genetic variability in rice (*Oryza sativa* L. and *O. glaberrima* Steud.) by breeders and farmers in Sierra Leone. Ph.D. thesis. Wageningen University and Research Centre, Wageningen, The Netherlands

Kalinowski ST (2002) How many alleles per locus should be used to estimate genetic distances? Heredity 88:62–65

Ko HL, Cowan DC, Henry RJ, Graham GC, Blakeney AB, Lewin LG (1994) Random amplified polymorphic DNA analysis of Australian rice (*Oryza sativa* L.) varieties. Euphytica 80:179–189

Kubo M, Purevdoj M (2004) The future of rice production and consumption. JFDRS 35:128–142

Linares OF (2002) African rice (*Oryza glaberrima*): history and future potential. Proc Natl Acad Sci USA 99:16360–16365

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y et al (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Ming H, Fang-min X, Li-yun C, Xiang-qian Z, Jojee L, Madonna D (2010) Comparative analysis of genetic diversity and structure in rice using ILP and SSR markers. Rice Sci 17:257–268

Mokuwa A, Nuijten E, Okry F, Teeken B, Maat H et al (2013) Robustness and strategies of adaptation among farmer varieties of african rice (*Oryza glaberrima*) and Asian rice (*Oryza sativa*) across West Africa. PLoS ONE 8(3):e34801. https://doi.org/10.1371/journal.pone.0034801

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. Nat Genet 30:194–200

National Research Council (1996) Lost crops of Africa. Volume 1: Grains. National Academy Press, Washington, p 380

Nisar M, Ghafoor A, Ahmad H, Khan MR, Qureshi AS, Ali H, Islam M (2008) Evaluation of genetic diversity of pea germplasm through phenotypic trait analysis. Pak J Bot 40:2081–2086

North A, Pennanen J, Ovaskainen O, Laine AL (2011) Local adaptation in a changing world: the roles of gene-flow, mutation, and sexual reproduction. Evolution 65:79–89

Nuijten E, Van Treuren R, Struik PC, Mokuwa A, Okry F, Teeken B, Richards P (2009) Evidence for the emergence of new rice types of interspecific hybrid origin in West African farmers' fields. PLoS ONE 4(10):e7335. https://doi.org/10.1371/journal.pone.0007335

Pham JL (1992) Evaluation des ressources génétiques des riz cultivés en Afrique par hybridation intra et interspécifique. Thèse Docteur et sciences, Université de Paris XI ORSAY (France), p 236

Portéres R (1962) Berceaux agricoles primaires sur le continent Africain. J Afr Hist 3:195–210

Portères R (1956) Taxonomie Agrobotanique des Riz cultivés O. sativa Linne. et O. glaberrima Steudel. J Agric Trop Bot Appl 3:341–384

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Pritchard J, Stephens M, Rosenberg N, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Roy Choudhury D, Singh N, Singh AK, Kumar S, Srinivasan K et al (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):e113094. https://doi.org/10.1371/journal.pone.0113094

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal spacer length polymorphisms in barley: mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 83:1757–1761

Saghai-Maroof MA, Yang GP, Biyashev RM, Maughan PJ, Zhang Q (1996) Analysis of the barley and rice genomes by comparative RFLP linkage mapping. Theor Appl Genet 92:541–551

Salem KFM, Sallam A (2016) Analysis of population structure and genetic diversity of Egyptian and exotic rice (*Oryza sativa* L.) genotypes. C R Biol 339:1–9

Sanchez PL, Wing RA, Brar DS (2013) The wild relative of rice: genomes and genomics. In: Zhang Q, Wing RA (eds) Genetics and genomics of rice, plant genetics and genomics. Springer, New York, pp 9–26

Sano Y (1989) The direction of pollen flow between two co-occurring rice species, *Oryza sativa* and *O. glaberrima*. Heredity 63:353–357

Sarla NM, Swamy BPM (2005) *Oryza glaberrima*: a source for the improvement of *Oryza sativa*. Curr Sci 89:955–963

Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Jpn J Genet 57:25–57

Semon M, Nielsen R, Jones MP, McCouch SR (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 169:1639–1647

Sexton JP, Strauss SY, Rice KJ (2011) Gene flow increases fitness at the warm edge of a species' range. Proc Natl Acad Sci USA 108:11704–11709

Shan X, Blake TK, Talbert LE (1999) Conversion of AFLP markers to sequence-specific PCR markers in barley and wheat. Theor Appl Genet 98:1072–1078

Sié M, Ogunbayo SA, Dakouo D, Sanou I, Dembélé Y, N'dri B, Dramé KN, Sanni KA, Toulou B, Glele RK (2010) Evaluation of intra and interspecific rice varieties adapted to valley bottom conditions in Burkina Faso. Afr J Plant Sci 4:308–318

Sié M, Sanni K, Futakuchi K, Manneh B, Mandé S, Vodouhé R, Dogbe S, Dramé KN, Ogunbayo A, Ndjiondjop MN, Traoré K (2012) Towards a rational use of African rice (*Oryza glaberrima* Steud.) for breeding in Sub-Saharan Africa. Genes Genomes Genom 6(Special Issue 1):1–7

Slatkin M (1985) Rare alleles as indicators of gene flow. Evolution 39:53–65

Tarang A, Gashti AB (2016) The power of microsatellite markers and AFLPs in revealing the genetic diversity of Hashemi aromatic rice from Iran. J Integr Agric 15:1186–1197

Teeken B, Nuijten E, Temudo MP, Okry F, Mokuwa A, Struik PC, Richards P (2012) Maintaining or abandoning African rice: lessons for understanding processes of seed innovation. Hum Ecol 40:879–892. https://doi.org/10.1007/s10745-012-9528-x

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000)

Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

USDA-ARS (2013) National Genetic Resources Program. Germplasm Resources Information Network - (GRIN) [Online Database]. National Germplasm Resources Laboratory, Beltsville, MA. http://www.ars-grin.gov/cgi-bin/npgs/html/tax_search.pl. Accessed 05th March 2018

Vaughan DA (1994) The wild relatives of rice: a genetic resources handbook. International Rice Research Institute, Manila, p 137

Vieira MLC, Santini L, Diniz AL, Munhoz CF (2016) Microsatellite markers: what they mean and why they are so useful. Genet Mol Biol 39:312–328

Vos P, Hogers R, Bleeker M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 23:4407–4414

Wambugu PW, Furtado A, Waters DL, Nyamongo DO, Henry R (2013) Conservation and utilization of African *Oryza* genetic resources. Rice 6:29

Wang MH et al (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. Nat Genet 46:982–988

Wright S (1951) The genetical structure of populations. Ann Hum Genet 15:323–354

Wu J, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225

Yelome OI, Audenaert K, Landschoot S, Dansi A, Vanhove W, Silue D, Van Damme P, Haesaert G (2018) Exploring genetic diversity and disease response of cultivated rice accessions against *Pyricularia oryzae* under rainfed upland conditions in Benin. Genet Resour Crop Evol 65:1615–1624