Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 9, Issue, 08, pp.56222-56232, August, 2017

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

MOLECULAR CHARACTERISATION AND DIVERSITY ANALYSIS OF BURKINA FASO RICE LANDRACES USING 23MICROSATELLITE MARKERS AND ESTABLISHMENT OF A CORE COLLECTION

***,[1]Honoré Kam, [2]Marie-NoëlleNdjiondjop, [3]Mark D. Laing and [4]NourollahAhmadi**

[1]Institut de l'environnementet de Recherches Agricoles (INERA), 01 BP 910 Bobo-Dioulasso, Burkina Faso
[2]AfricaRice, 01 BP 2031 Cotonou, Benin
[3]Univ. KwaZulu-Natal (UKZN), School of Agricultural Sciences & Agribusiness, Pietermaritzburg, South Africa
[4]Centre de Coopération international en recherche agronomique pour le développement (CIRAD), Montpellier, Cedex 5, France

ARTICLE INFO

Article History:
Received 13th May, 2017
Received in revised form
19th June, 2017
Accepted 25th July, 2017
Published online 31st August, 2017

Key words:

Admixture,
Collection,
O. Glaberrima,
O. Sativa,
SSR.

ABSTRACT

Understanding the genetic diversity of rice and its population structure is a key to the sustainable *in situ* and *ex situ* management of genetic resources. In this study, the molecular analysis of 330rice varieties collected in 59 villages across Burkina Faso was undertaken, using 23microsatellite markers. These markers were found to be polymorphic with an average allele number of 9.4and a mean PIC per locus of 0.58, ranging from 0.034 (RM338) to 0.838(RM1). The 23 SSR markers were powerful enough to separate the two species (*O. sativa and O. glaberrima*), and to identify a group of most probably resulting from natural crosses between the two species. The *glaberrima* group were divided into two subgroups in relation with their cultivation ecosystem, deep-water and lowland while *O. sativa japonica* accessions were clustered in G2. Accessions of *O. s.indica* species were clustered in three groups (G3, G4 and G5).G3 was composed of accessions of late maturing cropped in lowland environment. G5 accessions were tall and late maturing adapted to deep water condition.G4accessions were of short stature and early maturing adapted to lowland conditions. Using the molecular data and the framework it provided, a core collection of 52 accessions was established that represents the diversity of rice in Burkina Faso. This core collection represents 16% of the collection and account for 89% of the alleles of the whole collection. A core collection of this size could be easily manageable by the national agricultural research institute for conservation *in situ* or *ex situ*.

Citation: Honoré Kam, Marie-NoëlleNdjiondjop, Mark D. Laing and NourollahAhmadi, 2017. "Molecular characterisation and diversity analysis of burkina faso rice landraces using 23 microsatellite markers and establishment of a core collection", *International Journal of Current Research*, 9, (08), 56222-56232

INTRODUCTION

Africa is the only continent where the two cultivated rice species *O. glaberrima* Steud and *O. sativa* L. are grown. *O. glaberrima* originated from Africa (Chang, 1984; Chévalier, 1937). The primary centre of diversification of *O. glaberrima* is the inner delta of the River Niger, currently in the Republic of Mali. The regions of Sene-Gambia, covering South Senegal (Casamance), Gambia and Bissau Guinea, and the Guinea forest between Sierra Leone and western Cote d'Ivoire, are considered the secondary centres of diversification of *O. glaberrima* (Portères, 1950). The Asian rice, *O. sativa,* on the other hand, could have reached the African continent via three possible routes. The first entry might have been Madagascar, and the second through East Africa (Somalia) during the 10th century and the third entry may have happened during the 15th century through both the East and West coasts, brought in by the Portuguese, Dutch, French and the British sailors (Bezançon, 1993). The native African rice *O. glaberrima* was cultivated many centuries before the first Europeans arrived on the West African coasts (Linares, 2002). The introduction of modern varieties with high yield can result in the abandonment of landrace varieties, resulting in narrowed *in situ*genetic diversity, which further reduces the possibilities of genetic improvement. Despite their low yield, landrace varieties harbour important traits that can be used in breeding. Burkina Faso is a West African country bordered by Mali (the primary centre of diversification of *O. glaberrima*) in the North and by Cote d'Ivoire (believed to be one of the secondary centres of *O. glaberrima*) in the South. Both *O. glaberrima* and *O. sativa* are cropped in Burkina Faso(Sié, 1991). In a previous study, undertaken to preserve local rice varieties, especially *O. glaberrima,* an extensive collection of rice samples was

******Corresponding author:* **Honoré Kam,**
Institut de l'environnementet de Recherches Agricoles (INERA), 01 BP 910 Bobo-Dioulasso, Burkina Faso

conducted from November 1983 to February 1984 throughout the country (Sié, 1984). Isozyme markers were used for the genomic evaluation of 312 varieties of this collection, including 289 *O. sativa* and 23 *O. glaberrima* accessions. The *O. glaberrima* accessions were found to be the least diverse, whereas in the *O. sativa* subset, the phenol reaction revealed that *O. s. indica* accounted for 95% of the subset and the remaining 5% was *O. s. japonica*. Enzymatic polymorphisms identified a large number of intermediate forms within the *O. s. indica* subspecies. Two *O. sativa* varieties were found to belong to *Aus* group (Sié, 1991; Sié *et al*., 1999). This previous rice collection was not properly conserved because the Institute de l'Environnementet de Recherches Agricoles of Burkina Faso (INERA) lacks proper seed stores for long term storage. Furthermore, not all of the previous collection by Sié(1984) from Burkina Faso was available within the AfricaRicegenebank. Furthermore, there is a possibility that Burkina Faso rice landraces will start to disappear due to the introduction of modern varieties. Therefore, there was an urgent need to initiate the collection and conservation of rice landraces. During January to May, 2008, 330 accessions of rice were collected in farmers' stocks across the four main rice cropping regions of the country (Cascades, HautsBassins, Boucle du Mouhoun and Sud- Ouest).

Information on the genetic diversity within and amongst related crop varieties is essential for rational use of genetic resources in breeding programmes. Molecular markers are good tools to portray the structure and assess the genetic variability within and amongst different species. Therefore several marker types have been developed to characterize rice populations. Microsatellites, also called Simple Sequence Repeat (SSR) markers, have proved to be powerful tools in the assessment of genetic variation and in the elucidation of genetic relationships within and amongst African rice collections (Barry *et al*., 2007; Semon *et al*., 2005). The genotypic approach looks at neutral diversity, while the phenotypic approach reveals functional diversity. Therefore, agro-morphological characterization complements and illustrates the extent, the organisation and the specificities of genetic diversity revealed by molecular markers. Although abundantgenetic resources canprovide a broader genetic background for crop breeding,a large germplasm collection with numerous redundant accessions may be an obstruction to its conservation and utilisation by NARS in developing countries. The establishment of relatively small core collections are essential for the sustainable use and conservationof the collection. A core collection is a subset of a large germplasm collection that contains accessions chosen to represent the genetic variability of the given collection (Diwan *et al*., 1995). Brown (1989) described the core collection concept as a mean to enhance the efficiency in the utilisation of germplasm collections. Corecollections, with a minimum of entries, represent the diversity of the entire collection as much as possible (Liu *et al*., 1999). A small set of the collection, representing the diversity of the entire collection, can be manageable *in situ* and *ex situ* by NARS in developing countries like Burkina Faso. In this present study, the molecular diversity of 330 rice accessions was assessed in order to create a core collection of rice landraces.

MATERIALS AND METHODS

Plant material: The rice plant material was collected in 2008 in 59 villages of the four main rice cropping regions (Boucle du Mouhoun, Cascades, Hauts-Bassins and Sud-Ouest) of Burkina Faso (Figure 1).The collection team of each region comprised of an INERA staff and an agricultural agent at the departmental level. The agricultural agent of the concerned village typically informed farmers and summoned a meeting. An inventory of local rice varieties was drawn in an organized meeting in the presence of a group of farmers. Farmers were asked to list the rice varieties they have been cropping for several years and also those varieties that were cultivated by their ancestors and the present generation farmers are continuing to grow. Farmers who owned the requested varieties were identified and the collection team went and gathered the varieties in their stocks. Together, farmers identified the chosen varieties by consensus agreement. They provided information about each variety, like the name, the cycle, and the ecosystem. Data on longitude, latitude and altitude of each village were also collected with a GPS (Global Positioning System). Modern varieties released through the channel of agricultural and research networks were not collected. Overall, 330 rice varieties were collected, mostly bulkgrains and sometimes in panicles. Agro-morphological evaluation was conducted in the field at Africa Rice Center research station (6º25.415N and 2º19.684E; altitude: 21 m) located in Cotonou, Benin Republic. Each accession was phenotyped using 19 quantitative and 15 qualitative agro-morphological traits for rice descriptors (Table 1), which were extracted from "Descriptors of Wild and Cultivated Rice"(Bioversity-International *et al*., 2007). The *O. glaberrima* and *O. sativa* species were separated on the basis of ligule length and panicle features. Accessions with long ligules (> 10 mm) and drooping panicles were classified as *O. sativa* species, while accessions with short ligules (<7 mm) and erected panicles were classed as *O. glaberrima* species. Then 47 and 283 accessions, categorized as *O. glaberrima* and *O. sativa* respectively, were analysed for their molecular diversity. Figure 1 portrays the distribution of the two species in the collectionarea. For the purpose of easy identification, the samples collected in Boucle du Mouhoun, Cascades, Hauts-Bassins and Sud-Ouest regions were assigned the prefix BM, CC, HB and SO, respectively.

The code assigned to each measured trait is in accordance with the "Descriptors of Wild and Cultivated Rice"(Bioversity-International *et al*., 2007)

DNA extraction

Four seeds from panicle of each accession were sown and fresh young leaves of 21 days old of the four plants were harvested and put in a single tube for DNA extraction. DNA was extracted according to the method described by Risterucci *et al*. (2000). Two centimetres of frozen leaves were powdered in a grinding machine and mixed with 480 µl of extraction buffer (1.4 MNaCl, 100 mMTrisHCl pH 8.0, 20 mMEDTA, 10 mMNa_2SO_3, 1% PEG 6000, 2% MATAB) preheated at 75°C. The extract was then homogenized for 10 s with a vortex and incubated for 30 min at 75°C; after being cooled at room temperature (25°C), a volume of 480 µl of chloroform-isoamyl alcohol (24:1 v/v) was added. The tube was mixed gently before being centrifuged at 4000 rpmfor 15 min and the supernatant containing the DNA was transferred to new tubes and precipitated at –20°C after the addition of 270 µl of isopropanol. The tubes were centrifuged at 4000 rpm and the supernatant was discarded. A volume of 100 µl ethanol was added and centrifuged at a maximum speed to wash the DNA.

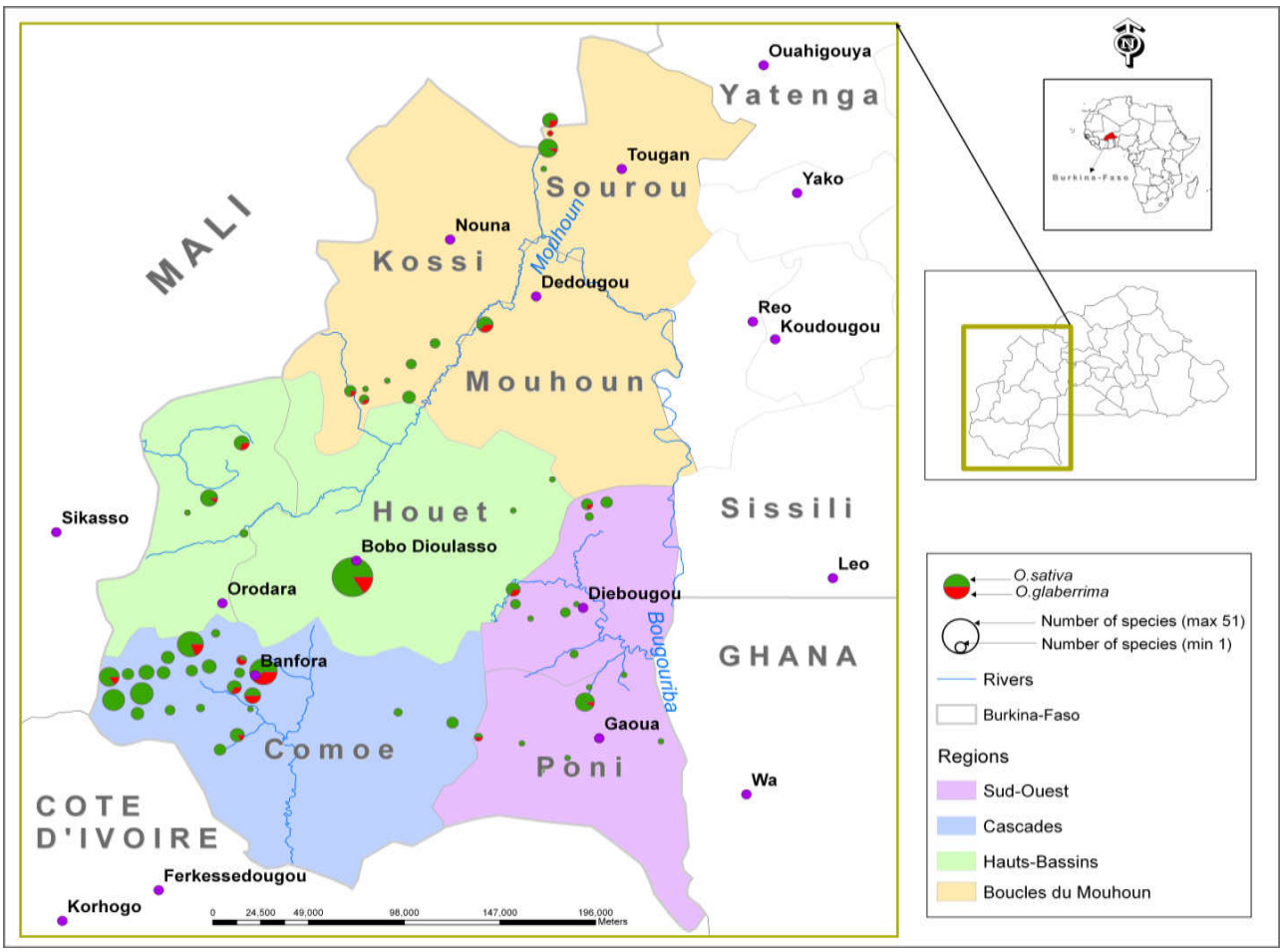


Figure 1. Map showing the four regions visited where *O. glaberrima* and *O. sativa* were collected in Burkina Faso (source: unpublished map, AfricaRice, GIS section)

Table 1. Nineteen quantitative and 15 qualitative phenotypic traits used to evaluate the collection in the field

Code	Quantitative trait measured	Code	Qualitative trait measured
7.2.2.1	Number of days from effective seedling to first heading	7.3.8	Leaf blade attitude
7.2.3.1	Number of days from effective seedling to main heading	7.3.9	Leaf blade pubescence
7.2.4.1.	Number of days from effective seedling to maturity	7.3.9.1	Leafblade pubescence on blade surface
7.3.13	Ligule length (mm)	7.3.22	Flag leaf attitude early observation
7.3.18	Leaf blade length (cm)	7.3.32	Lodging resistance
7.3.19	Leaf blade width (cm)	7.3.34	Flag leaf attitude late observation
7.3.20	Flag leaf length (cm)	7.4.2	Stigma colour
7.3.21	Flag leaf width (cm)	7.4.5	Lemma and palea colour
7.3.25	Culm length	7.4.9	Awn distribution
7.3.26	Culm number	7.4.19	Panicle attitude of main axis
7.4.14	Number of basal primary branches on panicle	7.4.20	Panicle attitude of branches
7.4.17	Number of panicle per plant	7.4.21	Panicle secondary branches
7.5.1	Panicle length (cm)	7.4.23	Panicle shattering
7.5.15	Grain Length (mm)	7.5.10	Sterile lemma length
7.5.16	Grain width (mm)	7.5.23	Caryopsis pericarp colour
7.5.17	Grain thickness (mm)		
7.5.18	100-grain weight (g)		
7.5.20	Caryopsis Length		
7.5.21	Caryopsis Width		

Finally, the DNA was suspended in 30 µl of TE buffer (0.7 M NaCl, 50 mMTRIS-HCl, 10 mMEDTA, pH 7.0). DNA concentration was checked with the fluoroscan apparatus (Labsystems, Vantaa, Finland) and the final concentration dilutedat 5 ng.µl^{-1}.

Genotyping

Twenty three SSR markers, already used for rice genetic diversity studies (Garris *et al*., 2005), were chosen. The Polymerase Chain Reaction (PCR) amplification was performed in a 384 well thermocycler (Mastercycler® 384, Eppendorf) on 5 ng of DNA in a 10 µl final volume containing buffer (10 mMTris-HCl pH 8, 100 mMKCl, 0.05 % w/v gelatin, and 2.0 mM MgCl2), 0.1 µM of reverse primer, 0.08 µM of forward primer, 200 µM of dNTP, 0.1 U of Taq DNA polymerase and 0.1 µM of M13 primer-fluorescent dye IR700 or IR800. The PCR process was: initial denaturation at 94°C for 4 min; 35 cycles of 94°C for 60 s; hybridization temperature for 60 s; elongation at 72°C for 60 s; and a final

elongation step at 72°C for 8 min. The PCR products were mixed with two desmiling reagents migrated in multiplex (two primer pairs on Dye 700 and 800) on polyacrylamide 6.5% gel on a DNA sequencer (Li-cor® DNA analyser Model 4300 (the genotyping platform at Génopole Montpellier Languedoc Roussilon, CIRAD)). Ten Generation Challenge Programme reference set rice varieties (ARC10317, Aswina, Basmati 310 (1), Basmati 310 (2), Dumsikalam, IR64, Khao Yang, Maybelle, Nipponbare and PadiRaoekang) were used as controls for allele size. A size marker was deposited in each side of the gel for allele size detection. The SSR markers were revealed following the protocol of Roy *et al*. (1996), applied with the automated infrared fluorescence technology of LICOR IR2 sequencers in the genotyping and robotics platform of "Centre de Coopération International enRecherche Agronomique pour le Développement" (CIRAD, Montpellier). For a given locus, the forward primer of the SSR was designed with a 5'-end M13 extension (5'-CACGACGTTG TAAAACGAC-3'). Allele sizes were determined using SAGA (version 3.2) software package, which encodes genes in base pairs. Tog5681, RAM63 and CG14 (*O. glaberrima*) were used as checks for the classification of the accessions.

Data analysis

Genotypic data was first assessed for matching genotypes at all loci with GenAlEx 6.2 software (Peakall and Smouse, 2006). The genotypes with identical alleles at all loci were deleted and a matrix of genotypes which differed at least in one locus was considered for further analysis. GenAlEx 6.2 software was also used to compute pairwise population Fst and AMOVA. PowerMarker 3.25 software package (Liu and Muse, 2005) was used to calculate allele number (NA) per locus, Major Allele Frequency (MAF), Heterozygosity (H_o), and Polymorphism Information Content (PIC). The gene diversity, which is referred to as expected heterozygosity (H_e) was calculated as

$H_e = n/(n-1)\ (1-\sum p_i^2-H_o/2n)$;

where n is the number of individuals, p_i the frequency of the i allele and H_o the number of observed heterozygotes (Nei, 1987).

The PIC was calculated as

$PIC = 1-\sum p_i^2-\sum_i\sum_{j>i} 2p_i^2p_j^2$;

where p_i and p_j are the frequencies of the i and j alleles respectively (Botstein *et al*., 1980).

A Bayesian clustering analysis was undertaken to determine the population structure using the Structure 2.3.3 software package (HUBISZ *et al*., 2009; Pritchard *et al*., 2000). The posterior probabilities were estimated using 500,000 iterations of Markov Monte Carlo Chain (MCMC) with a length of burn-in period of 50,000 iterations. The optimal population number (*K*) was determined after 20 independent runs, using the admixture model and the option of correlated allele frequencies between populations, as this configuration is considered best by Falush *et al*. (2003) in the case of subtle population structures. *K* was tested from *K=1* to *K=12* in order to determine the number of populations (*K*) supported by our data. The method of Evanno *et al*. (2005) was used to infer population number. The parameter $\Delta K = m(|L''(K)|)/s[L(K)]$ was calculated as the mean of absolute values of the second order change of the likelihood distribution $L''(K)$ divided by the standard deviation of the likelihood $L(K)$. The optimum value of this distribution was interpreted as the true number of populations, *K*(Evanno *et al*., 2005). To investigate the genetic relationships between accessions, a genetic dissimilarity matrix was computed in DARwin software 6(Perrier and Jacquemoud-Collet, 2006). The dissimilarity between samples was calculated by using simple matching based on the Sokal and Michener (1958) index. The dissimilarity formula is:

$d_{ij} = 1 - \frac{1}{L}\sum_{i=1}^{L}\frac{m_l}{\pi}$With *dij*: dissimilarity between units *I* and *j*, *L*: number of loci, π: ploidy and m_l: number of matching alleles for locus *l*

The dissimilarities were used to perform factorial coordinate analysis for graphical representations on Euclidean plans that conserve, at best, distances between units. The dissimilarities were also employed for the construction of an unweighted neighbour-joining tree (Perrier *et al*., 2003). The neighbour-joining tree was used to build the core collection by using stepwise techniques that proceeds by successive pruning of redundant genotypes. Redundancy means that some genotypes are very close and therefore bring, in part, the same information on their diversity. This procedure searches for a subset of genotypes minimizing redundancy between genotypes, but limits if possible the loss of diversity.

RESULTS

Genetic diversity

Within the 330 accessions, 282 accessions were found to be different in at least in one locus. The genetic diversity was then conducted on the 282unique genotypes, including 39 *O. glaberrima* and 243 *O. sativa* accessions. In total, 217alleles were found with the 23SSRs with a mean of 9.4 alleles per locus. The largest number of alleles was found with RM474 (26 alleles) and the smallest with RM338 (2 alleles). The former distinguished 31 varietal groups, whereas the latter distinguished only three varietal groups. The mean PIC per locus was 0.57 within a range of 0.03 (RM338) to 0.84 (RM1 and RM287). The heterozygosity rate varied from 0 to 6.5% with a mean value of 2.4% in the overall population (Table 2). When comparing SSRs with the different repeat motifs, those with the 2-bp motif had the highest PIC values, particularly the GA repeats (RM1, RM287 and RM514), whereas the 3-bp motif markers (RM338) had the lowest number of alleles and PIC values. Markers with high allele number (Na) values (RM514, RM287, RM1, RM19, RM552, RM154, and RM474) set the samples apart more distinctly in a population, resulting in a higher number of genotypes (varietal groups) and thereby depicting high genetic diversity. The diversity indexes of the two species were compared for the 23 SSR loci. The diversity observed in *O. sativa* was higher compared to that of *O. glaberrima*. The mean PIC was 0.53 in *O. sativa* and 0.25 in *O. glaberrima*. Similarly, the Na of *O. sativa* (8.4) was twice as high as that of *O. glaberrima* (4.0). Among the 23 markers, only eight (RM11; RM1227; RM215; RM237; RM447; RM474; RM5 and RM514) had more than four alleles in *O. glaberrima* accessions, whereas in *O. sativa* accessions, only 4 loci (RM124; RM338; RM452 and RM484) had less than five alleles. These four loci also showed lower allelic diversity in the overall population. Only two SSRs (RM237 and RM474)

had their PIC values higher in *O. glaberrima* than in *O. sativa* (Table 2).

Eight SSR markers (RM1, RM11, RM1227, RM14643, RM154, RM19, RM215 and RM5) separated very well *O. glaberrima* from *O. sativa*. These markers could be important in interspecific breeding between these two species for checking polymorphism for true crosses and could also help to identify *O. glaberrima* from *O. sativa*.

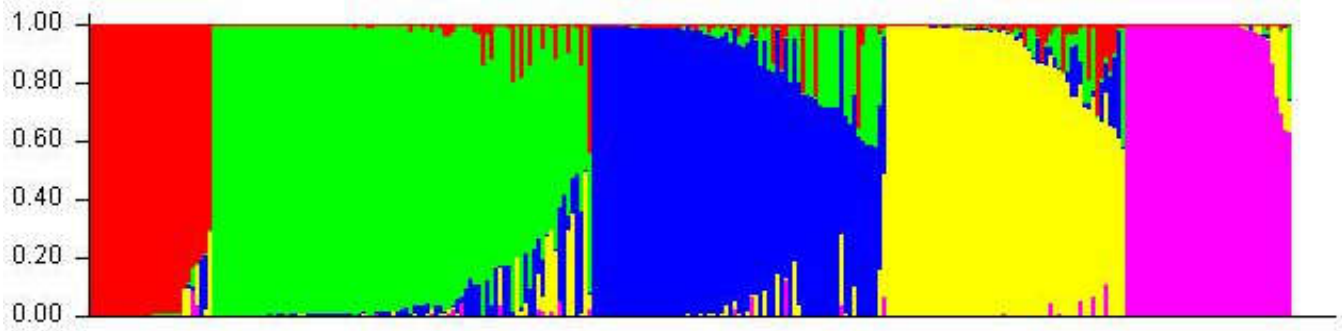
Population structure

The inference of population structure using the method of Evanno *et al*. (2005) in the overall collection strongly supported two populations (*K*=2) as the best number of the group. A clear peak was observed at the value of *K*=2. The first group was designated C1and the second group C2. C1was composed of 48 accessions, while C2had 243accessions. All the39 *O. glaberrima* accessions with short ligules (<7 mm) and erected panicles, including the check varieties CG14 and

Table 2. Comparison of the diversity in the overall collection and between the two species (*O. glaberrima* and *O. sativa*) by the 23 SSR markers

	Total N = 282				Glab N=39				Sat N=243			
Marker	MAF	Na	Het	PIC	MAF	Na	Het	PIC	MAF	Na	Het	PIC
RM1	0.25	17	0.0432	0.8377	0.96	3	0.0769	0.0731	0.30	16	0.0377	0.8149
RM11	0.33	13	0.0469	0.7865	0.60	6	0.1429	0.4362	0.38	10	0.0317	0.7243
RM1227	0.53	9	0.0249	0.5489	0.91	6	0.1282	0.1666	0.61	6	0.0083	0.4137
RM124	0.83	3	0.0072	0.2763	0.99	2	0.0263	0.0256	0.80	3	0.0041	0.3071
RM14643	0.72	10	0.0179	0.4402	0.94	3	0.0513	0.1182	0.83	9	0.0124	0.2998
RM154	0.38	13	0.0360	0.7617	0.97	2	0.0000	0.0487	0.44	12	0.0418	0.7066
RM19	0.32	10	0.0291	0.7836	1.00	1	0.0000	0.0000	0.37	9	0.0338	0.7350
RM215	0.62	7	0.0111	0.5537	0.89	5	0.0938	0.1973	0.70	5	0.0000	0.4502
RM237	0.63	7	0.0108	0.5221	0.39	5	0.0571	0.6239	0.72	6	0.0041	0.4219
RM271	0.54	8	0.0073	0.5863	0.66	3	0.0000	0.3797	0.63	6	0.0084	0.5168
RM287	0.20	11	0.0654	0.8362	0.82	4	0.1579	0.2836	0.24	11	0.0495	0.8095
RM316	0.80	7	0.0143	0.3055	0.82	4	0.0513	0.2781	0.80	5	0.0083	0.2941
RM338	0.98	2	0.0071	0.0343	0.99	2	0.0256	0.0250	0.98	2	0.0041	0.0358
RM431	0.49	6	0.0036	0.5905	0.74	2	0.0000	0.3086	0.53	6	0.0042	0.5304
RM447	0.60	9	0.0540	0.5506	0.67	5	0.1282	0.3903	0.69	9	0.0418	0.4561
RM452	0.77	3	0.0189	0.3141	0.90	2	0.1250	0.1692	0.76	3	0.0083	0.3259
RM474	0.37	26	0.0271	0.7841	0.30	15	0.0625	0.8407	0.42	20	0.0221	0.7366
RM484	0.77	3	0.0000	0.3184	0.77	2	0.0000	0.2885	0.77	3	0.0000	0.3216
RM5	0.45	7	0.0340	0.7001	0.94	5	0.1176	0.1118	0.52	7	0.0216	0.6324
RM510	0.80	5	0.0214	0.3048	0.95	3	0.1026	0.0963	0.92	5	0.0083	0.1458
RM514	0.25	11	0.0352	0.8218	0.37	8	0.0870	0.7636	0.24	11	0.0300	0.8218
RM538	0.39	10	0.0071	0.7086	0.99	2	0.0263	0.0256	0.45	10	0.0041	0.6612
RM552	0.33	20	0.0360	0.8237	0.96	3	0.0769	0.0731	0.22	20	0.0293	0.8651
Mean	0.54	9.4	0.0243	0.5735	0.80	4.0	0.0668	0.2489	0.58	8.4	0.0180	0.5229

MAF: Major Allele Frequency Na**:** Allele Number; Het: Heterozygosity and PIC: Polymorphism Information Content



Red=G2 (*O. s. japonica*); green=G3 (*O. s.indica*group 1); blue=G5 (*O. s. indica*group 3); Yellow=G4 (*O. s. indica*group 2) and purple =G1 (*O. glaberrima*)

Figure 2. Model based ancestries of 282rice genotypic accessions estimated from 23 nuclear SSR loci using Structure (Pritchard *et al*., 2000) showing five (5) groups

Table 3. Fst distances between groups and admixture in the collection

	Adm	Adm-GS	G1	G2	G3	G4	G5
Adm	0,000						
Adm-GS	0,183	0,000					
G1	0,286	0,101	0,000				
G2	0,217	0,272	0,364	0,000			
G3	0,040	0,248	0,369	0,260	0,000		
G4	0,102	0,258	0,381	0,284	0,168	0,000	
G5	0,065	0,285	0,412	0,329	0,123	0,189	0,000

G1: *O. glaberrima*; G2: *O. s. japonica*; G3: *O. s. indica*group 1; G4:*O. s. indica*group 2; G5: *O. s. indica*group 3; Adm: admixture *O. sativa*x *O. sativa* and Adm-GS: admixture *O. glaberrima*x *O. sativa*

Tog5681, were assigned toC1.However, 11 *O. s. japonica* accessions (BM15, BM18, BM24, BM29, CC31, CC145, SO3, HB5, HB53, HB62, and HB84) were falling intoC1.However, K=5 is the one capturing the major structure in the data giving five (5) groups (Figure 2). The *O. glaberrima* species were clearly assigned to group 1 (G1) and the *O. s. japonica* accessions clustered previously in C1 were well assigned in group 2 (G2). The*O. s. indica* in C2 were split in three groups G3, G4 and G5.

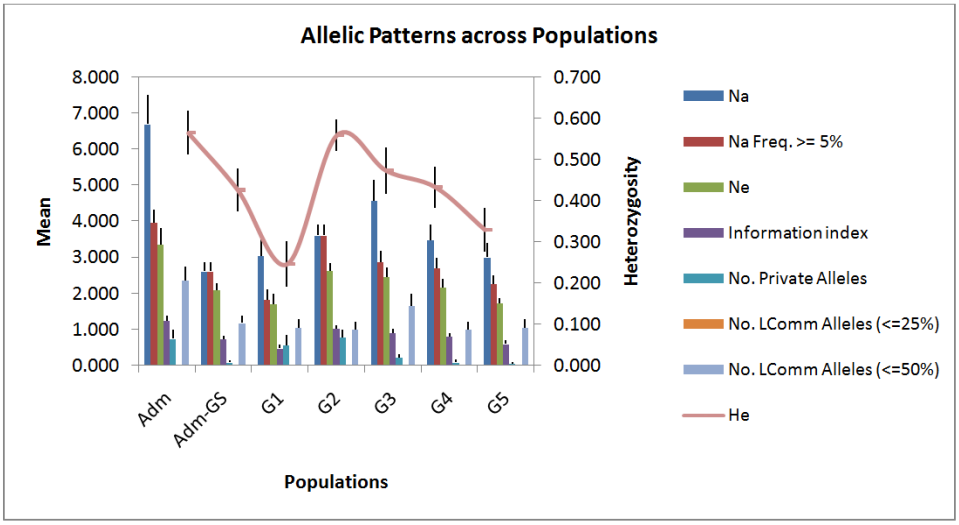
G1 included 36 accessions belonging to the *O. glaberrima* species, G2 included 11 accessions considered as *O. s. japonica* species, G3 included 100 accessions of *O. s. indica*, G4 included 28 accessions of *O. s. indica*, and G5 included 56 accessions of *O. s. indica*. Three admixtures between *O. glaberrima* and *O. sativa* were identified in the collection. They were: CC123A (67% of *O. glaberrima*), HB41 (61% of *O. glaberrima*) and HB46 (70% of *O. glaberrima*).The admixtures *O. glaberrima*by *O. sativa* shared at least 61% of *O .glaberrima* background and maximum 39% of *O. sativa*. These accessions were found in areas where *O. glaberrima* and *O. sativa* are cropped side by side.In total, 48 admixtures *O. sativa* by *O. sativa* were encountered in the collection, showing more diversity (PIC = 0.53) and heterozygosity (0.04) than the five other groups of the collection. The pairwise population Fst showed that G1 is highly different from the four other groups. The values of the Fst indicate that G1 which gather the *O. glaberrima* accessions of the collection has remarkable differences with the four other groups (Table 3). The Fst values confirmed that the admixtures *sativa* by *sativa* are very close to G3, G4 and G5 whereas different from the G1, G2 and Adm-GS.

observed (0.328) is with allele 147 on locus RM237. The population G2 had 18 private alleles with three alleles with a frequency more than 0.5; they are alleles 217, 260 and 132 on locus RM452, RM431 and RM14643 with frequencies 0.727, 0.545 and 0.545 respectively. These alleles contributed more to the differentiation of this population with the others. Each of the 11 accessions of G2 carried a private allele. Furthermore, these three loci RM452; RM14643 and RM447 contributed to differentiate G2 with the other group with a total frequency of private allele of 0.727; 0.727 and 0.700, respectively. G3 had five (5) private alleles with allele 262 at locus RM474 having the highest allele frequency of 0.021. G4 had only two private alleles 294 and 312 both on locus RM474 with the same frequency (0.038). G5 has only one private allele, allele 138 on locus RM11 with a frequency of 0.020. The allelic pattern across the different populations and the admixtures is given in Figure 3.

The AMOVA showed high variation within Admixture *O. sativa* by *O. sativa* populations. The summary AMOVA table (Table 4) showed more variation within population than among populations.

Table 4. Summary AMOVA table between the groups and the admixtures in the Burkina Faso rice collection

Source	df	SS	MS	Est. Var.	%
Among Pops	6	2156,832	359,472	9,257	31%
Within Pops	275	5762,203	20,953	20,953	69%
Total	281	7919,035		30,211	100%



G1: *O. glaberrima*; G2: *O. s. japonica*; G3: *O. s. indica* group 1; G4: *O. s. indica* group 2; G5: *O. s. indica* group 3; Adm: admixture *O. sativa*x *O. sativa* and Adm-GS: admixture *O. glaberrima*x *O. sativa*

Figure 3. Allelic pattern across the five populations and the admixture using 23 SSR markers

The Admixture *sativa* by *sativa* had 17 private alleles but with very low frequency confirming the high variation within that population and meaning that they carried more rare alleles. The highest frequency noted is with the allele 228 in locus RM552 with the frequency of 0.073 and the lowest allele 204 in locus RM154 with the frequency of 0.010. The admixtures *O. glaberrima* by *O. sativa* have only two private alleles: allele 150 and 155 on locus RM11 and RM1227 respectively, with the frequency of 0.167 carried only by the accession HB41.The population G1 has 13 private alleles. The highest frequency

The factorial analysis on axes 1/3 clearly split the five (5) groups identified by structure analysis (Figure 4). The accessions in the G1 group (36 accessions) were characterized by long panicle length (30.2 cm), large flag leaf (19.4 mm), short ligule (4.6 mm), short grains length (8.3 mm), and early maturing plants (107 days). Accessions in the G2 group were shorter plants (91.4 cm) compared to G1 (95.7 cm) but had in average the same cycle and grain length with G1 and were highly different in term of ligule and panicle length. Their ligule is in average 14.2 mm long while their panicle had 26.1

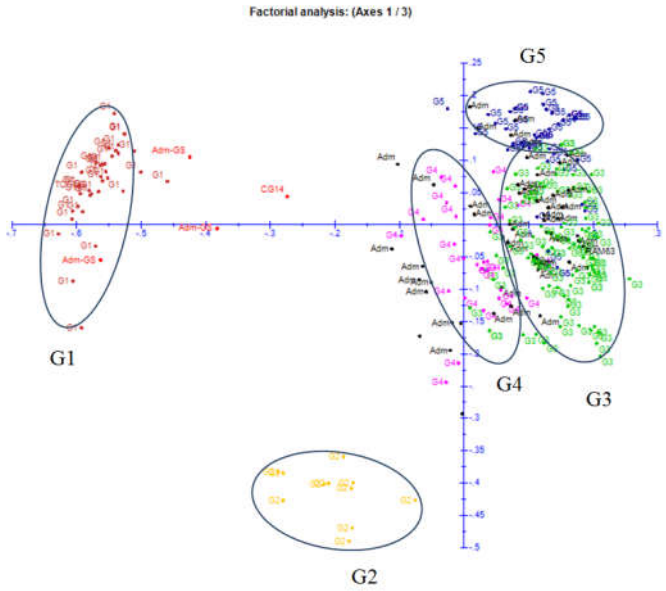
cm length. G3, G4 and G5 had long grain size (9.0 to 9.2 mm) and long cycle compared to G1 and G2. Their cycle was 131.5; 118.1 and 133.8 days for G3, G4 and G5, respectively (Table 5).

By analysing separately *O. glaberrima* accessions of the collection, the factorial analysis split the *O. glaberrima* accessions following regional pattern. The first axis partitioned *O. glaberrima* collected from Boucle du Mouhoun (BM) region from that of Hauts-Bassins (HB) region. The second axis separated *O. glaberrima* collected in Cascades (CC) region from that of the three other regions (Figure 5).

Table 5. Summary of the main phenotypical traits distinguishing the groups identified in the collection by molecular markers

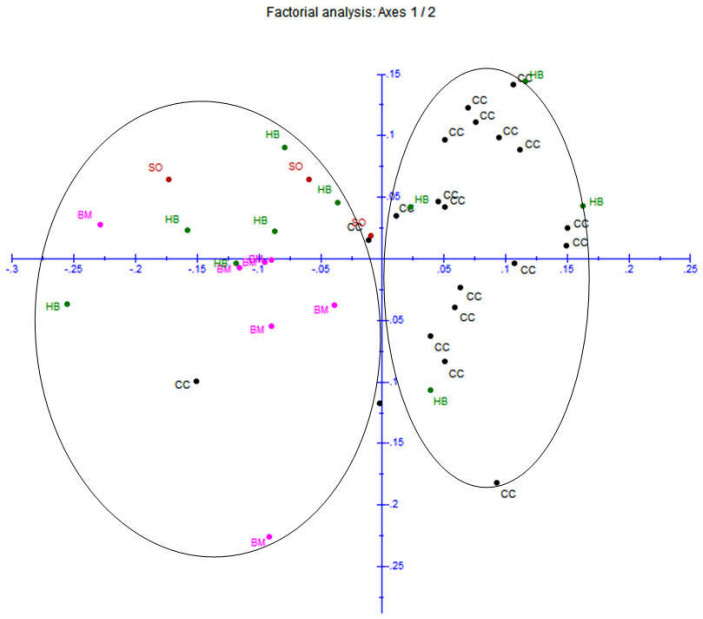
Group	CL (cm)	PL (cm)	FLBW (mm)	cycle (days)	Ligule	GL (mm)
G1 ± sd	95.7 ± 14.2	30.2 ± 2.3	19.4 ± 1.9	107.1 ± 12.9	4.6 ± 1.9	8.3 ± 0.3
G2 ± sd	91.4 ± 14.2	26.1 ± 3.1	15.8 ± 3.6	106.9 ± 16.0	14.2 ± 4.7	8.2 ± 1.9
G3 ± sd	99.9 ± 15.6	25.1 ± 2.1	14.1 ± 2.5	131.5 ± 16.1	17.4 ± 4.5	9.2 ± 0.9
G4 ± sd	82.6 ± 18.7	24.9 ± 1.7	14.6 ± 2.6	118.1 ± 15.8	14.9 ± 5.6	9.1 ± 1.0
G5 ± sd	101.3 ± 19.0	26.1 ± 2.0	14.1 ± 2.7	133.8 ± 23.5	17.6 ± 5.6	9.0 ± 1.0

CL= culm length; PL= panicle length; FLBW= flag leaf blade width; and GL = grain length



G1: *O. glaberrima*; G2: *O. s. japonica*; G3: *O. s. indica* group 1; G4: *O. s. indica* group 2; G5: *O. s. indica* group 3; Adm: admixture *O. sativa*x *O. sativa* and Adm-GS: admixture *O. glaberrima*x *O. sativa*

Figure 4. Factorial analysis performed on 282 accessions from Burkina Faso and 3 checks based on 23 SSRs data projected on axis 1/3 showing the 5 groups inferred by Structure



BM, CC, HB and SO referred to samples collected in Boucle du Mouhoun, Cascades, Hauts-Bassins and Sud-Ouest regions, respectively

Figure 5. Regional subdivision of 36 *O. glaberrima* accessions of Burkina Faso rice collection

Besides, an unrooted neighbour joining tree was performed on the *O. glaberrima* accessions and separated the deep water (DW) accessions from lowland accessions. *O. glaberrima* samples from CC region were mostly from deep water ecology while the *O. glaberrima* from the three other regions were concentrated in lowland ecology (Figure 6).

Blue lines represent accessions from deep water ecology and green line accessions from lowland ecology. Capital letters B, C, H and S stand for Boucle du Mouhoun, Cascades, Hauts-Bassins and Sud-Ouest regions. CG14 and Tog5681 were used as checks varieties and were highlighted in red colour. *O. glaberrima* samples from Cascades region were mostly from deep water ecology while the *O. glaberrima* from the three other regions were concentrated in lowland ecology.

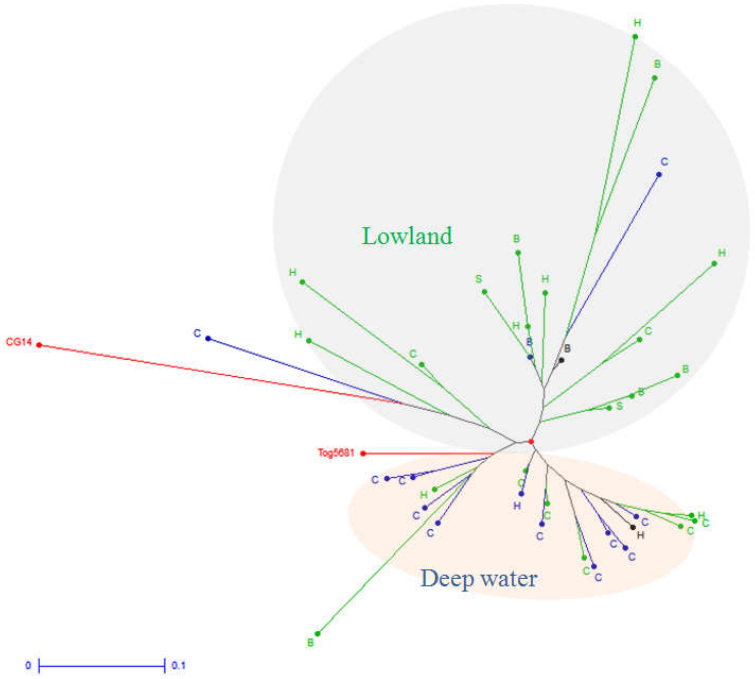


Figure 6. Unrooted neighbour-joining tree based on allelic data from 22 SSR loci among 39 *O. glaberrima* from Burkina Faso using the simple matching index

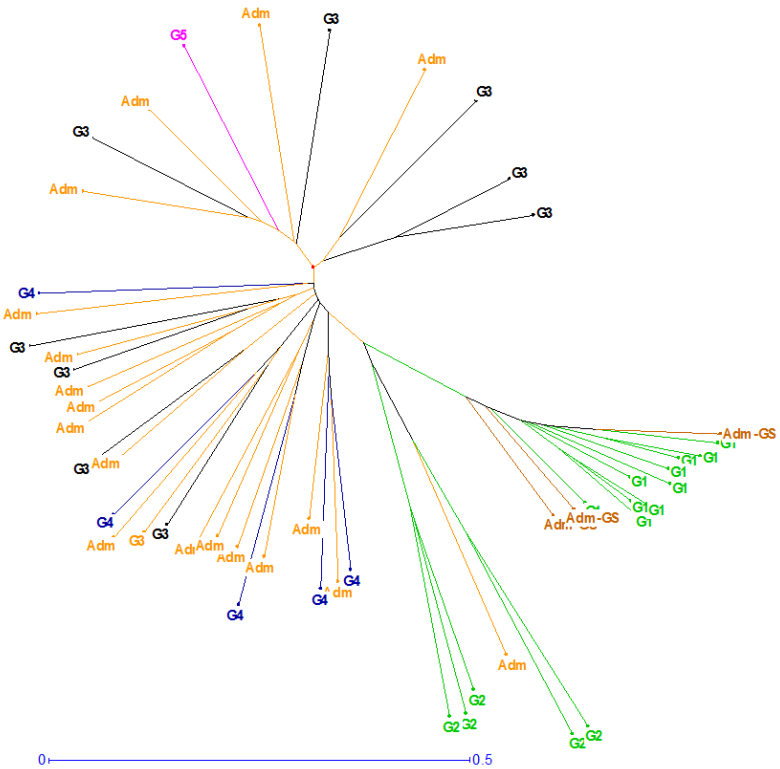


Figure 7. Neighbour-joining tree of 52 individuals of the core collection including all the groups identified by separately studying *O. glaberrima* and *O. sativa* based on 23 SSR data

G1: *O. glaberrima*; G2: *O. s. japonica*; G3: *O. s. indica* group 1; G4: *O. s. indica* group 2; G5: *O. s. indica* group 3; Adm: admixture *O. sativa*x *O. sativa* and Adm-GS: admixture *O. glaberrima*x *O. sativa*

Table 6. List of the accessions of the core collection and their respective groups

Accession	Group	Local name	Accession	Group	Local name
BM6A	G1	malotelemani	CC35	G3	Gnangnanhou
BM13B	G1	Sonksonke	CC37	G3	lassina
BM15A	G1	djoutchinmin	CC43	G3	dembele
BM34B	G1	lamizana junior	CC51	Adm	gnokpienkpao
CC2	G1	gnonre	CC70	G4	fitidjio
CC108	G1	tchiramahihou	CC111	G3	tiombie
CC109A	G1	gbarountchie	CC130	G3	maloba
CC114	Adm	lokofouo	CC131	Adm	kalonani
CC123A	Adm-GS	kamelekan	CC134	G4	were-were
CC149A	G1	malofiin	CC136	Adm	boriyo
HB9B	G1	melange	CC139	Adm	woloni
HB41	Adm-GS	boewo	CC156B	G3	malowile
HB46	Adm-GS	safati	SO2	G4	air-bouake
HB88	G1	Malodjalan	SO3	G2	mouimiougou
BM9	G4	badjimaloba	SO26	G3	marnii
BM11	G3	Kandeka	SO32	G3	maloba
BM13A	Adm	Bintoubalo	HB7	Adm	kankan-malo
BM15	G2	djoutchinmin	HB9	Adm	melange
BM24	G2	kiougayibu2	HB11	G4	X1
BM27A	Adm	melange	HB18B	Adm	dorado
CC1	Adm	dregbon	HB44	Adm	wereni
CC4A	G3	Toumihiennan	HB62	G2	zia
CC7	Adm	benkawili	HB69	Adm	kolognore
CC11	Adm	trouchaimin	HB77	Adm	Riz de banfora
CC17B	G5	djongolo	HB82	Adm	malble
CC31	G2	beregadougou	HB86	Adm	boanga

Core collection establishment

A core set with a limited allelic diversity reduction and a higher number of sample reductions was implemented separately for the *O. sativa* and *O. glaberrima* accessions, using the software package DARw in 6. A core set of 13 individuals was obtained for *O. glaberrima,* while 39 individuals were identified for *O. sativa*. The two were pooled together and a unique core of 52 accessions with a total of 182 alleles was established. A novel neighbour-joining tree with the 52 accessions was built. In the core collection, the sub-clusters obtained previously when investigating the entire collection were clearly represented (Figure 7). The alleles present in the core collection account for 89% of the alleles of the whole collection. The list of the accessions of the core collection are summarized in Table 6

DISCUSSION

The aim of molecular characterization of the 330 rice accessions was to study the population structure and the genetic diversity amongst and between the accessions of the collection, in order to establish a rice core collection of Burkina Faso landraces. The 23 SSR markers used for the diversity analysis were found to be polymorphic. The allele numbers per locus varied from 2 to 26, with a mean allele number of 9.4 alleles per locus. Barry *et al*. (2007), in their study of diversity of rice in maritime Guinea using 11 SSR markers, found a mean allele number value of 12, which was higher than ours. This could be due to low number of loci used in their studies. However, the main reason could be the diversity richness in Maritime Guinea rice collection compared to that of Burkina Faso rice collection. Guinea is believed to be the second centre of diversification of *O. glaberrima* and a port entry of *O. sativa* in West Africa (Portères, 1950). Markers based clustering of the 282 accessions was consistent with phenotypical traits that differentiate *O. glaberrima* from *O. sativa*. The G1 group included all the *O. glaberrima* of the collection confirming the larger genetic diversity of *O. sativa* in the Burkina Faso rice cultivation. Decades before, Chang (1984) had raised the alarm of the continually replacement of *O. glaberrima* by *O. sativa* in Africa. The replacement of the autochtonal rice (*O. glaberrima*) by modern varieties may endanger the conservation of local genetic resources (Barry *et al*., 2007). The high rate of exchange of *O. sativa* between farmers highlighted the necessity of development of strategies of conservation of *O. glaberrimain situ* and *ex situ*. Though endowed with smaller genetic diversity compared to *O. sativa*, *O. glaberrima* (4) the non-negligible allelic diversity of four alleles per loci, the same amount reported by Barry *et al*. (2007). Likewise, when considered alone *O. glaberrima* accessions split in two sub-groups according to their ecologies: flooded and lowland as it was reported also by Bezançon (1993). Indeed, the samples from Boucle du Mouhoun, Cascades and HautsBassins were from lowland ecology while those from Cascades region were in majority deep water *O. glaberrima*. This subdivision was reinforced by phenotypical traits highlighting the differences between the two groups. Cascades region harbour many traditional basin areas and this part of the country had rainfall exceeding 1,000 mm per year. Therefore, this region appeared to be the fief of flooded *O. glaberrima* with longer cycle and long culm. Conversely, lowland *O. glaberrima* with shorter cycle and short culm were encountered in Boucle du Mouhoun, Hauts-Bassins and Sud-Ouest regions. Accessions CC123A, HB41 and HB46 are believed to be an admixture of *O. sativa* and *O. glaberrima*. Admixed varieties between *O. sativa* and *O. glaberrima* were also reported by Barry *et al*. (2007) and Semon *et al*. (2005). Admixed varieties are important for breeding purposes. They can help to overcome crossing barriers between species. The organisation of genetic diversity revealed by the genotypic data was in agreement, to some extent, with the phenotypical traits. Genotypic structure took into account the differences between the ecosystems and growth duration in *O. glaberrima* and *O. sativaspecies*. The phenotypical traits highlighted the differences between groups in detail. However, analysis of the population structure using molecular markers pointed out three groups (G3; G4 and G5) in *O. s. indica* and 48 admixed accessions *O. sativa*x *O. sativa*. The ecology, the cycle and the culm length were characteristics splitting the three groups. The

plants from G5 with long cycle and long culm were cropped in deep water ecology. They must be varieties that resist submersion and adapted to flooded environments. Nevertheless, they might be very sensitive to water shortage. The early maturing *O. s. indica* with short culm and grown in lowland areas were gathered in group G4. They must be improved *O. s. indica* developed decade ago through research network. The late maturing group G3 were traditional *O. s. indica* grown in lowland areas. The abundance of admixed accessions *O. sativa* by *O. sativa* testified the importance of intraspecific breeding of this species since decade ago. Sié *et al*. (1999) using enzymatic polymorphism had confirmed the presence of intermediate form within the *O. sativa* type and had identified two varieties belonging to "*Aus*" cultivars.

Eleven accessions of the collection were belonging to *O. s. japonica* subspecies. The *japonica* cultivar accounted for only 3.3 % of the collection. Sié *et al*. (1999) noticed also the low frequency of var. *japonica* cultivar in their Burkina Fasorice collection. In reality, upland rice production is not very common in Burkina Faso. All the accessions of G2 had a private allele showing their singularity. BM15 is named "Djoutchèmè". In the rice collection of Maritime Guinea (Barry *et al*., 2007), a variety named "DjouKémé" was found to be intermediate between *O. s. indica* and *O. s. japonica*. The two accessions could be the same, given the homonymy and the atypical character of BM15 shown by our molecular study. These atypical accessions are precious resources for rice breeding. The phenotypic and genotypic data were complementary in describing the genetic diversity prevailing in the Burkina Faso rice collection. This study showed the efficiency of the Neighbour Joining Method, as described by Saitou and Nei(1987) and Takahashi and Nei(2000) in the evaluation of genetic diversity. The core collection of 13 *O. glaberrima* and 39 *O. sativa* spp reflects the sub-structures identified in *O. glaberrima* and in *O. sativa* spp, covering the extent of diversity of rice in Burkina Faso. The establishment of separate *O. glaberrima* and *O. sativa* core collections was followed to avoid reducing the representativeness of *O. glaberrima* in the core collection because *O. sativa* showed more allelic diversity than *O. glaberrima*. This sampling was effective in retaining89% of the alleles of the entire collection and accounted for 16% of the total collection. This size of the core is acceptable and represents the maximum genetic diversity of the whole collection. In several crop species the sampling percentage of core collections account for 10-30% of the entire resource (Zhang *et al*., 2009). The genetic diversity present in the entire collection, as well as in the core collection, deserves special attention, to be preserved *in situ* and *ex situ*. Such small size of core is manageable by NARS of developing countries and can be easily exploited by breeders and farmers for rice genetic improvement. The accessions excluded from this collection, plus a set of the core collection, were deposited in the AfricaRicegenebank for long term storage.

Conclusion

The molecular evaluation of the 282 unique samples out of the 330 samples, confirmed the phenotypical subdivision of cultivated rice species cropped in Burkina Faso. The *O. glaberrima*, *O. s. japonica* and *O. s. indica* were clearly clustered. The structure analysis, revealed the presence of three admixed varieties between *O. glaberrima* and *O. sativa* and 48 admixtures *O. sativa*x *O. sativa* in the collection. The core collection of 52 accessions including all the groups and subgroups identified in the collection account for 89% of the alleles of the whole collection. This core collection is a valuable materials for rice breeders for varieties development.

Acknowledgments

This research was financially supported by USAID and the Generation Challenge Program (GCP), which provided to H. Kam fellowship (G4009.02.01) for the genotyping of the accessions at CIRAD – Montpellier.

Conflicts of interest: The authors declare no conflict of interest

REFERENCES

Barry M.B., Pham J.L., Noyer J.L., Billot C., Courtois B., Ahmadi N. 2007. Genetic diversity of the two cultivated rice species (O. sativa & O. glaberrima) in Maritime Guinea. Evidence for interspecific recombination. Euphytica 154:127–137.

Bezançon G. 1993. Le riz cultivé d'origine africaine *Oryza glaberrima* Steud. et les formes sauvages et adventices apparentées : diversité, relations génétiques et domestication, Thèse de Doctorat, Université de Paris-Sud, Paris.

Bioversity-International, IRRI, WARDA. 2007. Descriptors for wild and cultivated rice (*Oryza* spp.). , Rome, Italy; Los Baños, Philippines; Cotonou, Benin.

Botstein D., White R.L., Skolnick M., Davis R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*., 32:314-31.

Brown A. 1989. Core collections: A practical approach to genetic resources management. *Genome*, 31:818 - 824.

Chang T.T. 1984. Conservation of Rice Genetic Resources: Luxury or Necessity?. *Science*, 224:251-256.

Chévalier A. 1937. Sur les riz africains du groupe *Oryza glaberrima*. *Rev. Bot. Appl*., 190:413-418.

Diwan N., McIntosh M.S., Bauchan G.R. 1995. Methods of developing a core collection of annual *Medicago* species. *Theor. Appl. Genet*., 90:755-761.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 1-10.

Falush D., Stephens M., Pritchard J.K. 2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics,* 164:1567-1587.

Garris A.J., Tai T.H., Coburn J., Kresovich S., McCouch S. 2005. Genetic Structure and Diversity in Oryza sativa L. Genetics 169:1631-1638. DOI:10.1534/genetics.104. 035642.

Hubisz M.J., Falush D., Stephens M., Pritchard J.K. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9:1322-1332.

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

Liu F., von Bothmer R., Salomon B. 1999. Genetic diversity among East Asian accessions of the barley core collection as revealed by six isozyme loci. *Theor. Appl. Genet*., 98:1226-1233. DOI: 10.1007/s001220051188.

Liu K., Muse S.V. 2005. PowerMarker: Integrated analysis environment for genetic marker data. *Bioinformatics,* 21:2128-2129.

Nei M. 1987. Molecular evolutionary genetics Columbia University Press, New York.

Peakall R., Smouse P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6:288-295.

Perrier X., Flori A., Bonnot F. 2003. Data analysis methods, in: P. Hamon, *et al*. (Eds.), Genetic diversity of cultivated tropical plants, Enfield, Science Publishers, Montpellier. pp. 43 - 76.

Perrier X., Jacquemoud-Collet J.P. 2006. DARwin software, http://darwin.cirad.fr/darwin.

Portères R. 1950. Vieilles agricultures de l'Afrique intertropicale. Centres d'origine et de diversification variétale primaire et berceaux de l'agriculture antérieurs au XVIième siècle. L'Agron. Trop. V:489-507.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155:945-959.

Risterucci A.M., Grivet L., N'Goran J.A.K., Pieretti I., Flament M.H., Lanaud C. 2000. A high-density linkage map of Theobroma cacao L. *Theor Appl Genet*., 101:948-955.

Roy M.S., Geffen E., Smith D., Wayne R.K. 1996. Molecular genetics of pre-1940 red wolves. *Conserv. Biol*., 10:1413-1424.

Saitou N., Nei M. 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic *Trees. Mol. Biol. Evol.*, 4:406-425.

Semon M., Nielsen R., Jones M.P., McCouch S.R. 2005. The Population Structure of African Cultivated Rice *Oryza glaberrima* (Steud.): Evidence for Elevated Levels of Linkage Disequilibrium Caused by Admixture with *O. sativa* and Ecological Adaptation. *Genetics*, 169:1639-1647.

Sié M. 1984. Prospection des variétés traditionnelles du riz au Burkina Faso., INERA/IBPGR. pp. 10.

Sié M. 1991. Prospection et évaluation génétique des variétés traditionnelles de riz (*Oryza sativa* L. et *O. glaberrima* Steud) du Burkina Faso, Faculté des Sciences et Techniques, Thèse de Doctorat, Université Nationale de Côte d'Ivoire, Abidjan. pp. 118.

Sié M., Ghesquiere A., Miezan K.M. 1999. Structure genetique des variétés traditionnelles de riz (*Oryza* sp.) du Burkina Faso. Agronomie Africaine 11:57-71.

Sokal R.R., Michener C.D. 1958. A statistical method for evaluating systematic relationships. *Univ Kans Sci Bull*., 38:1409-1438.

Takahashi K., Nei M. 2000. Efficiencies of Fast Algorithms of Phylogenetic Inference Under the Criteria of Maximum Parsimony, Minimum Evolution, and Maximum Likelihood When a Large Number of Sequences Are Used. *Mol. Biol. Evol*., 17:1251-1258.

Zhang C.-y., Chen X.-s., Zhang Y.-m., Yuan Z.-h., Liu Z.-c., Wang Y.-l., Lin Q. 2009. Method of constructing core collection for *Malus sieversii* in Xinjiang, China using molecular markers. Agri. Sci. China 8:276-284. DOI: Doi: 10.1016/s1671-2927(08)60210-2.

****