

Article

Assessment of the Genetic Diversity of Rice Germplasms Characterized by Black-Purple and Red Pericarp Color Using Simple Sequence Repeat Markers

Jae-Ryoung Park ¹, Won-Tae Yang ², Yong-Sham Kwon ², Hyeon-Nam Kim ²,
Kyung-Min Kim ^{1,*} and Doh-Hoon Kim ^{2,*}

¹ Division of Plant Biosciences, School of Applied Biosciences, College of Agriculture and Life Science, Kyungpook National University, Daegu 41566, Korea; icd92@naver.com

² Department of Molecular Genetic Engineering, College of Natural Resources and Life Science, Dong-A University, Busan 49315, Korea; wtyang@dau.ac.kr (W.-T.Y.); yskwon@dau.ac.kr (Y.-S.K.); wiseman311@naver.com (H.-N.K.)

* Correspondence: kkm@knu.ac.kr (K.-M.K.); dhkim@dau.ac.kr (D.-H.K.)

Received: 15 October 2019; Accepted: 1 November 2019; Published: 4 November 2019



Abstract: The assessment of the genetic diversity within germplasm collections can be accomplished using simple sequence repeat (SSR) markers and association mapping techniques. The present study was conducted to evaluate the genetic diversity of a colored rice germplasm collection containing 376 black-purple rice samples and 172 red pericarp samples, conserved by Dong-A University. There were 600 pairs of SSR primers screened against 11 rice varieties. Sixteen informative primer pairs were selected, having high polymorphism information content (PIC) values, which were then used to assess the genetic diversity within the collection. A total of 409 polymorphic amplified fragments were obtained using the 16 SSR markers. The number of alleles per locus ranged from 11 to 47, with an average of 25.6. The average PIC value was 0.913, ranging from 0.855 to 0.964. Four hundred and nine SSR loci were used to calculate Jaccard's distance coefficients, using the unweighted pair-group method with arithmetic mean cluster analysis. These accessions were separated into several distinctive groups corresponding to their morphology. The results provided valuable information for the colored rice breeding program and showed the importance of protecting germplasm resources and the molecular markers that can be derived from them.

Keywords: rice (*Oryza sativa* L.); simple sequence repeat (SSR); cultivar identification; molecular marker; plant variety protection

1. Introduction

Rice (*Oryza sativa* L.) is one of the most widely cultivated crops in the world; distributed across a variety of climates and regions, it has developed a diverse array of genotypes and phenotypes. Koreans have historically used rice as a major food source. However, in recent years, both the production and consumption of rice have been in decline. In 2017, rice production was 3.972 million tons, down 5.3% from 4.197 million tons the previous year, and the annual rice consumption per capita in 2017 was 70.9%, down 0.4% from the previous year. The continuous decrease in the production and consumption of rice in Korea means that the industry is not stable [1]. One reason for the annual decline in rice consumption is the increase in the number of people seeking meat and salad alternatives because of a preference for a westernized diet [2]. Therefore, to increase the consumption of rice in Korea, processed food products using a rice base have been developed. Colored rice varieties such as black-purple rice

are particularly popular for this purpose because of their unique color and aroma [3,4]. To increase rice consumption and production, high-quality rice containing the functional materials preferred by producers and consumers are also cultivated [5]. Rice varieties are largely classified according to the main areas of their cultivation, such as *Oryza sativa* L. ssp. *japonica*, *Oryza sativa* L. ssp. *indica*, and *Oryza sativa* L. ssp. *javanica* [6–8]; these can be utilized to further breed new, high-quality varieties. Despite the abundance of rice genetic resources globally, the genetic similarity of rice varieties has increased, and the variety of species has continued to decrease, since only a few specific varieties are utilized for breeding [9,10]. In the process of producing new varieties, genetic diversity was greatly reduced in both wild and cultivated rice. At present, various rice breeding institutes in Korea and abroad are trying to breed high-quality rice varieties; among the cultivars developed with the artificial hybridization method rather than the transgenic technique are “Tongilbyeo” and “Naepungbyeo” [11,12]. These are usually made by crossing several different varieties of ssp. *japonica*. It is assumed that the currently developed varieties are genetically very similar. As the minimum genetic distance between the varieties is very small because of the genetic resources used for cultivation, the varieties of rice in Korea are not very diverse, and seed disputes related to genotype use and the protection of breeders’ rights frequently occur. To overcome these problems, we have constructed a database for each species using molecular markers and established a genetic linkage to known varieties [13]. The International Union for the Protection of New Varieties of Plants (UPOV) also provides guidelines for database construction of varieties using molecular markers [14]. Assessing the diversity of rice genetic resources involves identifying phenotypes, analyzing biochemistry, and evaluating DNA diversity [15,16]. The evaluation methods involving the identification of phenotypes and biochemical characteristics are not necessarily reliable, as they are environmentally influenced, labor intensive, and costly [17,18]. However, evaluating genetic diversity based on DNA is the most widely used evaluation method in terms of repeatability, stability, and reliability. Several techniques have been developed to analyze the genetic diversity of rice varieties based on DNA, such as the analysis of restriction fragment length polymorphism (RFLP) using restriction enzymes, random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSR), and single nucleotide polymorphisms (SNP) utilizing the polymerase chain reaction (PCR) [19,20]. The RAPD and AFLP techniques are not easily reproducible and the markers are dominant, so there are many limitations in analyzing the relationship of genetic resources. However, since SSR markers exhibit high polymorphism and show co-dominant status, they can be used to identify heterozygosity. SSR markers, also called DNA microsatellites, are regions of DNA (often forming part of the non-coding regions) where sequences of one to five nucleotides are repeated, and they are uniformly distributed in the genomes of most eukaryotes. The SSR sequences found in plants are frequently made up of AT and GA nucleotide repeats. The relationships of genetic resources have been analyzed using SSR markers in other crops such as melon (*Cucumis melo* L.), watermelon (*Citrullus lanatus*), and corn (*Zea mays*) [21–23]. SSR markers representing the entire genome have been developed and commercialized for rice with an average distance of 3–4 cM between markers [17,24,25], but few have been used to identify rice varieties to protect the breeders’ rights. Colored rice has many antioxidants. At present, the development of colored rice with high antioxidant content is being made continuously. Therefore, in this study, if we constructed a DNA database of colored rice gene source by using SSR markers, this study was expected to be helpful for the improvement of colored rice varieties in the future. Also, in this study, we developed a core marker set for the identification of rice varieties using SSR markers and constructed a DNA profile database for rice varieties using the developed markers. To characterize the SSR markers with a high degree of polymorphism in the rice genetic resources, allele characteristics of 600 SSR markers reported in the rice genome database (<http://www.gramene.org/>) were used. The best markers were selected and analyzed for gene affinity with 548 colored rice genetic resources (376 black-purple rice samples, 172 red pericarp rice samples) collected both at home and abroad.

2. Results

2.1. Microsatellite Analysis

Assessing the genetic diversity of our resources is an essential process for efficiently preserving biodiversity and for characterizing and exploiting them. We used 600 microsatellite primers, reported in the rice genome database (<http://www.gramene.org/>), to select microsatellite markers suitable for rice germplasm identification and genetic diversity assessment. The microsatellite markers were amplified in 11 different varieties: “Hopum,” “Hwayeong,” “Dami,” “Sindongjin,” “Odae,” “Chucheong,” “Saechucheong,” “Junam,” “Ilpum,” “Ilmi,” and “Gopum.” Polymorphisms of primers and characteristics of the alleles were examined after removing those that showed the dominant form, had no obvious bands, or had low repeatability when the PCR was performed; 16 microsatellite markers were deemed suitable (KRF24, RM8085, RM72, KRF16, HvSSR02-86, RM18821, RM12834, RM20754, RM26063, RM333, RM15641, RM26730, RM19731, KRF15, KRF17, RM24946). The 16 microsatellite markers showed co-dominance, high repeatability, and high polymorphism rates and were selected to evaluate the diversity of the rice germplasm resources (Figure 1 and Table 1). To investigate the genetic polymorphisms of the chromosomal genetic resources, the fluorescent labels FAM, VIC, NED, and PET were attached to the forward 5'-ends of the 16 microsatellite markers selected above. PCR and electrophoresis were carried out with the rice germplasm DNA using an automatic nucleotide sequencer, and the degree of polymorphism was examined. A total of 409 alleles were amplified with the selected microsatellite markers. The number of alleles found in each variety ranged from 11 to 47, with an average of 25.6 alleles per variety. The polymorphism information content (PIC) values ranged from 0.855 to 0.964, and the average was very high at 0.913. High PIC values indicate that the selected microsatellite markers are efficient at evaluating a large number of genetic resources. In this study, the mean PIC value of the microsatellite markers used to assess the diversity of rice genetic resources was 0.913, which was much higher than the PIC values of microsatellite markers used in other studies [26–28]. These results indicate that the microsatellite markers selected for use in this study are very suitable and efficient for assessing the genetic diversity of many different rice genetic resources. Several studies have been conducted to evaluate genetic diversity using microsatellite markers in rice [17,24,25]. However, because of the abundance of rice genetic resources, and as only a fraction of these resources have been analyzed [29], it is necessary to conduct more such studies. In addition, since the PIC values of the microsatellite markers used in this study were much higher than those in other studies [30], it appears that the microsatellite markers used for the analysis of the 548 rice genetic resources in this study could be used more effectively to establish the genetic relationship of rice genetic resources and to create a database for each species.

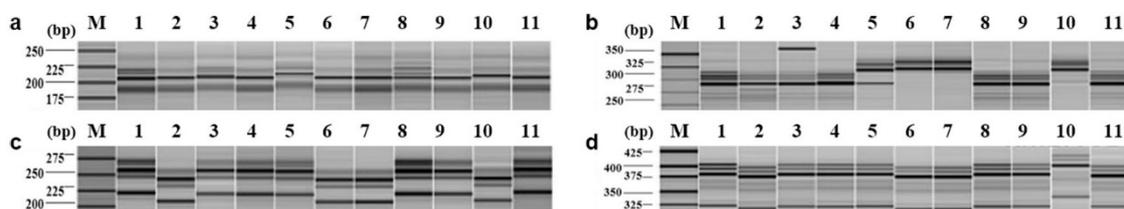


Figure 1. Polymorphism of four simple sequence repeat (SSR) markers, RM20754, HVSSR2-86, RM15641, and RM19731. The amplified polymerase chain reaction (PCR) products were loaded on the HAD-GT12™ Genetic Analyzer System and analyzed using Biocalculator Data Analysis Software. (a): RM20754, (b): HVSSR2-86, (c): RM15641, (d): RM19731; Lane 1: Hopum, 2: Hwayeong, 3: Dami, 4: Sindongjin, 5: Odae, 6: Chucheong, 7: Saechucheong, 8: Junam, 9: Ilpum, 10: Ilmi, 11: Gopum.

Table 1. Repeat motif, no. of alleles, and polymorphism information content (PIC) value of microsatellite markers selected for genetic characterization of rice cultivars and germplasms.

No.	SSR Primers	Repeat Motif	Chromosome Number	Annealing Temperature (°C)	PCR Product Size (bp)	No. of Alleles	PIC Value
1	KRF24	terminal repeat	7	55	213–238	22	0.902
2	RM8085	ag	1	55	105–144	20	0.915
3	RM72	(tat) ₅ c(att) ₁₅	8	55	151–207	20	0.916
4	KRF16	aac	9	55	160–253	27	0.963
5	HvSSR02-86	aac	2	55	187–328	47	0.970
6	RM18821	tct	5	55	150–205	17	0.876
7	RM12834	aga	2	55	186–260	25	0.862
8	RM20754	aat	6	55	151–219	22	0.955
9	RM26063	ct	11	55	225–283	27	0.952
10	RM333	aat	10	55	156–312	35	0.964
11	RM15641	aat	3	55	163–247	31	0.894
12	RM26730	ctt	11	55	98–157	11	0.781
13	RM19731	tta	6	55	339–429	31	0.964
14	KRF15	atg	8	55	126–175	18	0.878
15	KRF17	tatc	2	55	243–405	41	0.956
16	RM24946	atg	10	55	317–369	15	0.855
Total						409.00	14.604
Mean						25.56	0.913

2.2. Hierarchical Cluster Analysis

The unweighted pair-group method with arithmetical average (UPGMA) dendrogram was constructed using the MEGA7 program and Jaccard's distance coefficients using the results of the PCR with the 16 microsatellite markers selected from the 548 colored rice genetic resources (Figure S1). The analysis classified the 548 rice genetic resources into four major groups (I, II, III, and IV) with seven subgroups (I-1, I-2, I-3, II, III, IV-1, and IV-2). In addition, as the black-purple rice and red pericarp rice were not clearly distinguished from each other, they were mixed and classified according to the environmental conditions of the areas where each genetic resource was grown (Figure 2). The 548 rice genetic resources used in this study included 337 varieties in Group I, 13 varieties in Group II, 144 varieties in Group III, and 54 varieties in Group IV. Group I included 18 varieties from Malaysia, five from Myanmar, two from Vietnam, two from Bhutan, five from Sri Lanka, 95 from Indonesia, 43 from Japan, 78 from China, three from Cambodia, three from Taiwan, 26 from Thailand, two from Pakistan, 23 from the Philippines, and 32 from Korea. Group I included most of Asia and could be divided into three subgroups (I-1, I-2, I-3). Subgroup I-1 included mostly South Asian countries such as Malaysia, Indonesia, and Cambodia. Subgroup I-2 included East Asian countries such as Japan and Korea, and Subgroup I-3 included Malaysia, Bhutan, Sri Lanka, Indonesia, Japan, China, Thailand, the Philippines, and South Korea. Group II contained one variety from Indonesia, 11 from Japan, and one from Korea. Group III included one variety from Malaysia, two from Vietnam, 13 from Indonesia, 41 from Japan, 51 from China, two from Cambodia, one from Thailand, two from the Philippines, and 31 from Korea. Group III accounted for about 86% of the rice genetic resources of Japan, China, and Korea. In Group IV, Korean rice genotypes constituted the majority, with one variety from Myanmar, one from Indonesia, four from Japan, 10 from China, one from Cambodia, one from Thailand, one from the Philippines, and 35 from Korea. Group IV was also divided into two subgroups (IV-1, IV-2); Subgroup IV-1 had one variety from Indonesia, one from Japan, five from China, one from Cambodia, one from Thailand, and 31 from Korea. In Subgroup IV-2, there was one variety from Myanmar, three from Japan, five from China, one from the Philippines, and four from South Korea. Subgroup IV-2 contained rice genes from East Asian countries including Japan, China, and Korea. The UPGMA analysis classified most of the Southeast Asian rice genetic resources into Group I. Rice genetic resources of Japan were included in Group II, and Group III contained genetic resources from Japan, China, and Korea. In Group IV, most of the samples were confirmed as Korean rice genetic resources. These groupings were based on

areas where each rice genetic resource was grown, mainly Southeast Asia and East Asia. Southeast Asia, because of its proximity to the equator, represents a tropical climate with high temperatures throughout the year, and it is influenced by the monsoon season with frequent precipitation, except in the dry season. Therefore, it was confirmed that the *ssp. indica* varieties, suitable for this environment, were cultivated in Southeast Asia. On the contrary, the East Asian region has four distinct seasons, spring, summer, autumn, and winter, and has many plains; in this climate, the *ssp. japonica* varieties are cultivated rather than *ssp. indica*. An analysis of the dendrogram of the microsatellite marker genotypes from this study confirmed that similar rice varieties were in the same groups, and similar varieties of crosses were used for the cultivation of growth environments and breeds. When the dendrogram was made by the UPGMA analysis, Group I contained genetic resources cultivated in Southeast Asian countries, such as Indonesia, Malaysia, the Philippines, and Thailand. Groups I, III, and IV included genes cultivated in the East Asian countries of Japan, China, and Korea. These results were divided according to the genetic resources of the maternal and paternal families from which the varieties were developed as well as the results of other researchers and the genotype of rice or the environment of the region where the cultivars were grown [28,31]. Therefore, it is expected that more effective breeding results will be obtained by selection based on the maternal or paternal samples when cultivating a new breed for a new environment.

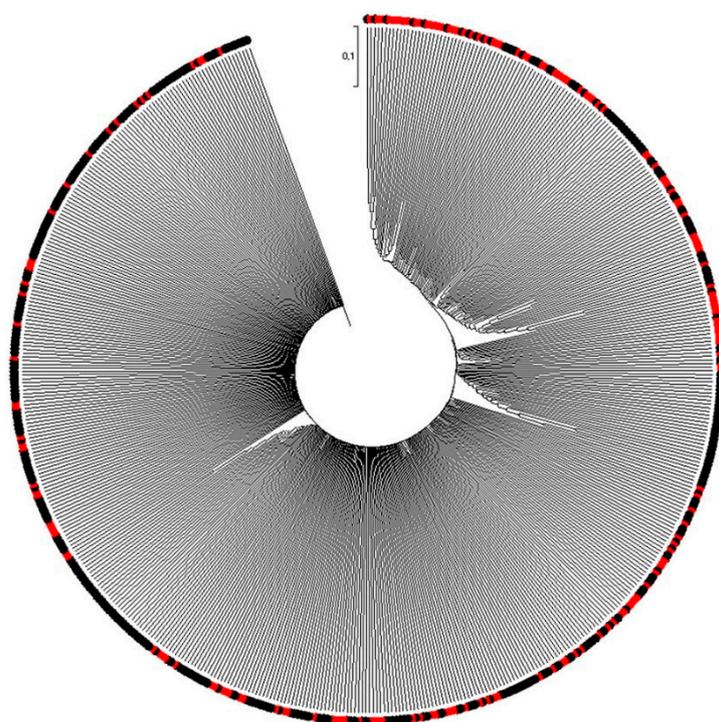


Figure 2. Unweighted pair-group method with arithmetical average (UPGMA) dendrogram of 548 rice germplasms based on 16 SSR markers using MEGA7 program. The red color in the boundary indicates rice genotypes that have red seeds and the black color indicates rice genotypes that have black seeds. Black and red rice are not clearly distinguished and mixed in large groups.

2.3. Population Structure

A model-based program called STRUCTURE v2.3.4 was used to determine the genetic relationship of the 548 colored rice genetic resources. In order to analyze the colored rice genetic resources, ΔK was chosen with the peak alpha parameter. Population structure of the 192 germplasm lines was analyzed by a Bayesian based approach. The estimated membership fractions of 192 accessions for different values of k ranged between 2 and 5 (Figure 3). The log likelihood revealed by structure showed the optimum value as 2 ($K = 2$). However, when the K value was set to 2, 548 colored rice gene sources

could not be effectively distinguished. Because 548 colored rice gene sources have a very diverse gene pool. Thus, when the K value was set to the next higher K = 3, 4, 7, the population structure was confirmed. When the K value was set to 7, 548 colored rice genes were classified into the optimal population structure. The seven groups were divided according to the environments in which the colored rice genetic resources were grown and the type of classification. In addition, analysis of the population structure distinguished whether the colored rice genetic resources belonging to the group were pure or admixtures; varieties with a probability value greater than 0.80 were considered pure. Most of the colored rice genetic resources belonging to cluster 7 were pure, and most of the colored rice genetic resources belonging to the remaining subgroups were admixtures (Figure 4). Most of the colored rice genetic resources used in this study were in the form of admixtures, meaning that the colored rice genetic resources were not high in purity but derived through many recombination and crossbreeding events. Population structure analysis showed that the grouping of colored rice genetic resources was like that of the dendrogram for the UPGMA analysis based on genetic distances. When the dendrogram was drawn through the UPGMA analysis, it was divided into four groups with seven subgroups. The population structure analysis showed the highest value at K = 7, so the 548 colored rice genetic resources were divided into seven groups. Population structure analysis showed that the classification of colored rice genetic resources, the geographical location, and the goal of breeding were significantly influenced by the genetic structure of rice genotypes.

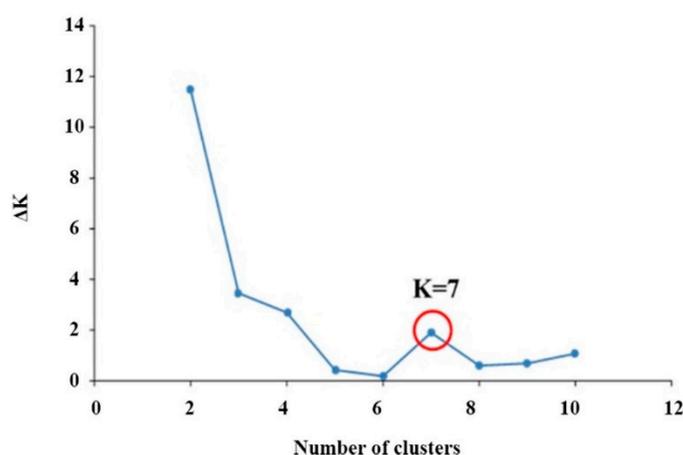


Figure 3. Inference of number of clusters in rice germplasm collection based on 16 SSR markers using STRUCTURE program. The Bayesian model in the STRUCTURE v2.3.4 program was used to test K (number of clusters) using the parameter of burn-in/Markov chain Monte Carlo (MCMC): 100,000/250,000 iterations. An ad hoc quantity (ΔK) [31] was calculated for each K to detect the best number of clusters (K = 7).

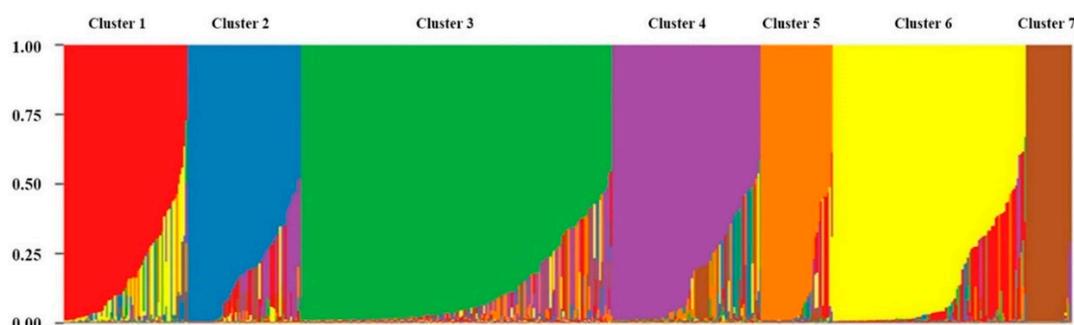


Figure 4. Structure plot presenting seven clusters of rice germplasm collection. The y -axis indicates the estimated membership coefficients for each individual. Each variety's genome is represented by a single vertical line, which is partitioned into colored segments in proportion to the estimated membership in the seven clusters.

2.4. Principal Component Analysis (PCA)

PCA is an analytical method that finds the best representations of the differences of each dataset and distinguishes the data by each element. In other words, PCA is a method in which when data are presented as axes, the axis with the greatest variance is set as the first main component, and the axis with the second largest variance is displayed as a diagram. When PCA was performed to confirm the degree of genetic diversity of the 548 colored rice genetic resources, all the genetic resources were uniformly distributed in four quadrants and were not shifted to any one place (Figure 5). Two-dimensional data obtained through PCA analysis resembled the results of the dendrogram and population structure analyses, and most varieties were distributed in all quadrants. When the UPGMA dendrogram was created, the colored rice genetic resources belonging to Subgroup I-1 were uniformly distributed in the first and second quadrants of the diagram, and those of Subgroup I-2 were distributed in the third quadrant. Subgroup I-3 appeared in the lower-left corner of the fourth quadrant, and Group II appeared in the upper-left quadrant. Group III was shown in the lower half of Quadrant 1 and Quadrant 4. Subgroup IV-1 was uniformly shown in the first and second quadrants like Group I, and Subgroup IV-2 was shifted to the lower right of the third quadrant. This PCA can be used to classify genotypes of colored rice genetic resources based on molecular morphological changes as well as plant morphological data. However, the PCA only finds the best representatives of the differences in genetic resources in the population; it does not provide information on the number of populations or subgroups of populations. In addition, PCA is based on morphological markers and it analyzes group structure. Therefore, PCA is less accurate than other methods. Since PCA showed very high variability in the principal component of this study, this information may be helpful in identifying related genotypes if used with population analysis. It can also be useful to distinguish between subgroups of populations involved in classifying genotypes. In addition, when comparing the PCA with the dendrogram, they showed similar tendencies and their general and group compositions were similar.

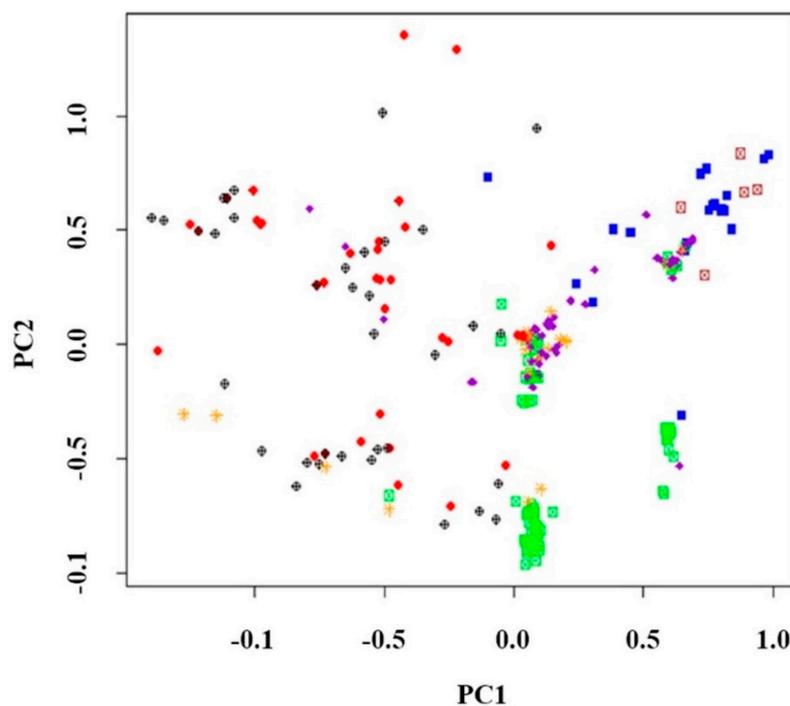


Figure 5. Principal component analysis (PCA) of 548 rice germplasms based on 16 SSR markers using R program. The colors shown in this figure represent the same genetic resources as the colors in Figure 5. The rice genotypes belonging to cluster 1 are indicated by red; cluster 2: blue; cluster 3: green; cluster 4: violet; cluster 5: orange; cluster 6: black; and cluster 7: brown.

3. Discussion

Sixteen SSR markers were used for the assessment of colored rice genetic diversity. Each marker can distinguish only one or two varieties, but using multiple marker set simultaneously can assess the genetic diversity of numerous gene resources. In this study, when UPGMA dendrograms of the 548 colored rice genetic resources were created using Jaccard's distance coefficients and the MEGA7 program, they were divided into four large groups (I, II, III, and IV) and seven subgroups (I-1, I-2, I-3, II, III, IV-1, and IV-2). Moreover, when analyzing the population structure using the STRUCTURE v2.3.4 program, it was confirmed that as in the previous results, optimal separation occurred when the 548 colored rice genetic resources were divided into seven subgroups [32]. In the analysis of the colored rice population structure using the UPGMA dendrogram and STRUCTURE programs, it was confirmed that it was divided into seven groups in all analysis. The results were obtained when assessment genetic diversity based on genetic distance and model-based population structure. In addition, when analyzing the population structure using the STRUCTURE program, it was confirmed that the colored rice gene sources were genetically mixed. The reason why the genetic material is so mixed is because breeding continues to develop varieties with high yield and quality. Since this breeding will not end now and will continue in the future, the mix of genetic resources will continue to increase. When the UPGMA dendrogram was created, Group I contained evenly colored rice genetic resources from Southeast Asian countries such as the Philippines, Indonesia, Thailand, Korea, China, and Japan. Group II included most of the Japanese colored rice genetic resources, and Group III included the rice genetic resources of East Asian countries including Japan, China, and Korea. Finally, Group IV included most of the Korean colored rice genetic resources. When the selected microsatellite markers were used to analyze the populations of rice genetic resources, some of the varieties were classified as identical. It was presumed that these results were obtained because rice genetic resources are composed of the same maternal or paternal material and the genetic resources should be judged by either comparing the morphological characteristics or increasing the number of microsatellite markers. In this study, 16 microsatellite markers were newly selected from the 600 pairs of microsatellite primers reported in the rice genome database (<http://www.gramene.org/>); they showed high polymorphism and excellent repeatability of the pattern of alleles from the PCR results. The results of the analysis of the 548 colored rice genetic resources, the black-purple rice and red pericarp rice, using these microsatellite markers showed that the colors were not clearly distinguishable but that the resources were clearly grouped according to plant taxonomic characteristics, breeding lineage, and cultivation environment of rice. Therefore, the database of DNA profiles constructed through the selected microsatellite markers can be used not only to evaluate the characteristics of rice genetic resources but also to select the controls for cultivar-protected varieties. There are many genetic resources that have not been identified by the microsatellite markers used in this study and there have been many recent reports of new SNP markers [33], highlighting the necessity to continue to develop new markers, specific to varieties, by comparing and analyzing morphological characteristics.

4. Materials and Methods

4.1. Plant Material

In this study, 548 colored rice genetic resources (376 black-purple rice samples, 172 red pericarp rice samples) collected from national and international sources were used as analytical materials for microsatellite markers (Table S1).

4.2. Genomic DNA Extraction

Genomic DNA was isolated using a NucleoSpin®Plant II (Cat. 740 770.250, Macherey–Nagel GmbH & Co., KG, Deutscher, Düren, Germany) kit. The extracted genomic DNA was quantitated at 260 nm using a spectrophotometer, and the DNA concentration was confirmed by electrophoresis on 1.5% agarose gel. The DNA was diluted to a final concentration of 20 ng/μL using sterilized water

and then used for SSR analysis. The concentration and quality of the obtained genomic DNA samples were estimated with an ultramicrospectrophotometer Nano Drop (Cat. ND-2000, ThermoFisher SCIENTIFIC, Seoul, South Korea).

4.3. SSR Genotyping

In order to select SSR markers effective for colored rice genetic resources, the 600 SSR primer sets reported in the rice genome database (<http://www.gramene.org/>) were used to carry out PCR on the “Hopum,” “Hwayeong,” “Dami,” “Sindongjin,” “Odae,” “Chucheong,” “Saechucheong,” “Junam,” “Ilpum,” “Ilmi,” and “Gopum” rice varieties. The primers with high repeatability, high band reproducibility, and a high degree of polymorphism were first selected. Primers having a high PIC value and a high heterozygosity value were selected among the primers selected first. Finally, the selected SSR primer sets were distributed evenly on the rice chromosome. Gene amplification products by PCR were analyzed by electrophoresis using Genetic Analyzer (HAD-GT12™ System, eGENE, Irvine, CA, USA), and the differences in the various alleles were analyzed using a computer program (Biocalculator 2.0). The PCR amplification results were used to construct a DNA profile database for the 548 colored rice genetic resources. The PCR reaction mixture contained 50 µg of genomic DNA, 0.1 µM of SSR primer, and 2.0 µL of dNTP mixture (2.5 mM), and 1.0 U of HS Prime Taq polymerase (Cat. G-7002, Genet Bio, Daejeon, South Korea, 50 mM KCl, 20 mM Tris-HCl, pH 8.0, 2.0 mM MgCl₂) was added with distilled water to adjust the total volume to 30 µL. PCR amplification (C1000, BioRad, Hercules, CA, USA) was performed by denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 45 s, for a total of 40 cycles. After completion of the PCR, 3.0 µL of the PCR amplification product was electrophoresed on a 2.5% agarose gel to confirm the PCR amplification. After confirming amplification, 1.5 µL of the PCR amplification product was diluted in 150 µL of distilled water. To estimate the size of the alleles of the chromogenic genes according to the SSR markers, 1.0 µL of the PCR amplification product, 10 µL of deionized formamide, and 0.25 µL of size marker (LIZ500 size standard) were mixed. After denaturation at 94 °C for 2 min, electrophoresis was performed using an automatic sequencer (Genetic Analyzer 3130XL, Applied Biosystems, Foster City, CA, USA). And the size of amplified alleles was measured using the Gene Mapper (version 3.7) program (Applied Biosystems, Foster City, CA, USA). The number of alleles and the size of alleles were determined for each marker.

4.4. Genetic Diversity Analysis

PIC values were calculated using the following formula to investigate the diversity of microsatellite markers:

$$PIC = 1 - \sum_{i=1}^n p_i^2, \quad (1)$$

where n is the number of alleles analyzed per marker and P_{ij} is the frequency of the j th common band pattern among the bands of marker i . [34]. Microsatellite analysis was used to select alleles with high reproducibility and high polymorphism as markers. NTSYSpc (version 2.10) [35] was performed according to dominant marker scoring (present = 1, absent = 0), and genetic similarity values were calculated according to the Jaccard’s method [36], followed by population analysis using UPGMA [37]. The population structure was deduced from the model-based program STRUCTURE v2.3.4 [38] and performed over the range of $K = 2$ to 10. The final K value was determined using Evanno’s ΔK method [39]. Allelic frequencies of each microsatellite marker and each variety were calculated and used for PCA. PCA analysis was performed using the R program (ver. 3.2.3).

Supplementary Materials: The following are available online at <http://www.mdpi.com/2223-7747/8/11/471/s1>, Figure S1: Dendrogram of 548 rice varieties based on 16 SSR markers. The scale at the bottom is Jaccard's coefficient of similarity. Table S1: List of rice cultivars and germplasms assayed for genetic characterization using microsatellite markers.

Author Contributions: Conceptualization, J.-R.P.; methodology, J.-R.P. and W.-T.Y.; formal analysis, W.-T.Y.; investigation, J.-R.P., W.-T.Y., Y.-S.K., H.-N.K.; writing-original draft preparation, J.-R.P. and W.-T.Y.; writing-review and editing, J.-R.P. and W.-T.Y.; project administration, K.-M.K. and D.-H.K.

Funding: This work was supported by a grant from the Next-Generation BioGreen 21 program (Project No. PJ01368401), Rural Development Administration, Republic of Korea and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018RC1B5047863).

Acknowledgments: This work was supported by a grant from the Next-Generation BioGreen 21 program (Project No. PJ01368401), Rural Development Administration, Republic of Korea and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2018RC1B5047863).

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

References

1. Sakong, Y. Measures of price response to supply shock in Korean rice market. *J. Korean Agric. Econ. Assoc.* **2006**, *47*, 1–16.
2. Chung, H.K. Changes in the Korean food culture with the change of life style. *J. Korean Soc. Food Sci. Nutr.* **2001**, *5*, 17–24.
3. Ichikawa, H.; Ichiyana, T.; Xu, B.; Yoshii, Y.; Nakajima, M.; Konishi, T. Antioxidant activity of anthocyanin extract from purple black rice. *J. Med. Food* **2001**, *4*, 211–218. [[CrossRef](#)] [[PubMed](#)]
4. Sutharut, J.; Sudarat, J. Total anthocyanin content and antioxidant activity of germinated colored rice. *Int. Food Res. J.* **2012**, *19*, 215–221.
5. Lee, J.H.; Yeo, U.S.; Cho, J.H.; Lee, J.Y.; Song, Y.C.; Shin, M.S.; Kang, H.W.; Sohn, J.K. Marker assisted selection of brown planthopper resistance and development of multi-resistance to insect and diseases in rice (*Oryza sativa* L.). *Korean J. Breed Sci.* **2011**, *43*, 413–421.
6. Khush, G.S. Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **1997**, *35*, 25–34. [[CrossRef](#)]
7. Garris, A.; Tai, T.; Coburn, J.; Kresovich, S.; McCouch, S.R. Genetic structure and diversity in *Oryza sativa* L. *Genetics* **2005**, *169*, 1631–1638. [[CrossRef](#)]
8. Lu, H.; Redus, M.A.; Coburn, J.R.; Rutger, J.N.; McCouch, S.R.; Tai, T.H. Population structure and breeding patterns of 145 US rice cultivars based on SSR marker analysis. *Crop Sci.* **2005**, *45*, 66–76. [[CrossRef](#)]
9. Nanda, J.S.; Sharma, S.D. *Monograph of Genus Oryza*; Science Publishers: Enfield, NH, USA, 2003; pp. 283–309.
10. Das, B.; Sengupta, S.; Parida, S.K.; Roy, B.; Ghosh, M.; Prasad, M.; Ghose, T.K. Genetic diversity and population structure of rice landraces from Eastern and North Eastern States of India. *BMC Genet.* **2013**, *14*, 71. [[CrossRef](#)]
11. Chung, D.H.; Lee, H.Y. Studies on the improvement of milling, quality and storage of Tongil (indica type) rice. *Korean J. Food Sci. Technol.* **1976**, *8*, 179–184.
12. Lee, K.Y.; Park, S.T.; Kim, J.H.; Lee, J.S.; Shin, S.O.; Lim, S.J.; Kim, S.C.; Oh, Y.J.; Ko, J.C.; Jin, Y.D.; et al. Information (brief introduction of newly released crop cultivars in 1996): A new high yielding rice variety with good grain quality and resistance to wind, “Naepungbyeon”. *Korean J. Breed Sci.* **1996**, *28*, 483.
13. Kwon, Y.S.; Oh, Y.H.; Yi, S.I.; Kim, H.Y.; An, J.M.; Yang, S.G.; Ok, S.H.; Shin, J.S. Informative SSR markers for commercial variety discrimination in watermelon (*Citrullus lanatus*). *Genes Genom.* **2010**, *32*, 115–122. [[CrossRef](#)]
14. UPOV UPOV/INF/17/1 *Guideline for DNA-Profiling: Molecular Marker Selection and Database Construction (“BMT Guideline”)*; World Intellectual Property Organization: Geneva, Switzerland, 2010.
15. Second, G. Origin of the genic diversity of cultivated rice (*Oryza* spp.): Study of the polymorphism scored at 40 isozyme loci. *Jap. J. Genet.* **1982**, *57*, 25–57. [[CrossRef](#)]
16. McNally, K.L.; Childs, K.L.; Bohnert, R.; Davidson, R.M.; Zhao, K.; Ulat, V.J.; Stokowski, R. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12273–12278. [[CrossRef](#)] [[PubMed](#)]
17. Song, M.T.; Jong, S.K.; Lee, J.H.; Cho, Y.S.; Gul, J.H.; Lee, S.B.; Hwang, H.G. Comparison of DNA-based and pedigree-based genetic similarity among Korean rice cultivars. *Korean J. Genet.* **2003**, *48*, 74–75.

18. Thomson, M.J.; Septiningsih, E.M.; Suwardjo, F.; Santoso, T.J.; Silitonga, T.S.; McCouch, S.R. Genetic diversity analysis of traditional and improved Indonesian rice (*Oryza sativa* L.) germplasm using microsatellite markers. *Theor. Appl. Genet.* **2007**, *114*, 559–568. [[CrossRef](#)] [[PubMed](#)]
19. Saini, N.; Jain, N.; Jain, S.; Jain, R.K. Assessment of genetic diversity within and among Basmati and non-Basmati rice varieties using AFLP, ISSR and SSR markers. *Euphytica* **2004**, *140*, 133–146. [[CrossRef](#)]
20. Jeung, J.U.; Hwang, H.G.; Moon, H.P.; Jena, K.K. Fingerprinting temperate japonica and tropical indica rice genotypes by comparative analysis of DNA markers. *Euphytica* **2005**, *146*, 239–251. [[CrossRef](#)]
21. Palomares-Rius, F.J.; Viruel, M.A.; Yuste-Lisbona, F.J.; López-Sesé, A.I.; Gómez-Guillamón, M.L. Simple sequence repeat markers linked to QTL for resistance to Watermelon mosaic virus in melon. *Theor. Appl. Genet.* **2011**, *123*, 1207. [[CrossRef](#)]
22. Li, Y.; Du, J.; Wang, T.; Shi, Y.; Song, Y.; Jia, J. Genetic diversity and relationships among Chinese maize inbred lines revealed by SSR markers. *Maydica* **2002**, *47*, 93–101.
23. Monforte, A.J.; Garcia-Mas, J.; Arús, P. Genetic variability in melon based on microsatellite variation. *Plant Breed.* **2003**, *122*, 153–157. [[CrossRef](#)]
24. Kwon, S.J.; Ahn, S.N.; Hong, H.C.; Hwang, H.G.; Choi, H.C. SSR diversity in Japonica rice cultivars and its association to several traits. *Korean J. Breed* **2002**, *34*, 57–63.
25. Song, M.T. Narrow genetic background of Korean rice germplasm as revealed by DNA fingerprinting with SSR markers and their pedigree information. *Korean J. Genet.* **2002**, *24*, 397–403.
26. Kwon, Y.S.; Park, E.K.; Park, C.U.; Bae, K.M.; Yi, S.I.; Cho, I.H. Identification of rice variety using simple sequence repeat (SSR) marker. *Life Sci.* **2006**, *16*, 1001–1005.
27. Singh, S.K.; Sharma, S.; Koutu, G.K.; Mishra, D.K.; Singh, P.; Prakash, V.; Kumar, V.; Namita, P. Genetic diversity in NPT lines derived from indica × japonica sub-species crosses of rice (*Oryza sativa* L.) using SSR markers. *Sch. J. Agric. Sci.* **2014**, *4*, 121–132.
28. Surapaneni, M.; Balakrishnan, D.; Mesapogu, S.; Raju, A.K.; Rao, Y.V.; Neelamraju, S. Genetic characterization and population structure of Indian rice cultivars and wild genotypes using core set markers. *3 Biotech* **2016**, *6*, 95. [[CrossRef](#)] [[PubMed](#)]
29. Babu, B.K.; Meena, V.; Agarwal, V.; Agrawal, P.K. Population structure and genetic diversity analysis of Indian and exotic rice (*Oryza sativa* L.) accessions using SSR markers. *Mol. Biol. Rep.* **2014**, *41*, 4329–4339. [[CrossRef](#)]
30. Zhang, P.; Li, J.; Li, X.; Liu, X.; Zhao, X.; Lu, Y. Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS ONE* **2011**, *6*, e2756. [[CrossRef](#)]
31. Nachimuthu, V.V.; Muthurajan, R.; Duraijalaguraja, S.; Sivakami, R.; Pandian, B.A.; Ponniah, G.; Gunasekaran, K.; Swaminathan, M.; Suji, K.K.; Sabariappan, R. Analysis of population structure and genetic diversity in rice germplasm using SSR markers: An initiative towards association mapping of agronomic traits in *Oryza sativa*. *Rice* **2015**, *8*, 30. [[CrossRef](#)]
32. Jin, L.; Lu, Y.; Xiao, P.; Sun, M.; Corke, H.; Bao, J. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor. Appl. Genet.* **2010**, *121*, 475–487. [[CrossRef](#)]
33. Singh, N.; Choudhury, D.R.; Tiwari, G.; Singh, A.K.; Kumar, S.; Srinivasan, K.; Tyagi, R.K.; Sharma, A.D.; Singh, N.K.; Singh, R. Genetic diversity trend in Indian rice varieties: An analysis using SSR markers. *BMC Genet.* **2016**, *17*, 127. [[CrossRef](#)] [[PubMed](#)]
34. Smith, J.S.C.; Chin, E.C.L.; Shu, H.; Smith, O.S.; Wall, S.J.; Senior, M.L.; Ziegler, J. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): Comparisons with data from RFLPs and pedigree. *Theor. Appl. Genet.* **1997**, *95*, 163–173. [[CrossRef](#)]
35. Rohlf, F.J. *NTSYSpc: Numerical Taxonomy and Multivariate Analysis System*, version 2.10b; Applied Biostatistics Inc.: New York, NY, USA, 2000.
36. Jaccard, P. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* **1908**, *44*, 223–270.
37. Kumar, S.; Tamura, K.; Jakobsen, I.B.; Nei, M. MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* **2001**, *17*, 1244–1245. [[CrossRef](#)] [[PubMed](#)]

38. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.
39. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).