Electronic Journal of Biotechnology 30 (2017) 48–57

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

Research article

Molecular characterization and genetic diversity of different genotypes of *Oryza sativa* and *Oryza glaberrima*

Caijin Chen [a,1], Wenchuang He [a,1], Tondi Yacouba Nassirou [a], Athanase Nsabiyumva [a], Xilong Dong [a], Yawo Mawunyo Nevame Adedze [b], Deming Jin [a,*]

[a] MOA Key Laboratory of Crop Ecophysiology and Farming System in the Middle Reaches of the Yangtze River, College of Plant Science and Technology, Huazhong Agricultural University, 430070 Wuhan, China
[b] State Key Laboratory of Rice Biology, China National Rice Research Institute, 310006 Hangzhou, China

ARTICLE INFO

Article history:
Received 16 February 2017
Accepted 9 August 2017
Available online 16 August 2017

Keywords:
African rice
Asian rice
Fingerprinting
Food security
Genetic relationship
Microsatellite markers
Molecular profiling
Phylogenetic tree
Polymorphic alleles
Rice breeding
SSR

ABSTRACT

Background: Availability of related rice species is critical for rice breeding and improvement. Two distinct species of domesticated rice exist in the genus Oryza: *Oryza sativa* (Asian rice) and *Oryza glaberrima* (African rice). New rice for Africa (NERICA) is derived from interspecific crosses between these two species. Molecular profiling of these germplasms is important for both genetics and breeding studies. We used 30 polymorphic SSR markers to assess the genetic diversity and molecular fingerprints of 53 rice genotypes of *O. sativa*, *O. glaberrima*, and NERICA.

Results: In total, 180 alleles were detected. Average polymorphism information content and Shannon's information index were 0.638 and 1.390, respectively. Population structure and neighbor-joining phylogenetic tree revealed that 53 genotypes grouped into three distinct subpopulations conforming to the original three groups, except three varieties (IR66417, WAB450-4, MZCD74), and that NERICA showed a smaller genetic distance from *O. sativa* genotypes (0.774) than from *O. glaberrima* genotypes (0.889). A molecular fingerprint map of the 53 accessions was constructed with a novel encoding method based on the SSR polymorphic alleles. Ten specific SSR markers displayed different allelic profiles between the *O. glaberrima* and *O. sativa* genotypes.

Conclusions: Genetic diversity studies revealed that 50 rice types were clustered into different subpopulations whereas three genotypes were admixtures. Molecular fingerprinting and 10 specific markers were obtained to identify the 53 rice genotypes. These results can facilitate the potential utilization of sibling species in rice breeding and molecular classification of *O. sativa* and *O. glaberrima* germplasms.

1. Introduction

Rice is one of the most important crops in the world; and the *Oryza* germplasm serves as the essential resource for rice breeding and contributes significantly to global food security [1]. There are two distinct cultivated species in the genus *Oryza* that both belong to the AA genome with diploid level (2n = 24), *O. sativa* (Asian rice) and *O. glaberrima* (African rice). They were independently domesticated from divergent progenitor wild species in different geographic locations, South Asia and West Africa, respectively [2]. The *O. glaberrima* has many useful traits such as resistance to pests and diseases, and drought tolerance but has lower grain yield potential than *O. sativa* [3,4,5]. Recently, breeders have turned their attention to select elite Asian rice parents because the widespread adoption of similar but improved varieties has decreased genetic diversity of rice gene pool. This might be a major contributing factor for the increased vulnerability to various biotic/abiotic stresses and the yield plateau witnessed in rice production [6]. To overcome these challenges, introduction of new favorable genetic material from a closely related species within the genus *Oryza* is considered a promising approach [7]. African rice could be an important donor germplasm to enrich the rice genetic pool [8,9]. By overcoming the reproductive barriers that existed between the two species, introgression lines such as New Rice for Africa (NERICA) have been developed to combine the superior traits of *O. glaberrima* and *O. sativa,* and bridge the genetic gap between the two distinct species [10,11]. Furthermore, utilization of interspecific heterosis *via* partial interspecific hybrid rice between *O.*

* Corresponding author.
E-mail address: djin@mail.hzau.edu.cn (D. Jin).
[1] These authors contributed equally to this work.
Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

http://dx.doi.org/10.1016/j.ejbt.2017.08.001

sativa and introgression lines carrying *O. glaberrima* genes could be a promising novel approach to raise the grain yield potential [12].

Several previous researches have revealed the genetic diversity of different rice germplasms. An assessment of the molecular diversity of 79 *O. glaberrima* germplasms from Mali (West Africa) revealed that the populations from different locations were highly differentiated [13]. Genetic diversity and population structure were investigated using 93 simple sequence repeats (SSR) markers for 198 accessions of *O. glaberrima* collected from 12 different countries in West Africa. Genetic evidence indicated that 67% of *O. glaberrima* accessions carry some level of admixture with *O. sativa* and that natural interspecific outcrossing might occur because *O. glaberrima* was often grown in combination with *O. sativa* in West Africa [14]. Similar results were reported by Barry et al. [15] who found a close genetic relationship between *O. sativa* and some *O. glaberrima* accessions using 11 SSR markers and 26 morpho-physiological descriptors [15]. Similarly, reports also describe the genetic diversity of *O. sativa*. Two subgroups including *indica* and *japonica* as well as six sub-subgroups were found within a primary *O. sativa* core collection [16]. Eight subpopulations were found to correspond to major geographic regions among 103 *O. sativa* accessions studied [17]. However, to our knowledge, the comparison of genetic diversity of different cultivated species of rice genotypes has not been reported. A better understanding of the extent and distribution of genetic diversity within and between different genotype groups is essential not only to assist plant breeders in the selection of parents but also to provide a more rational basis for expanding the gene pool and for identifying materials that harbor alleles valuable for plant improvement.

Based on the phenotypic and genotypic data, some rice core collections were evaluated for germplasm with similar genetic diversity as the entire rice collection; this can be an effective and convenient tool for rice breeders. However, it is difficult to unambiguously identify cultivars from other groups using conventional morphological characteristics owing to effects of environmental factors [18]. Fingerprinting with molecular markers allows a precise, objective, and rapid cultivar identification, which has been proven to be an efficient tool for crop germplasm characterization, collection, and management [19]. A variety of molecular markers can be used to evaluate the genetic diversity and establish the fingerprint of rice genotypes, such as SSRs and single nucleotide polymorphisms (SNPs). Until now, SSR markers have been widely used for assessment of genetic diversity and establishment of unique fingerprint owing to their abundance, codominant inheritance, high polymorphism, reproducibility, ease of assay by polymerase chain reaction (PCR), and relatively low cost [20,21]. Recently, SNPs have received increased attention because they occur at a much higher frequency in the genome than SSRs. However, most SNPs are biallelic; thus, an SNP marker has less information content than an SSR marker [22]. Furthermore, SSR markers have their own advantages as compared to SNP markers for population genetics analysis [23] and are still used widely in the construction of molecular fingerprinting databases. A molecular fingerprinting database of 49 rice cultivars was constructed using 24 SSR markers [24]. Unique DNA profiles of conventional *japonica* rice from Taihu Lake area were established based on 24 SSR markers [25]. SSR markers have been used in many crop species for cultivar identification, such as wheat [26], maize [27], bean [28], tomato [29,30], and other crops. Additionally, the fingerprint can be used for the protection of plant genetic resources.

The objectives of this study were to (1) investigate the genetic diversity, variability and molecular phylogeny; and (2) establish an effective and low-cost encoding method for the molecular fingerprinting of 53 rice genotypes including *O. glaberrima, O. sativa*, and NERICA, using SSR markers. This study will enhance our understanding of genetic diversity among different genotypes of *O. sativa* and *O. glaberrima* and facilitate the use of diverse germplasm in rice breeding.

2. Materials and methods

2.1. Experiment materials

A total of 53 rice genotypes were collected for this study, including 18 African rice accessions, 10 NERICA varieties, 23 Asian rice varieties, and two tropical weedy rice accessions that were collected from six different countries (Table S1). Twenty-five *O. sativa* genotypes in the germplasm pools, as well as their known ancestors, were distributed in nine Southeast Asian countries such as China, India, and Indonesia (Fig. 1; Table S1).

2.2. Pedigree analysis

The origin and pedigree information of rice genotypes were investigated using the International Rice Information System database (http://irri.org/tools-and-databases/international-rice-information-system) and the China Rice Data Center (http://www.ricedata.cn/variety/) resources. Pedigree plotting was performed using Pedigraph 2.2 software [31].

2.3. DNA extraction and SSR analysis

Genomic DNA was extracted from young and healthy leaves of plants, using the cetyltrimethylammonium bromide method [32] with minor modifications. Quality of the isolated DNA was checked on 1% agarose gel and quantity was determined using ND-1000 spectrophotometer (NanoDrop Technologies, USA).

A total of 118 SSR primers were used to evaluate the polymorphism of 12 genotypes comprising four *O. glaberrima*, four *O. sativa*, and four NERICA genotypes. To screen the 53 genotypes, 30 SSR primers were selected based on their polymorphism and distribution on the chromosomes (Table S2).

The polymerase chain reaction (PCR) was performed in 20 μL volume, containing 2 μL (50 ng μL^{-1}) DNA, 0.2 μL (5 U μL^{-1}) *Taq* polymerase, 2 μL 10X PCR buffer, 1.2 μL (10 mM) $MgCl_2$, 0.4 μL dNTP, 2 μL primer pairs, and 12.2 μL sterile double distilled water. The PCR cycling parameters were as follows: initial preheating and initial denaturation at 94°C for 4 min, followed by 10 cycles of denaturing at 94°C for 30 s, annealing at 55°C to 65°C for 30 s, decreasing 1°C per cycle, and extension at 72°C for 45 s. The last 22 cycles were at 94°C for 30 s, 55°C for 30 s and 72°C for 45 s, followed by a final extension at 72°C for 10 min. The SSR products were separated on a denaturing 6% polyacrylamide gel and visualized by silver staining. The electrophoretic bands were scored as 1 (present) or 0 (absent) to form a raw data matrix for further analysis.

2.4. Data analysis

POPGENE 3.2 software was used to calculate the number of effective alleles [33], expected heterozygosity [34], Shannon's information index [35], and allele frequency. Allelic polymorphic information content (PIC) was calculated using the following formula:

$$PIC = 1 - \sum_{i=1}^{n} P_i^2$$

where n is the total number of alleles detected for a given marker locus and P_i is the frequency of the ith allele in the set of genotypes investigated [36].

The genetic distance (GD) was calculated using the following formula:

$$GD = 1 - GS$$

where GS was calculated using the NTSYS-pc ver. 2.1e software with option of DICE coefficient.

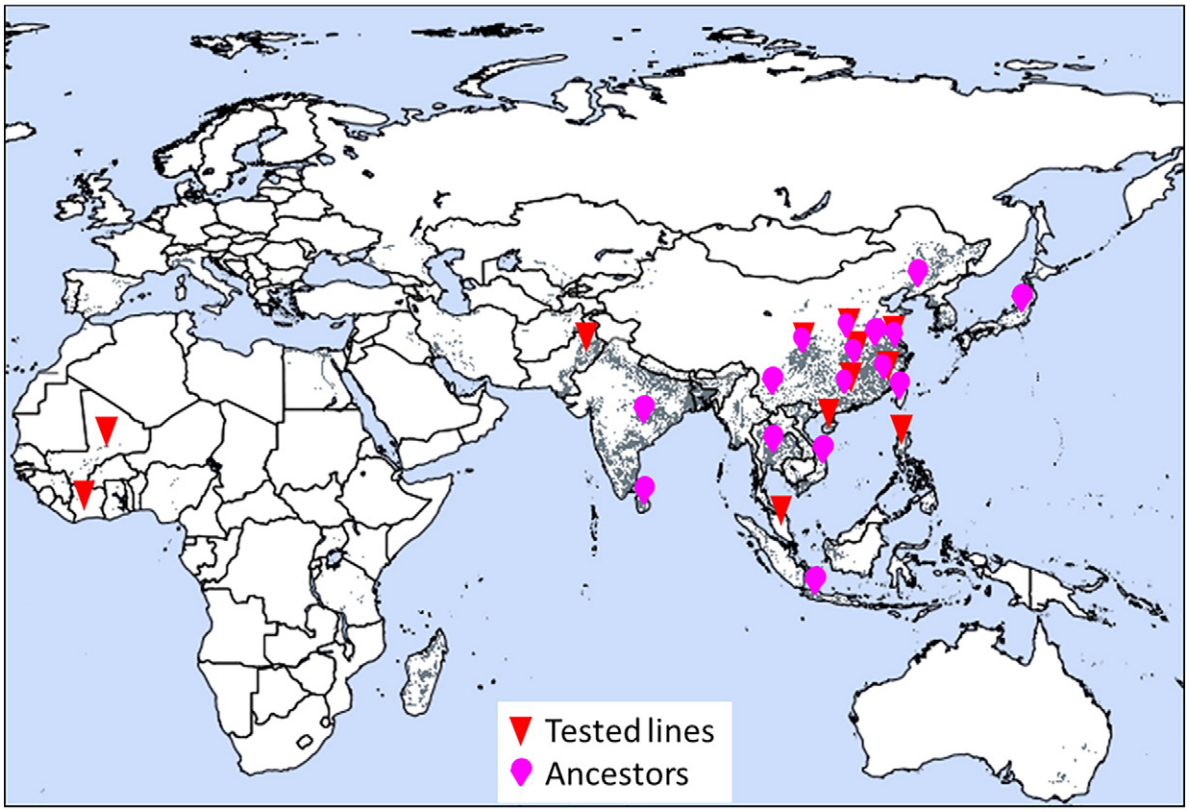

Fig. 1. Distribution of origins of rice genotypes used in this study.

Principal component analysis (PCA) was performed using the NTSYS-pc options D-center and Eigen algorithms. The STRUCTURE software program was used to infer the structure of the population [37]. Ten independent runs were performed with the number of sub-populations (*K*) ranging from 1 to 10 with the admixture model, a burn-in period length of 10,000, a run length of 10,000, and 10 replications for each *K* value. Both LnP(D) and Evanno's Δ*K* methods were used to estimate the number of sub-populations. The run of the estimated numbers of sub-populations showing the maximum likelihood was used to assign genotypes with membership probability ≥0.7 to sub-populations. Genotypes with membership probability <0.7 were assigned to an admixed group. Genetic differentiation in the population was estimated by constructing a neighbor-joining (NJ) tree based on genetic distance in MEGA 5.0. The lengths of amplified fragments were estimated using the Quality One software.

3. Results

3.1. SSR diversity

A total of 180 polymorphic bands were produced from 30 SSR primers (average of 6 bands per primer). The number of polymorphic bands for each primer showed large variation ranging from three (RM60 and RM55) to 11 (RM163) as shown in Table 1. The effective number of alleles ranged from 1.391 to 8.011 (average of 3.766) and the Shannon information index varied from 0.615 to 2.203 (with an average of 1.390). PIC value provides an estimate of discriminatory power of the markers by taking into account not only the number of alleles per locus but also their relative frequencies in the population studied [38]. In this study, PIC values ranged from 0.268 for RM6543 to 0.862 for RM163 with an overall mean of 0.638, indicating a diverse distribution of polymorphic information throughout the bands.

Table 1
Genetic diversity analysis based on 30 SSR markers.

Marker	Na	Ne	I	PIC
RM312	8	4.812	1.776	0.766
RM250	9	4.047	1.734	0.729
RM60	3	2.788	1.062	0.569
RM255	6	2.931	1.343	0.622
RM163	11	8.011	2.203	0.862
RM1370	10	5.729	1.948	0.804
RM3743	5	2.293	1.002	0.484
RM223	6	4.392	1.569	0.736
RM6543	5	1.391	0.615	0.268
RM6364	9	5.903	1.947	0.810
RM202	5	4.378	1.534	0.734
RM1227	6	3.271	1.325	0.638
RM220	6	2.857	1.242	0.596
RM225	5	2.204	1.093	0.511
RM224	6	4.613	1.629	0.750
RM251	9	5.121	1.856	0.782
RM287	5	4.643	1.572	0.750
RM318	6	3.523	1.413	0.669
RM55	3	2.200	0.885	0.458
RM218	8	5.2115	1.797	0.782
RM197	4	1.9852	0.812	0.406
RM5814	4	1.748	0.815	0.394
RM1369	9	7.706	2.116	0.856
RM3183	7	3.318	1.416	0.652
RM3609	5	4.080	1.492	0.715
RM6396	4	2.806	1.115	0.573
RM259	4	2.575	1.070	0.546
RM1302	4	3.153	1.243	0.624
RM6306	4	3.218	1.217	0.625
RM327	4	2.061	0.831	0.417
Mean	6	3.766	1.390	0.638

Na = Observed number of alleles.
Ne = Effective number of alleles.
I = Shannon's Information index.
PIC = Polymorphism information content.

Table 2
Genetic diversity analysis in different types of rice based on 30 SSR markers.

	Type		
	O. sativa	*O. glaberrima*	*O. sativa* × *O. glaberrima*
Sample number	23	18	10
Number of alleles	3.600	2.370	2.830
Effective number of alleles	2.212	1.780	1.930
Shannon's Information index	0.771	0.565	0.700
Expected heterozygosity	0.413	0.342	0.411
Polymorphism information content	0.367	0.288	0.351

Expected heterozygosity was computed using Levene's (1949) method [38].

3.2. Genetic diversity of different rice genotypes

Genetic diversity of three different rice groups, Asian rice, NERICA, and African rice are shown in Table 2. The value of average number of effective alleles of Asian rice, NERICA, and African rice were 2.212, 1.930, and 1.780, respectively. The number of alleles can be affected by numbers of tested samples given that the sample number of Asian rice group was higher in this study. Asian rice group and NERICA group had higher averages for Shannon's information index of 0.771 and 0.700, respectively. African rice had a lower average Shannon's Information index of 0.565. The group consisting of Asian rice and NERICA had similar values for expected heterozygosity, i.e., 0.413 and 0.411, respectively. Simultaneously, the PIC also revealed similar results for the group consisting of *O. sativa* (0.367) and NERICA (0.351). The Asian rice group showed the most diverse variation, probably due to the wide and complex genes flowing between germplasms from different geography sources including from wild species (e.g., *Oryza nivara*) considering the origins and pedigrees of the 23 Asian rice varieties (Fig. 1; Fig. S1).

Pairwise genetic distances of the 53 rice genotypes showed consistent results with that of diversity coefficients (Fig. 2). Pairwise genetic distances within Asian rice, NERICA, and African rice groups ranged from 0.097–0.735, 0.097–0.739, and 0.065–0.548, with an average of 0.421, 0.419 and 0.331, respectively. Furthermore, Asian rice showed a higher average distance with African rice group (0.886) than with the NERICA group (0.744). Although NERICA was developed from interspecific crosses between Asian and African rice, it was interesting to note that it showed relative genetic proximity to Asian rice group (0.744) than with the African rice group (0.889). Additionally, weedy rice MZCD73 also showed relatively closer distance with Asian rice group (0.447) whereas MZCD74 showed greater distances from all three groups (0.849, 0.844 and 0.760, respectively).

3.3. Population structure and molecular phylogeny

The NJ phylogenetic tree (Fig. 3a) and population structure (Fig. 3b) were constructed based on the 30 SSR markers fingerprinting data to analyze the molecular phylogeny of 53 genotypes. Two sub-populations were detected when $K = 2$ according to the STRUCTURE results (Fig. 3b). Eighteen African rice accessions were assigned into a separate sub-population, whereas the remaining 10 NERICA, two weedy rice, and 23 Asian rice genotypes grouped into another sub-population. However, the optimum population structure inferred using the admixture model approach in STRUCTURE was subdivided into three sub-populations according to both LnP(D) and Evanno's ΔK method ($K = 3$; Fig. 4). With membership probabilities ≥0.7, eighteen

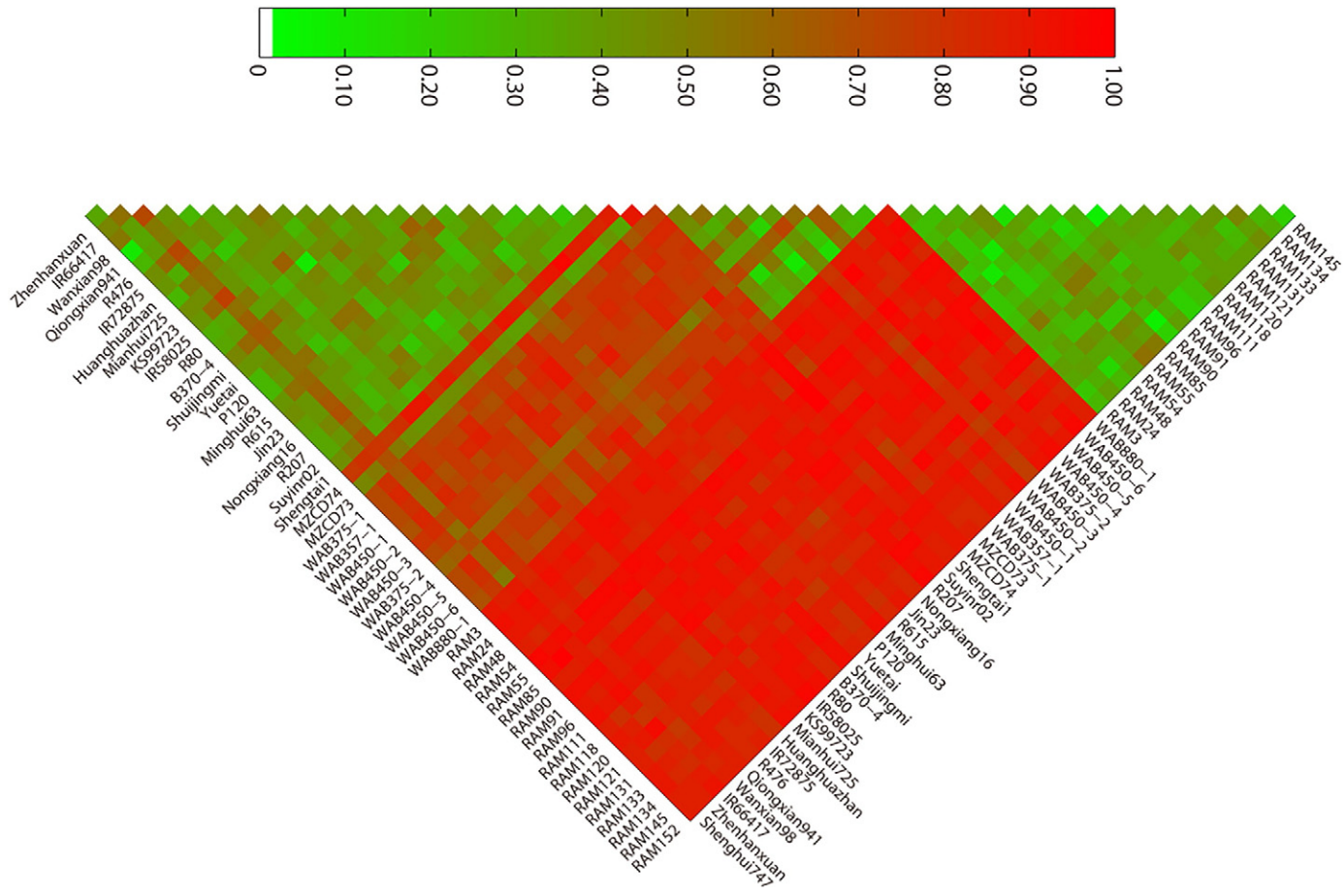

Fig. 2. Genetic distances between different rice genotypes based on 30 SSR markers.

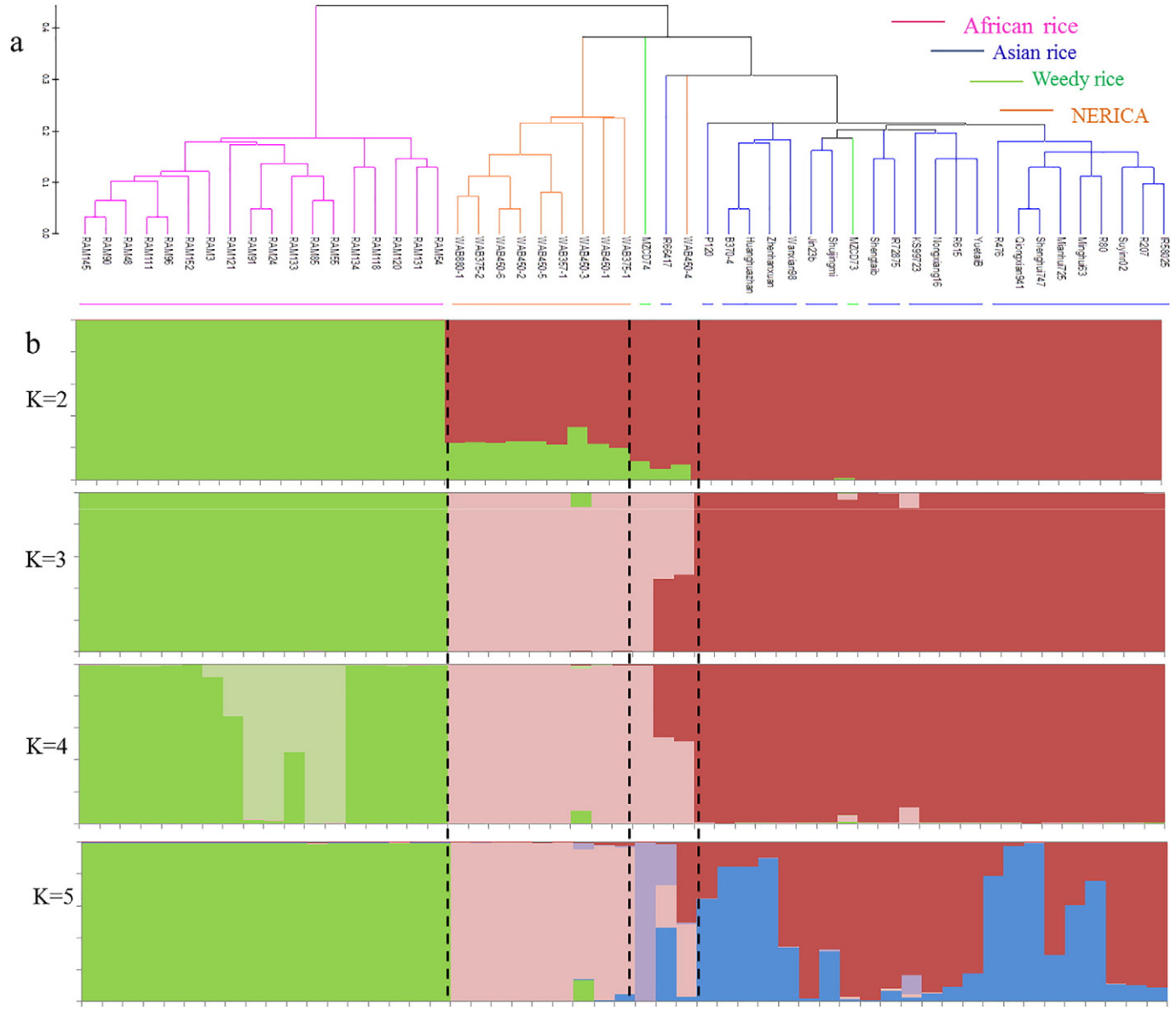

Fig. 3. Genetic relationship among the 53 rice accessions. (a) The neighbor-joining phylogenetic tree of the rice genotypes based on 30 SSR markers. The bars indicate African rice (Red), NERICA (Orange), Asian rice (Blue), and weedy rice (Green) lines. (b) Model-based clustering analysis with different numbers of clusters (K = 2, 3, 4, and 5). The y-axis quantifies cluster membership, and the x-axis lists the different accessions. The orders of these accessions on the x-axis are consistent with those for the neighbor-joining tree.

O. glaberrima accessions, nine of 10 NERICA varieties, and 23 of 25 *O. sativa* varieties conformed to their original grouping, except three accessions IR66417, WAB450-4, and MZCD74 ($K = 3$; Fig. 3b).

The NJ phylogenetic tree (Fig. 3a) based on genetic distance further supported the classification of STRUCTURE results. The *O. sativa* accessions and NERICA varieties were assigned to one cluster at a coefficient value of 0.4, whereas *O. glaberrima* accessions were allocated to another clade in the NJ tree. However, with a few exceptions, the three model-based groups clearly separated in the NJ phylogenetic tree at the coefficient value of 0.34. The NERICA accession WAB450-4, the Asian rice cultivar IR66417, and the weedy rice accession MZCD74 were assigned to three separate clades.

Principal component analysis was carried out to determine the genetic variation and components among the 53 genotypes (Fig. 5). This result showed that the first two principal components could explain 30.94% and 14.74% of the genetic variation. The PCA analysis showed a clear distinction between the genotypes in *O. glaberrima, O. sativa*, and NERICA groups, although some discrepancies between NERICA and *O. sativa* groups are expected owing to shared ancestral variation and historical gene flow among these very closely related genotypes. For instance, both MZCD74 and IR66417 were collected from South Asia and probably shared the same ancestry (IRAT216) with one of the *O. sativa* donor(s) used to develop WAB450-4.

3.4. SSR markers for identifying the genotypes of O. glaberrima and O. sativa

The SSR polymorphisms amplified using 30 SSR markers showed a rich diversity among the accessions which allowed the SSR markers to be used to identify the different types of rice genotype and classify rice accessions. Ten of the 30 specific markers, RM3743, RM250, RM1227, RM251, RM197, RM1369, RM3609, RM6396, RM6306, and RM1302, that amplified the different alleles in the *O. glaberrima* and *O. sativa* were selected to distinguish *O. glaberrima* and *O. sativa* groups. Four electrophoresis gels of PCR products amplified using the SSR primers (RM3743, RM197, RM6306 and RM1227) are presented in Fig. 6. Only RM1227 could be used to identify NERICA genotypes from *O. glaberrima* and *O. sativa* groups because NERICA rice was derived from interspecific crosses between *O. glaberrima* and *O. sativa*.

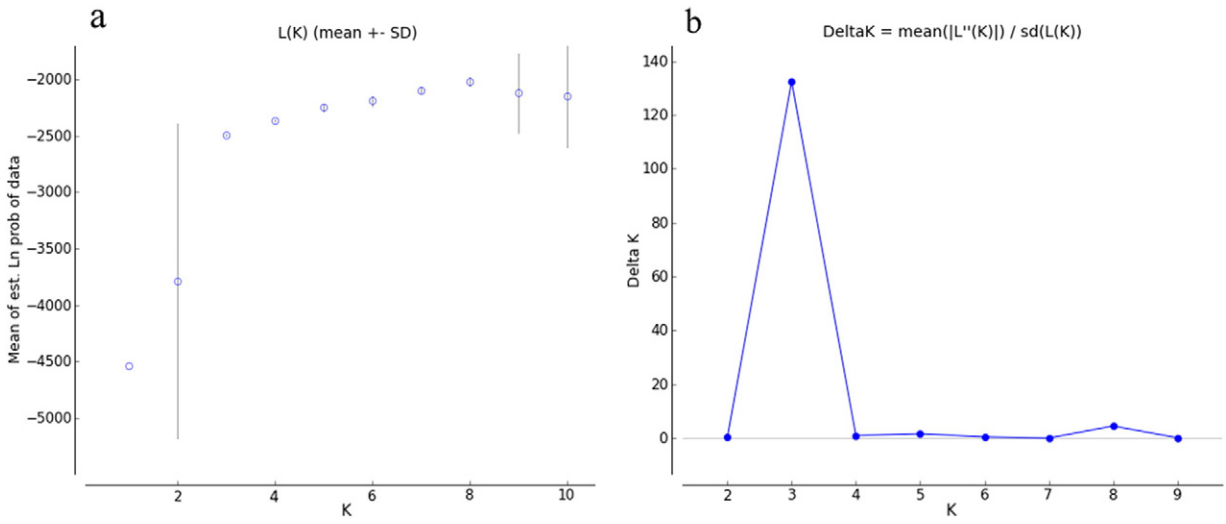

Fig. 4. The L(K) and ΔK values with different K-value calculated by STRUCTURE Harvester. (a) K versus mean−−in likelihood. (b) Magnitude of ΔK as a function of K. Note the steady increase in likelihood from K = 1 to K = 10, and the relatively highest value of ΔK, with peaks at K = 3.

3.5. DNA fingerprinting for the 53 rice cultivars

A DNA fingerprinting database of the 53 rice cultivars was successfully constructed with 13 selected SSR markers, using 5 selected polymorphic bands for each genotype (Table 3). Thirteen selected SSR markers were labeled A to M, and each band was coded with its marker label and the estimated amplified fragment length. The 5-code fingerprinting for each genotype comprised five specific bands, of which the first code is used to distinguish the three groups of *O. sativa*, *O. glaberrima*, and NERICA rice whereas the remaining four codes were selected based on the SSR polymorphism for each tested genotype. Fingerprints of Shenghui747, RAM3, and WAB375-1 were L150D137E151F129J153, L167A115B189E163J159, and L174A100C158F148J159, respectively, where L is the primer RM1227, as this primer can be used to distinguish *O. sativa*, *O. glaberrima*, and NERICA groups; 150 refers to the length of the amplified band for *O. sativa*; 167 refers to the length of the amplified band for the *O. glaberrima*, whereas 174 refers to the length of the amplified band for NERICA; D is the primer RM255, 137 refers to a specific 137-bp amplified band of cultivar Shenghui747, and so on.

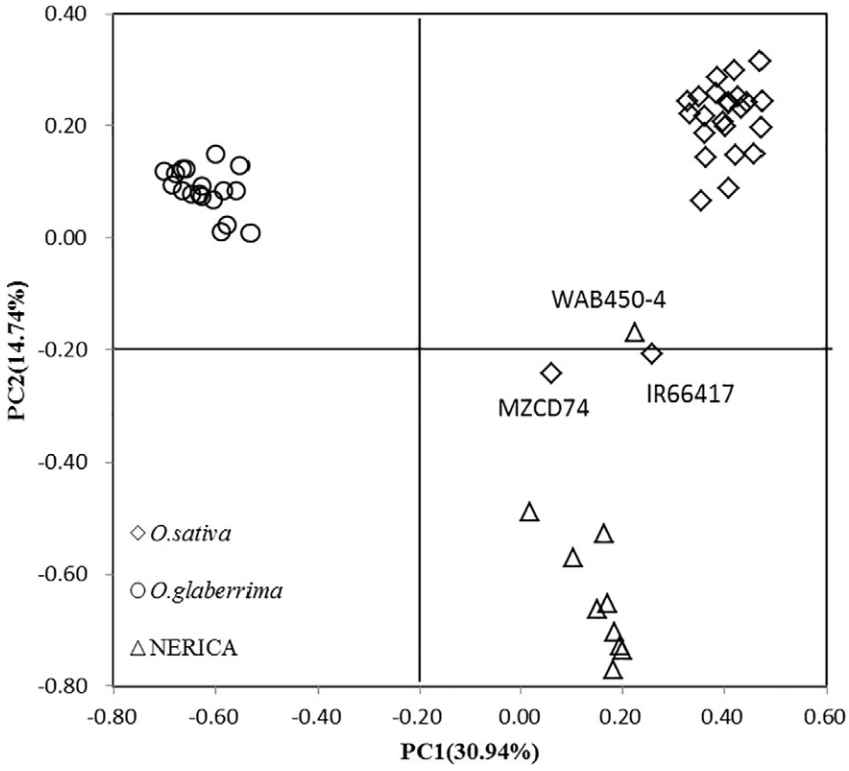

Fig. 5. Principle component analysis of SSR genotypes of 53 rice accessions. The 53 genotypes were divided into three clusters and the first two principal components could explain 30.94% and 14.74% of the genetic variation.

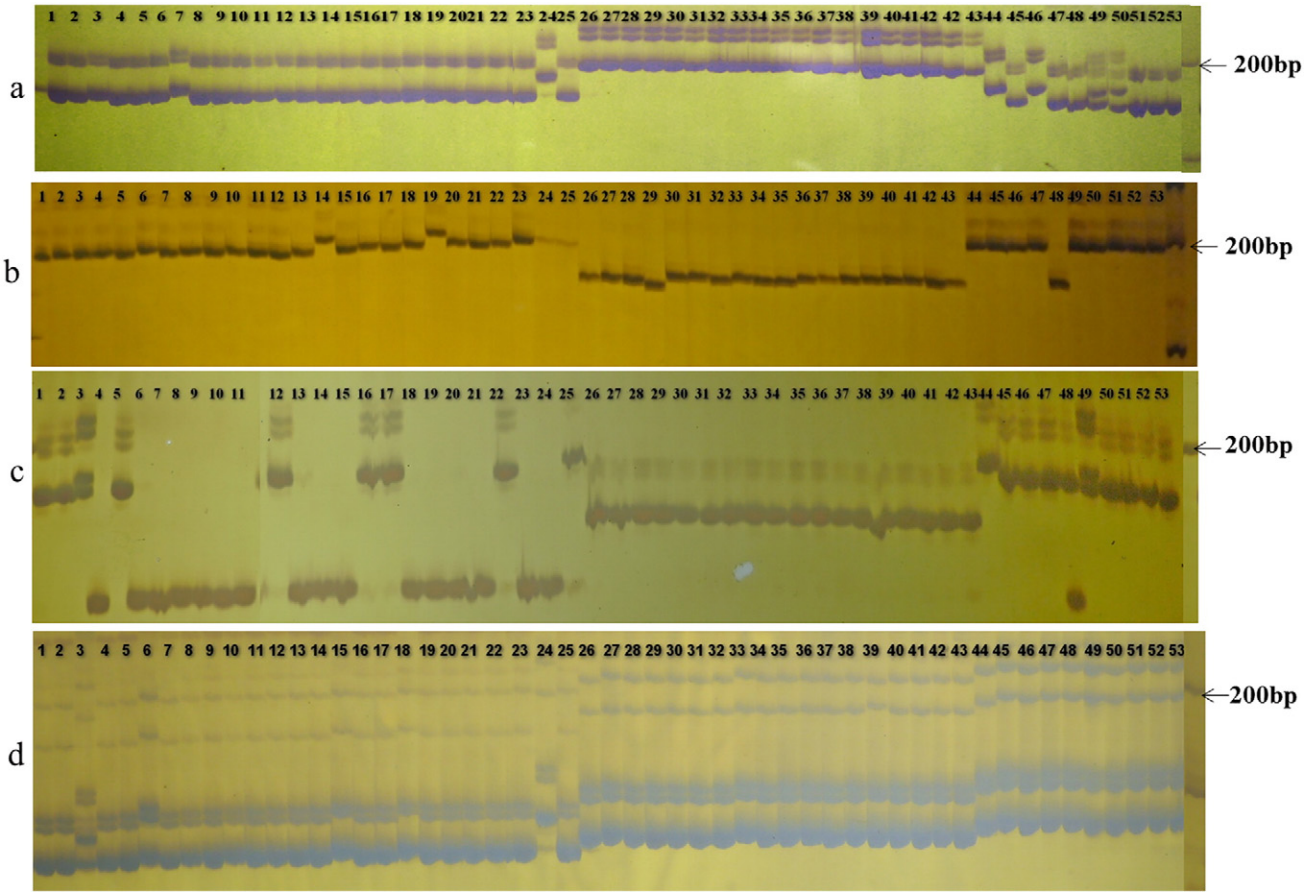

Fig. 6. SSR profiles of 53 rice accessions amplified by primers RM3743 (a), RM197 (b), RM6306 (c), and RM1227 (d). Lanes 1–25: *O. sativa* genotypes; Lanes 26–43: *O. glaberrima* genotypes; Lanes 44–53: NERICA genotypes.

Table 3
Fingerprints of 53 rice cultivars.

Accessions	Fingerprint	Accessions	Fingerprint
Shenghui747	L150D137E151F129J153	RAM48	L167A117B173H145M127
Zhenhanxuan	L150E158F129I148J151	RAM54	L167A119B170F90J166
IR66417	L150B155F143I148M97	RAM55	L167A115B187H139K140
Wanxian98	L150E140F125J151K140	RAM85	L167A115B187H145K138
Qiongxian941	L150D131E151H153J151	RAM90	L167A117B173H149M131
R476	L150E158F125J153K153	RAM91	L167A117B189E163H145
IR72875	L150A98G174J153K156	RAM96	L167A115B170F117I153
Huanghuazhan	L150A98E166F129J140	RAM111	L167A115B170F90I153
Mianhui725	L150D131E138J157K153	RAM118	L167B180E163I159K140
KS99723	L150B142D127F122I148	RAM120	L167A121E166J166M131
IR58025	L150A110E140H145J157	RAM121	L167B180C158E163F119
R80	L150A105E158F117J153	RAM131	L167A115B173K140M127
B370-4	L150A98E172F129J140	RAM133	L167B170D125J159K140
Shuijingmi	L150A98D137E172J157	RAM134	L167A110B180I153J140
YuetaiB	L163D129E140F129K140	RAM145	L167A119B173H149I142
P120	L150D129F129H153J153	RAM152	L167A117E151I153J155
Minghui63	L150D131E158F125J153	WAB375-1	L174A100C158F148J159
R615	L163E177H153K153J157	WAB357-1	L174B170D133H149K153
Jin23B	L150E163H158J157K140	WAB450-1	L174D131E140F161H149
Nongxiang16	L150A105E177H139J157	WAB450-2	L174B163D133E166K149
R207	L150D131E140H153K156	WAB450-3	L174B163F157H158J159
SuyinR02	L150E140F125J155K153	WAB375-2	L174C161D133F152K149
ShengtaiB	L150A105E166J155K156	WAB450-4	L174F152H163I136J162
MZCD74	L180B155F143G185M121	WAB450-5	L174B173C164J153K156
MZCD73	L150A98E180J162K140	WAB450-6	L174B163E158H153K149
RAM3	L167A115B189E163J159	WAB880-1	L174B170E138F161K153
RAM24	L167D125E158H139K138		

Note: Labels for 13 primers: RM312(A), RM250(B), RM60(C), RM255(D), RM163(E), RM1370(F), RM3743(G), RM223(H), RM6543(I), RM6364(J), RM202(K), RM1227(L), RM220(M).

4. Discussion

4.1. Genetic diversity and application potential of the tested rice genotypes

It is reported that genetic diversity of modern rice cultivars has been reduced due to intensive breeding [39]. A large amount of genetic diversity has been eroded both in *indica* and *japonica* subspecies during domestication of Asian rice [13]. Although 25 *O. sativa* cultivars used in this study were collected from six different provinces in China and four other countries (Fig. 1, Table S1), at least 10 cultivars had common genetic lineage (Fig. 7). Six of the 10 cultivars were used as cytoplasmic male sterility (CMS) restorers, whereas two of them were used as cytoplasmic male sterile lines for developing hybrid rice. A total of 22 CMS restorer lines could be found in the ten tested cultivars and their ancestors, in which six restorers were considered six of the 10 most widely used CMS restorer lines in China (Fig. S1; Table S1). For example, Minghui63 is the restorer line widely used in China due to its high general combining ability, and the parent of other four restorer lines R80, Minghui725, R476, and R615, as shown in the pedigree tree (Fig. S1). These varieties show shorter genetic distance in the molecular NJ phylogenetic tree (Fig. 3a) indicating that the genetic diversity of Asian rice cultivars is becoming relatively narrow, although wild species and *japonica* germplasms were used to broaden the genetic bases of hybrid and inbred *indica* varieties (Fig. S1).

Genetic variation provides the potential for a species to adapt to adverse environmental changes [40] and it is critical to introduce novel germplasm to enrich the genetic diversity of rice cultivars. *O. glaberrima* genotype is a closely related species within the genus *Oryza* and could be useful for broadening the genetic basis of *O. sativa* and increasing the rice genetic diversity. The large average genetic distance between *O. glaberrima* and *O. sativa* could become a reproductive barrier and cause interspecific hybrid sterility. However,

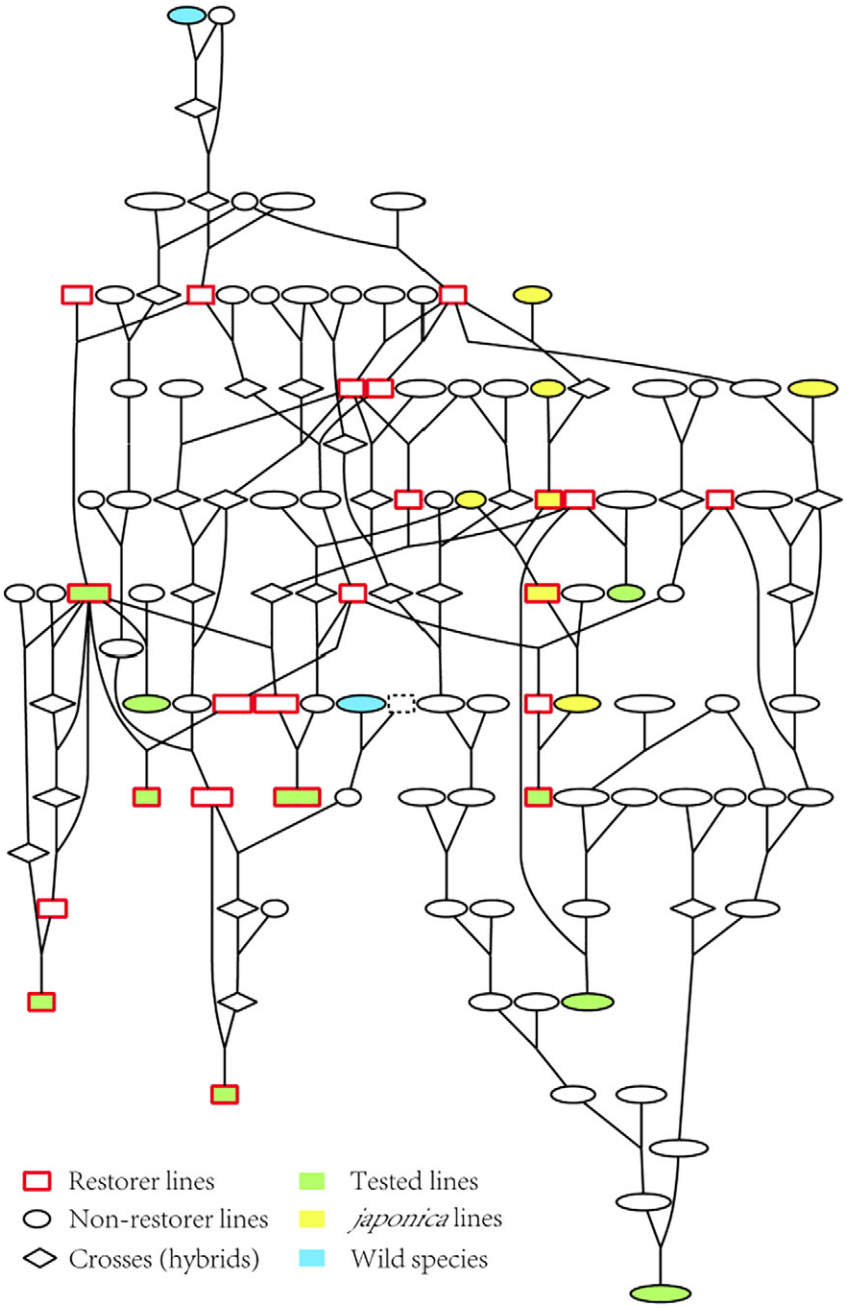

Fig. 7. Pedigree tree of related rice genotypes used in this study.

different genotypes within a species group may have different genetic distances to those in another species group. For example, the genetic distance (0.774) between the *O. sativa* variety Jin23B and *O. glaberrima* variety RAM3 was relatively smaller than the average genetic distance (0.886) of the two species (Fig. 2). In addition, the interspecific cross between RAM3 and Jin23B was one of the few successful combinations from several hundreds of crosses performed [41]. Our results suggest that analyzing the genetic distance between individual genotypes of two species may be helpful when selecting potential parents for interspecific crosses. Introgression lines such as NERICA varieties containing genes of *O. glaberrima* may also be valuable germplasm material, considering the optimal genetic distance with *O. sativa*. It could be used as a bridge germplasm for facilitating the introduction of *O. glaberrima* genes into *O. sativa* genotype to enrich their genetic diversity. Furthermore, we used 30 out of the 53 genotypes including four African rice, three NERICA, and 23 Asian rice cultivars to develop a novel CMS type of African rice cytoplasmic male sterility lines [41], introgression lines carrying African rice genes [5], and partial interspecific hybrid rice [12].

The introduction of genes from distant germplasm into the *O. sativa* genome is an effective way to enrich genetic diversity and a promising approach for breeding novel rice varieties. Information on population structure and genetic diversity of the germplasm is essential for both genetics and breeding projects [42]. Information on genetic diversity among the *O. sativa* and *O. glaberrima* genotypes in this study will enhance our understanding of genetic relationship among different genotypes and facilitate the use of diversified germplasms in rice breeding.

4.2. Genetic differentiation of the three rice groups

O. glaberrima was thought to have been domesticated from the wild ancestor *O. barthii,* whereas *O. rufipogon* was the progenitor of *O. sativa* [1,43]. NERICA cultivars were developed via an interspecific cross between *O. glaberrima* and *O. sativa* [44], whereas weedy rice (*Oryza sativa* f. *spontanea*) was considered to be generated by introgression of *O. sativa* and related wild species such as *O. rufipogon* accessions [45]. Our results corroborated the classification proposed by all the above studies. In our study, the longest genetic distance observed was between the *O. glaberrima* and *O. sativa* groups, whereas NERICA group was placed in between the two groups and was closer to the *O. sativa* group. Although only two weedy rice accessions were studied, our investigation supported the hypothesis that weedy rice likely has different origins. MZCD73 clustered closely to the Asian rice group whereas MZCD74 mapped further away from the Asian rice group, suggesting two different origins: the former might be derived from mutations of tropical cultivars, and the latter might be the result of introgression events between cultivated rice and wild rice species.

4.3. Fingerprinting for classification and identification of rice collections

Fingerprinting with molecular markers allows precise, objective, and rapid variety identification. In previous studies, special bands or multi-polymorphic bands were used to construct fingerprinting of the target variety; however, universal standards to code the molecular markers remain unavailable. Some researchers proposed digital coding in which bands were scored as present (1) or absent (0), and then combined the number of the selected bands, for example, 0101001000100011001 [46]. However, it would be complex to apply these in identifying a genotype in large rice collections, and impractical to classify due to lack of precise length information of the polymorphic SSR bands for target genotypes. Additionally, it is complex to understand and may lead to errors in management of germplasm collections. In this study, we tried to construct a fast, effective, and easy to understand encoding method to use SSR fingerprinting for classification and identification in rice collections. This method proposes (i) screening and numbering the markers with high polymorphism and labeling them in capital letters; (ii) estimating the length of the polymorphic fragment; (iii) selecting a typical polymorphic marker with a specific band capable of distinguishing different groups (such as *O. glaberrima*, *O. sativa*, and NERICA rice) as the first code for all genotypes; and (iv) choosing four additional representative polymorphic bands for each genotype, to follow the first code, to set up a five-code fingerprint for each genotype. In this study, 53 cultivars could be effectively distinguished with 13 SSR primers using this encoding method. The data obtained and recorded via this method can be used effectively for the management of rice germplasm collection.

Conflict of interest

The authors declare no conflict of interest.

Financial support

This study was supported by the Fundamental Research Funds for the Central Universities (2013PY134); Specialized Research Fund for the Doctoral Program of Higher Education (20130146110026).

Supplementary data

http://dx.doi.org/10.1016/j.ejbt.2017.08.001

References

[1] Zhang QJ, Zhu T, Xia EH, et al. Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. Proc Natl Acad Sci U S A 2014;111(46):E4954–62. http://dx.doi.org/10.1073/pnas.1418307111.
[2] Sweeney M, McCouch S. The complex history of the domestication of rice. Ann Bot 2007;100(5):951–7. http://dx.doi.org/10.1093/aob/mcm128.
[3] Johnson DE, Dingkuhn M, Jones MP, et al. The influence of rice plant type on the effect of weed competition on *O. sativa* and *O. glaberrima*. Weed Res 1998;38(3): 207–16. http://dx.doi.org/10.1046/j.1365-3180.1998.00092.x.
[4] Linares OF. African rice (*Oryza glaberrima*): History and future potential. Proc Natl Acad Sci U S A 2002;99(25):16360–5. http://dx.doi.org/10.1073/pnas.252604599.
[5] Chen CJ, He WC, Nassirou TY, et al. Genetic diversity and phenotypic variation in an introgression line population derived from an interspecific cross between *Oryza glaberrima* and *Oryza sativa*. PLoS One 2016;11(9):e0161746. http://dx.doi.org/10.1371/journal.pone.0161746.
[6] Cheng GP, Feng JH, Lang GH, et al. Identification of QTLs for agronomic traits associated with yield in a BC_2F_2 population between *Oryza sativa* and *Oryza rufipogon*. Chin J Rice Sci 2006;20(5):553–6. http://dx.doi.org/10.3321/j.issn:1001-7216.2006.05.017.
[7] Brar DS, Singh K, Oryza. In: Kole C, editor. Wild crop relatives: Genomic and breeding resources, cereals. Berlin: Springer Berlin Heidelberg; 2011. p. 321–65. http://dx.doi.org/10.1007/978-3-642-14228-4.
[8] Xu P, Tao DY, Hu FY, et al. Interspecific hybridization of cultivated rice for breeding japonica rice in Yunnan Province. Chin J Rice Sci 2005;19(1):41–6. http://dx.doi.org/10.3321/j.issn:1001-7216.2005.01.008.
[9] Li F, Liu FH, Morinaga D, et al. A new gene for hybrid sterility from a cross between *O. sativa* and *O. glaberrima*. Plant Breed 2011;130(2):165–71 http://dx.doi.org/10.1111/j.1439-0523.2010.01845.x.
[10] Sarla N, Swamy BPM. *Oryza glaberrima*: A source for the improvement of *Oryza sativa*. Curr Sci 2005;89(6):955–63.
[11] Nassirou TY, He YQ, NERICA: A hope for fighting hunger and poverty in Africa. Mol Plant Breed 2011;2(11):75–82. http://dx.doi.org/10.5376/mpb.2011.02.0011.
[12] Adedze YMN, He WC, Samoura AD, et al. Genomic composition and yield heterosis of the partial interspecific hybrid rice between *Oryza sativa* L. and *Oryza glaberrima* Steud. J Agric Sci 2016;154(3):367–82. http://dx.doi.org/10.1017/S002185961500026X.
[13] Ndjiondjop MN, Cisse F, Girma G, et al. Morpho-agronomic and molecular characterization of *Oryza glaberrima* germplasm from Mali. Afr J Biotechnol 2010;9(44): 7409–17. http://dx.doi.org/10.5897/AJB2010.000-3312.
[14] Semon M, Nielsen R, Jones MP, et al. 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 2005; 169(3):1639–47. http://dx.doi.org/10.1534/genetics.104.033175.
[15] Barry MB, Pham JL, Noyer JL, et al. Genetic diversity of the two cultivated rice species (*O. sativa* and *O. glaberrima*) in Maritime Guinea. Evidence for interspecific recombination. Euphytica 2007;154(1–2):127–37. http://dx.doi.org/10.1007/s10681-006-9278-1.
[16] Zhang DL, Zhang HL, Wang MX, et al. Genetic structure and differentiation of *Oryza sativa* L. in China revealed by microsatellites. Theor Appl Genet 2009;119(6): 1105–17. http://dx.doi.org/10.1007/s00122-009-1112-4.

[17] Agrama HA, Eizenga GC, Yan W. Association mapping of yield and its components in rice cultivars. Mol Breed 2007;19(4):341–56. http://dx.doi.org/10.1007/s11032-006-9066-6.
[18] Rahman MS, Molla MR, Alam MS, et al. DNA fingerprinting of rice (*Oryza sativa* L.) cultivars using microsatellite markers. Aust J Crop Sci 2009;3(3):122–8.
[19] Zhu YF, Qin GC, Jin H, et al. Fingerprinting and variety identification of rice (*Oryza sativa* L.) based on simple sequence repeat markers. Plant Omics 2012;5(4):421–6.
[20] Kuleung C, Baenziger PS, Dweikat I. Transferability of SSR markers among wheat, rye, and triticale. Theor Appl Genet 2004;108(6):1147–50. http://dx.doi.org/10.1007/s00122-003-1532-5.
[21] Xie RJ, Zhou J, Wang GY, et al. Cultivar identification and genetic diversity of Chinese bayberry (*Myrica rubra*) accessions based on fluorescent SSR markers. Plant Mol Biol Report 2011;29(3):554–62. http://dx.doi.org/10.1007/s11105-010-0261-6.
[22] Stich B, Van ID, Melchinger AE, et al. Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 2010;120(7):1289–99. http://dx.doi.org/10.1007/s00122-009-1256-2.
[23] Hamblin MT, Warburton ML, Buckler ES. Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS One 2007;2(12):e1367. http://dx.doi.org/10.1371/journal.pone.0001367.
[24] Tang H, Yu HY, Zhang XM, et al. Analysis on the diversity of DNA fingerprinting of the example varieties used for the test of rice new varieties. J Plant Genet Resour 2015;16(1):100–6. http://doi.org/10.13430/j.cnki.jpgr.2015.01.015.
[25] Luo B, Sun HY, Yang ZG, et al. DNA fingerprints construction and genetic similarity analysis of conventional japonica rice from Taihu Lake area based on SSR markers. J South Agric 2015;46(1):9–14. http://dx.doi.org/10.3969/j:issn.2095-1191.2015.1.9.
[26] Li L, Wang JF, Yan YJ, et al. Establishment of DNA fingerprinting for wheat in Shandong Province by SSR markers. J Plant Genet Resour 2013;14(3):537–41. http://doi.org/10.13430/j.cnki.jpgr.2013.03.017.
[27] Ping HU, Liu T, Yang EQ, et al. Screening of SSR Core primer pairs suitable for establishing DNA fingerprinting pool of Guizhou maize cultivars. Guizhou Agric Sci 2012; 40(7):1–6. http://dx.doi.org/10.3969/j.issn.1001-3601.2012.07.001.
[28] Xue JF, Tan ML, Yan MF, et al. Genetic diversity and DNA fingerprints of castor bean cultivars based on SSR markers. Chin J Oil Crop Sci 2015;37(1):48–54. http://dx.doi.org/10.7505/j.issn.1007-9084.2015.01.008.
[29] Scarano D, Rao R, Masi P, et al. SSR fingerprint reveals mislabeling in commercial processed tomato products. Food Control 2015;51:397–401. http://dx.doi.org/10.1016/j.foodcont.2014.12.006.
[30] Ruiz GJI, Barandalla L, Jose RD, et al. Genetic relationships among local potato cultivars from Spain using SSR markers. Genet Resour Crop Evol 2011;58(3):383–95. http://dx.doi.org/10.1007/s10722-010-9583-3.
[31] Garbe JR, Da Y. A software tool for the graphical visualization of large and complex populations. Acta Genetica Sinica. Decubitus 2003;30(12):1193–5.
[32] Stein N, Herren G, Keller B. A new DNA extraction method for high throughput marker analysis in a large-genome species such as *Triticum aestivum*. Plant Breed 2001;120(4):354–6. http://dx.doi.org/10.1046/j.1439-0523.2001.00615.x.
[33] Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. Genetics 1964;49(4):725–38.
[34] Levene H. On a matching problem arising in genetics. The annals of mathematical statistics. 1949;20(1):91–4. http://dx.doi.org/10.1214/aoms/1177730093.
[35] Lewontin RC. Testing the theory of natural selection. Nature 1972;236(5343):181–2. http://dx.doi.org/10.1038/236181a0.
[36] Botstein D, White RL, Skolmick M, et al. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 1980; 32(3):314–31.
[37] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155(2):945–59. http://dx.doi.org/10.1111/j.1471-8286.2007.01758.x.
[38] Mackill DJ. Classifying japonica cultivars with RAPD markers. Crop Sci 1995;35(3): 889–94. http://dx.doi.org/10.2135/cropsci1995.0011183X003500030043x.
[39] Carvalho RC, Guedes-Pinto H, Igrejas G, et al. High levels of genetic diversity throughout the range of the Portuguese wheat landrace 'Barbela'. Ann Bot 2004; 94(5):699–705. http://dx.doi.org/10.1093/aob/mch194.
[40] Meng ZN, Zhuang ZM, Jin XS, et al. Analysis of RAPD and mitochondrial 16S rRNA gene sequences from *Trichiurus lepturus* and *Eupleurogrammus muticus* in the Yellow Sea. Prog Nat Sci 2003;13(11):1170–6. http://dx.doi.org/10.1080/10020070412331343251.
[41] Huang F, Fu X, Efisue A, et al. Genetically characterizing a new indica cytoplasmic male sterility with *Oryza glaberrima* cytoplasm for its potential use in hybrid rice production. Crop Sci 2013;53(1):132–40. http://dx.doi.org/10.2135/cropsci2012.07.0444.
[42] Nie X, Tu J, Wang B, et al. A BIL population derived from G-hirsutum and G-barbadense provides a resource for cotton genetics and breeding. PLoS One 2015;10(10):e0141064. http://dx.doi.org/10.1371/journal.pone.0141064.
[43] Cheng C, Tsuchimoto, Ohtsubo SH, et al. Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. Genes Genet Syst 2002; 77(5):323–334. http://dx.doi.org/10.1266/ggs.77.323.
[44] Semagn K, Ndjiondjopd MN, Cissoko M. Microsatellites and agronomic traits for assessing genetic relationships among 18 New Rice for Africa (NERICA) varieties. Afr J Biotechnol 2006;5(10):800–10.
[45] Chen LJ, Dong SL, Zhi PS, et al. Gene flow from cultivated rice (*Oryza sativa*) to its weedy and wild relatives. Ann Bot 2004;93(1):67–73. http://dx.doi.org/10.1093/aob/mch006.
[46] Cheng BS, Wei J, Xu WJ, et al. Establishment of SSR fingerprint map and genetic similarity of varieties (lines) in japonica Rice(*Oryza sativa* L.). Jiangsu J Agric Sci 2010; 26(5):897–903. https://doi.org/10.3969/j.issn.1000-4440.2010.05.001.