Available online at www.sciencedirect.com

ScienceDirect

Rice Science, 2019, 26(6): 393−403

Research Paper

Genetic Diversity and Allelic Frequency of Selected Thai and Exotic Rice Germplasm Using SSR Markers

Wanwarang PATHAICHINDACHOTE[1, 2, 3], Natjaree PANYAWUT[4], Kannika SIKAEWTUNG[4], Sujin PATARAPUWADOL[1, 2, 5], Amorntip MUANGPROM[1, 2, 4]

([1]*Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand;* [2]*Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Education, Bangkok 10900, Thailand;* [3]*Department of Agricultural Sciences, Faculty of Agriculture, Natural Resources and Environment, Naresuan University, Phitsanulok 65000, Thailand;* [4]*National Center for Genetic Engineering and Biotechnology, Thailand Science Park, Pathum Thani 12120, Thailand;* [5]*Department of Plant Pathology, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom 73140, Thailand*)

Abstract: A collection of 167 Thai and exotic rice accessions was subjected for evaluation of genetic diversity and assessment of relationship by simple sequence repeat (SSR) markers. Among a total of 49 SSR markers, 13 markers distributing over 12 rice chromosomes showed clear polymorphic band patterns, and they were selected for genetic assessment. A total of 110 alleles were detected with an average of 8.46 alleles per locus. The averages of gene diversity, heterozygosity and polymorphic information content were 0.59, 0.02 and 0.56, respectively. The unweighted-pair group method with arithmetic averages (UPGMA) clustering analysis was performed for genetic distance, and phylogenetic tree was constructed. The result showed that this rice collection was divided into two major groups, classified as *japonica* and *indica* subspecies. Within the *japonica* group, *temperate japonica* and *tropical japonica* subgroups can be clearly separated. Three-dimensional principal component analysis projection and model-based population structure analysis showed consistent clustering results with two major groups of UPGMA analysis, supporting the classification of *japonica* and *indica* subspecies. The *indica* allelic frequency was also investigated to provide an indicative guide for breeders to overcome the practical problems on sterility of inter-subspecies hybrid offspring. This rice collection and information obtained in this study will be useful for rice breeding programs.

Key words: rice; genetic diversity; *indica*; *japonica*; allelic frequency; simple sequence repeat

Rice is one of the most important food crops in the world (Subudhi et al, 2006; Global Rice Science Partnership, 2013). Rice crop can be cultivated in several regions of the world due to its diversity and resourcefulness. However, only two species, *Oryza sativa* L., Asian cultivated rice, and *Oryza glaberrima* Steud., African cultivated rice, are generally cultivated (Linares, 2002; Maclean et al, 2002; Nayar, 2014). *O. sativa* is more widely grown than *O. glaberrima* because of its broad spectrum of growth climates, high-yielding features and market-favored characteristics (Chang, 1976; Jones et al, 2004; Subudhi et al, 2006).

Asian rice cultivars are divided into two major subspecies, *Oryza sativa* L. subsp. *indica* and *japonica* (Oka, 1958, 1988). *indica* is mainly lowland rice and typically grown in tropical and subtropical Asia, particularly in Thailand, India, Bangladesh, the Philippines and southern China. *japonica* is generally grown in

Received: 10 July 2018; **Accepted:** 6 November 2018
Corresponding author: Amorntip MUANGPROM (amorntip.mua@biotec.or.th)

Peer review under responsibility of China National Rice Research Institute
http://dx.doi.org/10.1016/j.rsci.2018.11.002

temperate or cooler climate regions such as East Asia, mountainous or high-altitude areas of Southeast Asia and South Asia (Glaszmann, 1987; Khush, 1997; Garris et al, 2005; Wang et al, 2014). However, some minor differences in several features such as morphological variations or geographical cultivated areas can further separate *japonica* into two subgroups, *temperate japonica* and *tropical japonica* (Mackill, 1995; Ni et al, 2002; Yonemaru et al, 2015; Hori et al, 2017). *temperate japonica* is generally cultivated in temperate regions such as China, Japan and Korea. *tropical japonica*, also known as *javanica*, is usually locally cultivated in upland areas of tropical regions such as mountain of Indonesia and the Philippines (Khush, 1997; Bhattacharya and Ali, 2015; Hori et al, 2017).

Hybridization between *indica* and *japonica* rice varieties is one of the most excellent applications for high-yielding rice production (Khush, 2001). Genetic relationships between *indica* and *japonica* subspecies are typically slightly interrelated; therefore, inter-subspecies hybridization will generate significant genetic recombination and variation (Lu et al, 2009). Successful inter-subspecies hybridization will result in greater yields of hybrids with significantly higher heterosis than that of inbred varieties (Virmani et al, 1982; Khush, 1995; Peng et al, 2004). Degree of heterosis depends on genetic difference between the pairs of parents; however, the limitation of utilizing the great heterosis from *indica* and *japonica* hybridization is a strong sterility trait of inter-subspecies hybridization (Oka, 1957; Virmani et al, 1982; Lu et al, 2009; Chen et al, 2013; Liu et al, 2015). Therefore, selection of appropriate pairs of parents becomes the most critical strategy for an effective rice breeding program. Interestingly, recent study revealed that the balance between a higher degree of heterosis and increased reproductive isolation is correlated to the genetic divergence index (GDI) of the parents. This investigation provides guidance for breeders to tackle sterility burdens of the hybrid offspring in rice breeding programs (Dan et al, 2014). Understanding of *indica-japonica* differentiation through molecular genotyping index based on *indica* or *japonica* allelic frequency may contribute to the effective selection of parent combinations for successful hybridization (Peng et al, 2004; Lu et al, 2009; Liu et al, 2015).

Thailand is located in a tropical climate region with differences in topography that is suitable for rice cultivation. Several Thai rice research institutes have put in a collective effort to compile a lot of native, cultivated and wild rice accessions in Thailand (Wuthiyano, 2000; Thatsanabanjong and Phoasavadi, 2017). In 2000, Thai Department of Agriculture reported that approximately 24 000 rice accessions are found in Thailand. They are collected and have been preserved in several Thai gene bank institutes for decades. There are approximately 17 000 rice accessions reported as native Thai rice. Among these native Thai rice collections, approximately 6 000 accessions are distinguished by their difference in names and phenotypes. However, evaluation of genetic divergence and relationship among Thai rice germplasm collections has been limited (Wuthiyano, 2000; Thatsanabanjong and Phoasavadi, 2017). The replacement of local rice varieties with a few market preferred varieties leads to dramatic decrease in the genetic diversity of rice crop in several regions due to a consequence from green revolution for decades. The collections of conserved local rice germplasm are valuable genetic resources for future rice improvement programs intended for agro- or market-preferential features (Heal et al, 2004; Sakamoto and Matsuoka, 2004).

Up to date, genetic diversity of several rice collections has been studied worldwide but only few Thai rice germplasm are represented in these studies. In the previous study, 43 accessions from Thai and 57 from International Rice Research Institute (IRRI) germplasm were clustered by 19 polymorphic InDel markers selected from a total of 98 InDel markers. Cluster, structure and differentiation analyses revealed six distinct groups corresponding to *indica* and *japonica* subspecies, which were quite well correlated to their ecologies and their development information (Chakhonkaen et al, 2012). Classification of the genetic variation and their relationship are critically required for enhancing the management and utilization of the germplasm resource (Ramanatha Rao and Hodgkin, 2002). Understanding of the genetic relationship within germplasm resource provides an insight for breeders to increase efficiency of parental selection for future rice breeding programs (Singh et al, 2016).

Several studies have shown application of using molecular markers for genetically diverse genotyping. Up to now, there are various molecular markers for studying genetic diversity such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), InDel and single nucleotide polymorphism (SNP) markers (Powell et al, 1996; Kanawapee et al, 2011; Chakhonkaen et al, 2012; Sutoro et al, 2015; Salem

and Sallam, 2016; Singh et al, 2016; Xu et al, 2016). Simple sequence repeat (SSR) marker, also known as microsatellite marker, is one of the most commonly used and powerful molecular markers for the genetic characterization in plants. SSR markers can efficiently facilitate the establishment of the genetic relationship due to their co-dominant inheritance, high level of allelic diversity, relative polymorphic abundance and extensive distribution across the genome (Powell et al, 1996; Mondini et al, 2009; Hoshino et al, 2012).

The objectives of this study were to study the genetic diversity and relationship of Thai upland landraces and some other rice varieties with special traits using SSR markers and to investigate *indica* allele frequency, providing an indicative guide for selecting parent combinations.

MATERIALS AND METHODS

Rice materials

A total of 167 rice accessions could be divided into two major groups according to the countries of origins (144 Thai rice accessions and 23 exotic rice accessions). Approximately 83% of Thai rice accessions was upland rice collected from nationwide. Of them, 82 accessions were labeled as upland non-colored rice and 38 were labeled as upland colored landraces. A total of 46 Thai colored rice accessions were used. In addition, 21 Thai lowland accessions were also included. The details of each accession are described in Supplemental Table 1.

Genomic DNA extraction and amplification

Total genomic DNA was extracted from 14-day-old young fresh leaves using the standard CTAB extraction protocol (Murray and Thompson, 1980). The genomic DNA samples were subjected to PCR amplification using SSR markers as primers. The PCR amplification was performed in 10 μL of reaction mixture, containing 15 ng genomic DNA, 1× *Taq* DNA polymerase buffer, 1.5 mmol/L $MgCl_2$, 0.2 mmol/L dNTPs, 0.25 U *Taq* DNA polymerase (Invitrogen, USA) and 1 mol/L of each primer. Thermal profiling was setup with initial denaturation temperature of 95 ºC for 5 min followed by 35 cycles of denaturation (95 ºC for 45 s), annealing (55 ºC for 45 s) and extension (72 ºC for 1 min), and a final extension (72 ºC for 8 min). PCR products (3 μL) were quality checked on 1.5% agarose gel electrophoresis. The amplified PCR products together with a 100 bp DNA ladder (Thermo Scientific, USA) were then size separated on the 3% vertical polyacrylamide gel electrophoresis, and DNA patterns were observed by silver staining (Bassam et al, 1991; Panaud et al, 1996).

SSR allele scoring and statistical analysis

The clear products of SSR-PCR were scored according to their allelic band patterns. The unique band positions of PCR products from each polymorphic SSR primers were coded by digital numbers depending on the numbers of the unique fragments. These assigned numbers were used as genotype codes for genotyping different co-dominant alleles. Accordingly, all rice germplasms were then scored for each polymorphic SSR markers as a result of bi-allelic combination. Numbers of genotypes, numbers of alleles, genetic diversity, heterozygosity and polymorphism information content (PIC) of each markers were calculated using PowerMarker version 3.25 (Liu and Muse, 2005).

A data matrix file for clustering analysis was generated by NTedit version 1.04. Similarity coefficients based on Simple Matching (SM) coefficients were estimated using SIMQUAL module in NTSYS-pc version 2.10e software (Rohlf, 1998). A phylogenetic dendrogram was then constructed using the unweighted-pair group method with arithmetic averages (UPGMA) clustering analysis, through SAHN module in NTSYS software. Principal component analysis (PCA) was also carried out using EIGEN module available in NTSYs software (Rohlf, 1998).

A model-based Bayesian analysis was also performed with Structure software version 2.3.4 (Pritchard et al, 2000). The admixture model and correlated allele frequency analysis were executed with the length of burn-in period set at 100 000 and number of Markov Chain Monte Carlo (MCMC) replications after burn-in set at 100 000. The population structure was analyzed for 10 iterations at each K value, when a number of populations (K) ranged from 1 to 10. The optimum K value was defined at a particular K giving a maximum K value according to an ad hoc statistic K suggested by Evanno et al (2005). The Evanno's Ks were determined based on the second order rate of change of the likelihood function with respect to K. The Evanno statistic parameters were calculated and demonstrated in a graphical plot using a web-based program Structure Harvester (Evanno et al, 2005; Earl and vonHoldt, 2012). The posterior probability of membership coefficient (Q) 0.8 was used to classify

rice germplasm into clusters (Travis et al, 2015).

Estimating allelic frequency

The SSR-PCR banding pattern of Non Rai, a representative *indica* rice variety for this rice collection, was used as a reference for estimating the *indica* genotype at a particular SSR locus. Frequency of *indica* alleles (F_i) of an inspected germplasm was calculated (Lu et al, 2009).

RESULTS

SSR marker-based divergence and molecular characterization

A total of 49 SSR markers were first used to test four rice accessions (Nipponbare, Azucena, KDML105 and PTT1) representing *temperate japonica*, *tropical japonica*, Thai breeding line and Thai improved line to identify the polymorphic SSR markers. Among these SSR markers, 13 SSR markers distributing across the 12 rice chromosomes produced clear single-locus polymorphic band patterns in at least three out of the four accessions. Thus, they were selected for genetic assessment of the 167 rice accessions. The details of the 13 SSR markers, including their sequences, product sizes and the repeat motifs, are given in Supplemental Table 2.

A total of 132 genotypes and 110 alleles were detected by 13 polymorphic SSR markers with an average of 10.2 genotypes per locus and 8.5 alleles per locus, respectively. The number of genotype per locus generated by a single marker varied from 4 (RM108) to 29 (RM1381), and the number of alleles per locus ranged from 4 (RM108, RM313 and RM588) to 19 (RM1381). The gene diversity ranged from 0.27 (RM542) to 0.87 (RM21) with an average of 0.59, and the average PIC value was found to be 0.56 with the range from 0.26 (RM542) to 0.86 (RM21). Heterozygosity varied from 0 (RM7, RM108, RM169 and RM317) to 0.07 (RM588) with an average of 0.02 (Table 1).

Genetic structure analysis

The genetic relationship among germplasm was estimated by calculating simple matching's similarity coefficients from all the 110 alleles based on their shared allele ratio. The resulting similarity coefficient matrix was used for UPGMA clustering analysis to generate a dendrogram to access the overall genetic relationships among the rice germplasm. The UPGMA dendrogram clustered the 167 rice germplasm into two major groups, G1 and G2, with additional subgroups as showed in Fig. 1.

The first group (G1) composed of 20 rice accessions including well known *japonica* rice varieties such as Nipponbare and Azucena. Nipponbare, a representative genotype for *temperate japonica* subspecies (G1.1), was clearly separated from Azucena, a representative genotype for *tropical japonica* subspecies (G1.2). IR68544-29-2-1-3-1-2 and Dawk Pah Yawm, which are also *tropical japonica*, were included in G1.2. The other rice germplasm classified in G1.2 are 1 upland colored landrace and 15 upland non-colored landraces.

The second major group (G2) comprised 147 rice accessions classified as *indica* cluster. G2 could be subdivided into six subgroups which consisted of three multi-genotypic subgroups (G2.1, G2.2 and G2.4), two mono-genotypic subgroups (G2.3 and G2.5) and one di-genotypic subgroup (G2.6). The subgroup G2.1 contained 26 rice accessions, of which 13 (50%) are exotic germplasms with special traits. The other 13 accessions included 7 Thai upland non-colored rice accessions, 3 Thai colored rice accessions and 3 Thai lowland elite improved lines. Approximately 56.5% of the exotic rice accessions were classified in G2.1 subgroup. The largest multi-genotypic subgroup G2.2 consisted of 102 rice accessions, of which 53 (52%) are Thai upland landraces. G2.3 and G2.5 were mono-genotypic subgroups which were composed of IR52 and Khiaw Nok Gra Ling, respectively. The subgroup G2.4 contained 15 rice accessions, of which 7 (46.7%) were be Thai colored landraces. However, two exotic accessions, i.e. Pokkali

Table 1. Genetic parameters of 13 simple sequence repeat markers in 167 rice accessions.

Marker	Chromosome	*Ng*	*Na*	*He*	*Ho*	PIC
RM283	1	7	5	0.39	0.02	0.35
RM5529	2	7	6	0.62	0.01	0.58
RM7	3	6	6	0.64	0.00	0.60
RM313	3	5	4	0.43	0.01	0.38
RM317	4	9	9	0.69	0.00	0.64
RM169	5	11	11	0.80	0.00	0.77
RM588	6	5	4	0.39	0.07	0.33
RM542	7	9	8	0.27	0.01	0.26
RM1381	8	29	19	0.83	0.07	0.82
RM108	9	4	4	0.58	0.00	0.51
RM496	10	8	7	0.44	0.01	0.40
RM21	11	19	16	0.87	0.02	0.86
RM247	12	13	11	0.77	0.01	0.75
Mean		10.2	8.5	0.59	0.02	0.56
Total		132	110			

Ng, Number of genotypes; *Na*, Number of alleles; *He*, Gene diversity; *Ho*, Heterozygosity; PIC, Polymorphic information content.

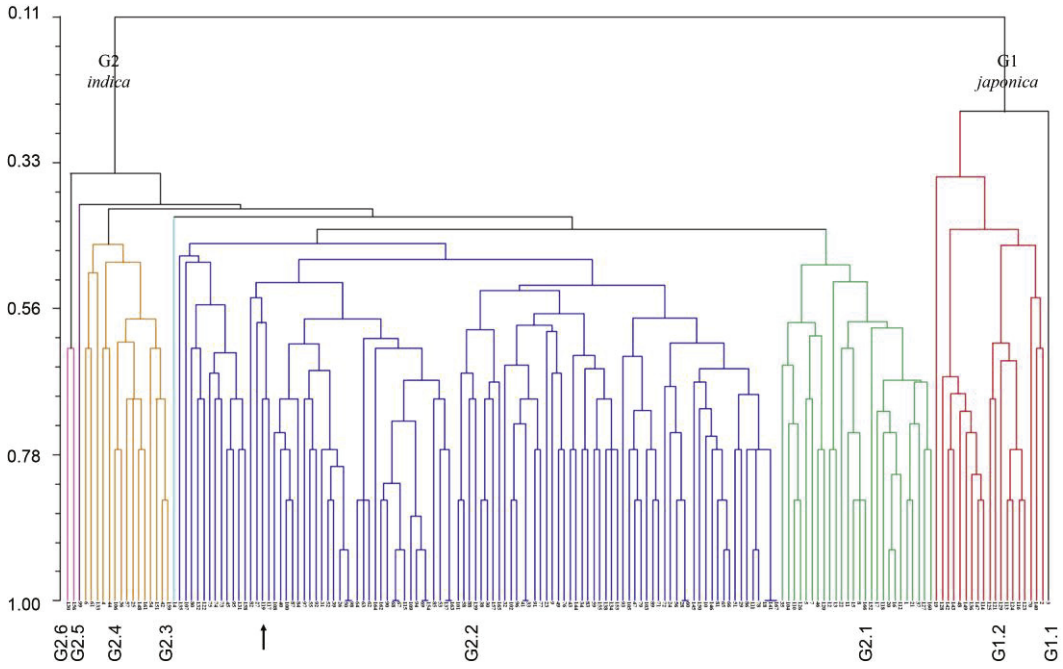

Fig. 1. Unweighted-pair group method with arithmetic averages cluster analysis of 167 rice accessions based on 13 SSR markers.

harboring the salt-resistant special trait and Taichung Native 1 with high-yielding and semi-dwarf traits, were grouped into this subgroup. The other six accessions were four upland landraces and two lowland landraces. The subgroup G2.6 was a di-genotypic group containing two accessions, Ngaw Pin (upland) and Khao Luang San Pah Tong (lowland).

PCA result revealed that the first three principle components explained 31.28% of the total genetic variation in the population. The projection of 3-dimentional PCA (Fig. 2) showed that two major distinct groups were clearly separated. Among 167 accessions, 20 accessions formed the G1 group together, obviously separating from the G2 group. Nipponbare and Azucena were composed in the PCA clustered G1 group, and Nipponbare was clearly distinct from Azucena. The other 18 rice accessions were grouped together with Azucena. The major group G2 consisted of 147 rice accessions and could be concisely classified as *indica* group. However, the subgroups within PCA clustering G2 group could not be apparently predicated.

The population structure based on model-based simulation showed the observed maximum Delta *K* value at $K = 2$ (Fig. 3). This suggested the genetic structure of 167 rice accessions had the most probable number of populations at $K = 2$ (two distinct populations). The first population assigned in red bars (Fig. 4) was composed of 20 rice accessions. The second population assigned in green bars was composed of 139 rice accessions. The admixture, showing estimated clustering membership coefficient (Q) below 0.8, was composed of Pokkali, Taichung Native 1, Lueang Noi, Niaw Ga Am Mah Ka, Ngaw Pin, Dok Mak, Dam Hawm and Khao Luang San Pah Tong.

Estimation of *indica* allelic frequency

The *indica* allelic frequencies of each individual were calculated based on scoring of fingerprint patterns generated by 13 SSR markers. Non Rai, indicated by the arrow in UPGMA phylogenetic dendrogram (Fig.

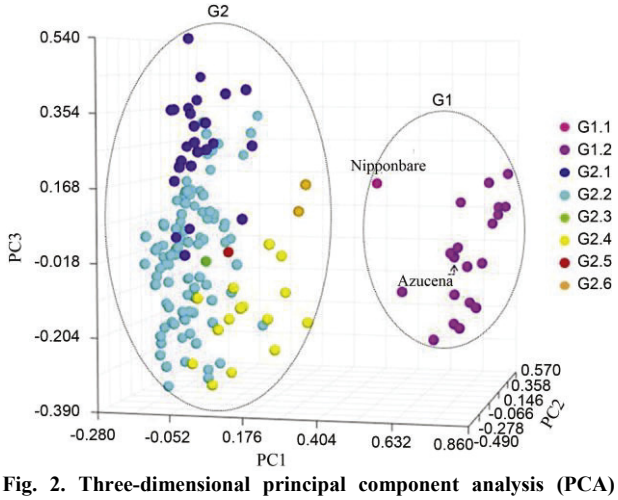

Fig. 2. Three-dimensional principal component analysis (PCA) projection of 167 rice accessions based on 13 SSR markers.

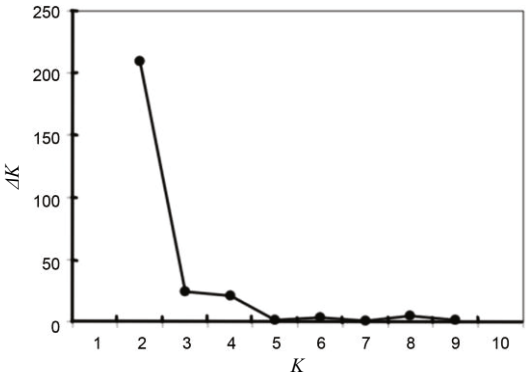

Fig. 3. Correlation between *K* and Delta *K* showing the maximum Delta *K* at *K* = 2.

1), was selected as a reference *indica* variety for *indica* allelic frequency estimation. According to di-allele form of co-dominant SSR marker of rice, if the homozygous *indica* allele appeared was scored as 2, the heterozygous *indica* allele appeared was scored as 1 and non *indica* allele appeared was scored as 0. *indica* allele frequency based on SSR loci of *indica* rice represented by Non Rai variety was calculated, ranging from 0.04 (Azucena) to 1.00 (Non Rai) as showed in Supplemental Table 3.

DISCUSSION

Although Thailand has divers rice germplasm collections, evaluations of genetic divergence and relationship among Thai rice germplasm have been limited (Wuthiyano, 2000; Thatsanabanjong and Phoasavadi, 2017). Most of the rice accessions used in this study are different from the germplasm previously reported (Chakhonkaen et al, 2012). The current set of rice germplasm composes of *indica* and *japonica* rice lines having diverse agronomic desirable traits such as high antioxidant, fragrance, resistance to diseases and insects, etc., which will be useful for future rice breeding programs. Understanding the genetic variation and their relationship among these germplasm is important for utilization of these germplasm for future rice improvements (Ramanatha Rao and Hodgkin, 2002; Singh et al, 2016).

Allelic richness and SSR marker analysis

SSR markers were applied for genotyping characterization because of their polymorphism abundance, simplicity and high reproducibility. The current study evaluated the genetic relatedness among a collection of 167 rice accessions of Thai and exotic rice germplasm using 13 polymorphic SSR markers. A total number of 132 genotypes with 110 alleles detected in the present study were slightly higher than 121 genotypes with 103 alleles of 75 cultivated Asian rice accessions using 32 polymorphic loci of RFLP markers reported by Sun et al (2001), probably due to genotypes of rice

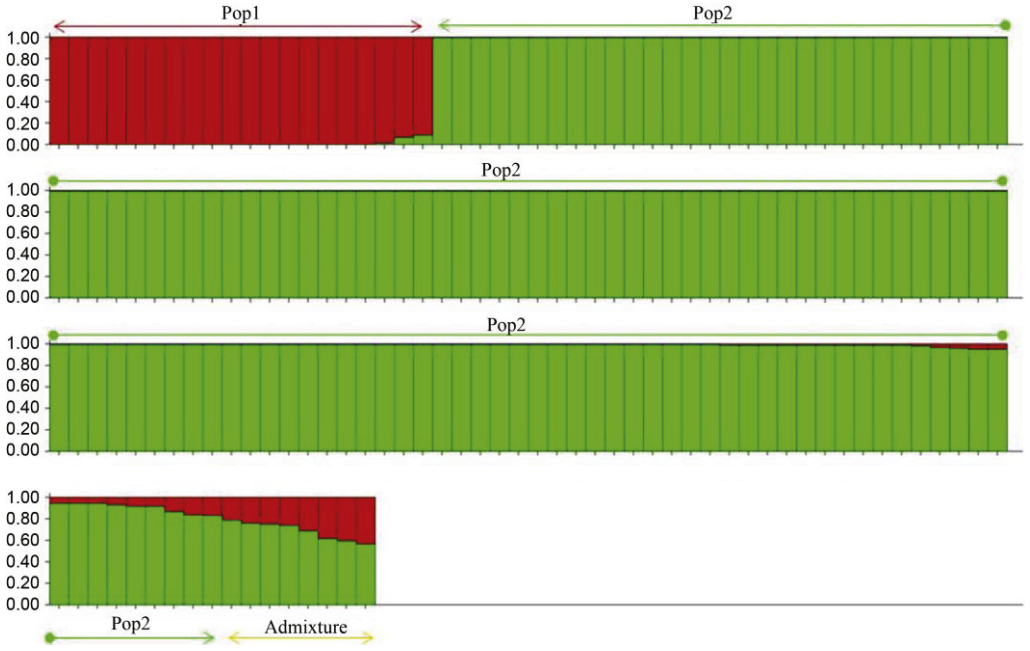

Fig. 4. Admixture model-based population structure based on *K* as 2.

Each thin bar represents one accession. Red color code represents *japonica* proportion of the individuals and green color code represents *indica* proportion of the individuals. The length of color code in each bar indicates the proportion of the estimated membership coefficient.

germplasm and markers used in the studies. The average number of allele detected in this study (8.5 per locus, ranging from 4 to 19) corresponded well with that obtained in the study of 69 Argentine rice accessions accessed by 26 SSR markers (8.4 per locus, ranging from 3 to 21) (Giarrocco et al, 2007), and slightly lower than that reported in the study of 21 Thai- and 9 exotic-rice cultivars using 20 SSR markers with an average of 9.5, ranging from 5 to 18 per locus (Kanawapee et al, 2011). A higher mean number of allele (11.8 per locus) was reported in the study of genetic structure and diversity among 234 rice accessions obtained from all over the world using 169 nuclear SSR markers (Garris et al, 2005). However, some studies reported lower numbers of allele per locus. For example, an average number of alleles (3 per locus) ranging from 2 to 7 was detected in the study of 192 rice accessions from India, Argentina, Bangladesh, Brazil, Bulgaria, China, Colombia, Indonesia, the Philippines, Uruguay, Venezuela and United States using 61 SSR markers (Nachimuthu et al, 2015). Lower average alleles per locus of 3.9, ranging from 2 to 12 per locus, were also reported in the study among 150 rice accessions 274 SSR markers (Zhang et al, 2011). The variation on the number of alleles per locus was possibly due to differences in rice genotypes, variability of markers used and methods to detect and score PCR products.

Gene diversity or expected heterozygosity is the most widely used parameter to estimate genetic variability within populations (Nei, 1973; Toro and Caballero, 2005). PIC values reflect the relative allelic polymorphism of a particular marker and their potential to differentiate the genotypes based on their genetic relationships (Guo and Elston, 1999). The results from our study with the average gene diversity of 0.59 (range of 0.27 to 0.87) and the average PIC of 0.56 (range of 0.26 to 0.86) agree well with several previous reports (Wang et al, 2013; Babu et al, 2014; Salem and Sallam, 2016). The average gene diversity and PIC value from the study of 151 Bangladesh rice accessions using 47 SSRs were 0.58 (range of 0.26 to 0.84) and 0.52 (range of 0.24 to 0.82), respectively (Wang et al, 2013). The average gene diversity at 0.55 (range of 0.02 to 0.80) and the average PIC at 0.50 (range of 0.02 to 0.77) were reported from the study of 82 rice accessions obtained from different parts of the Asian countries, including India, characterized by 39 SSR markers (Babu et al, 2014). The average gene diversity of 0.62 (range of 0.38 to 0.79) and the PIC value of 0.57 (range of 0.34 to 0.76) were reported in the study of 22 rice germplasm from India, the Philippines and Egypt using 23 SSRs (Salem and Sallam, 2016). All 13 polymorphic SSR markers used in this study are considered informative markers, of which 8 are highly informative (PIC > 0.5) and 5 are reasonably informative markers (0.25 < PIC < 0.50) (Botstein et al, 1980). The results suggest that the marker used in this study was suitable to assess the genetic relationship and fingerprinting studies among this rice collection.

Heterozygosity is one of the most popular parameters to determine the proportion of heterozygous individuals at a locus in populations (Liu and Muse, 2005). In the present study, the average heterozygosity value of 0.02 is slightly lower than the average heterozygosity (0.07) reported by Babu et al (2014). Interestingly, it is comparable with that (0.019) of 82 upland rice landraces from Java Island, Indonesia, using 16 SSR markers (Sutoro et al, 2015). A low level of heterozygosity with an average of 0.02 was likely due to inbreeding status and low genetic variability of individuals within this rice collection.

Genetic relationships revealed by SSR

The genetic relationship assessment among rice germplasm is a significant aspect to manage the rice collections and to explore the potential parent combinations. Information on genetic relatedness of germplasm collection could be used in selection of parental genotypes to broaden genetic material and improve rice varieties (Ramanatha Rao and Hodgkin, 2002). In the present study, the genetic relatedness based on informative SSR loci of the 167 rice accessions were assessed by three clustering approaches, which are distance-based UPGMA dendrogram, PCA projection and population structure simulation.

UPGMA is one of the simplest clustering algorithms to generate distance-based phylogenetic dendrogram. The generated UPGMA dendrogram relatively represented genetic diversity and relationships among germplasm within the collection (Sneath and Sokal, 1973). Based on the UPGMA dendrogram, a collection of 167 rice accessions was classified into two major groups corresponding to *japonica* and *indica* subspecies, similar to other reports (Giarrocco et al, 2007; Tang et al, 2010; Chakhonkaen et al, 2012; Zhang et al, 2013). The *japonica* major group was composed of two subgroups which corresponded well to *temperate japonica* and *tropical japonica* rice groups, whereas the *indica* major group was composed of six additional subgroups.

The detected six *indica* subgroups in this study was slightly higher than five *indica* subgroups in the study of Chinese landraces reported by Zhang et al (2013) and five *indica* subgroups in the study of exotic and Thai elite lines reported by Chakhonkaen et al (2012). This might be due to differences in landrace genotypes and markers. It should be remarked that some rice accessions which have identical primary native names might be different in their genotypes such as Khiaw Nok Gra Ling classified in subgroup G2.5, and Khiaw Nok Gra Ling classified in subgroup G2.2. In addition, Daw Dam obtained from the colored rice collection of Biotechnology Research and Development office, Department of Agriculture, Thailand, and Daw Dam which was as a very famous fragrance LOX3-null variety might have same primary native name but differences in their genotypes and phenotypes. Daw Dam in this study is colored rice which has green with dark purple stripe leaves, dark purple stalk, brown speckle pericarp (shell) and black grain characteristics while Daw Dam is not colored rice which has green leaves, light yellow pericarp (shell) and white grain characteristics according to Ma et al (2015). It should be noted that 'Dam' (in Thai) means 'black' (in English).

Based on the PCA projection, the first three principle components showed similar major grouping patterns to that inspected by UPGMA clustering with consistently clustered 167 rice accessions into two major groups of *japonica* and *indica* subspecies. The 20 accessions classified within the PCA-clustered *japonica* group were identical to those found within UPGMA-clustered *japonica* group. The two additional subgroups of *temperate japonica* and *tropical japonica* could be obviously differentiated within the major group *japonica* corresponding to those classified by UPGMA analysis. The additional subgroups within *indica* group were slightly formed and fairly similar to the additional *indica* subgroups classified by UPGMA clustering. In particular, Ngaw Pin and Khao Luang San Pah Tong, which formed a di-genotypic G2.6 assessed by UPGMA clustering, were slightly separated. However, the other additional *indica* subgroups assessed by PCA clustering were not clearly determined.

In addition, the results of model-based structure analysis are in agreement with the major clustering by UPGMA and PCA analysis. The population structure analysis clustered a collection of 167 rice accessions into two major groups classified as *indica* and *japonica* subspecies. The rice accessions classified in *japonica* group assessed by population structure analysis were corresponding to those assessed by UPGMA and PCA analysis. Apart from 20 accessions of the *japonica* and 139 accessions of the *indica*, 8 accessions were defined as the admixtures. Of them, Dam Hawm, an upland rice landrace from southern of Thailand, was classified in the UPGMA-clustered *indica* subgroup G2.2. Five accessions, Pokkali, Taichung Native 1, Lueang Noi, Niaw Ga Am Mah Ka and Dok Mak, were the members of the UPGMA-clustered *indica* subgroup G2.4. The other two admixture accessions, Ngaw Pin and Khao Luang San Pah Tong, were the members of the di-genotypic G2.6 *indica* subgroup assessed by UPGMA clustering.

In rice global classification, Glaszmann (1987) suggested six varietal rice groups namely *indica*, *japonica*, *aus*, *aromatic*, *rayada* and *ashina* based on variation of 15 isozyme loci. Garris et al (2005) proposed five distinct rice groups corresponding to *indica*, *temperate japonica*, *tropical japonica*, *aus* and *aromatic* rices, which become the most commonly accepted classification of global rice cultivars. However, some classification studies of rice landraces were found to be different in their subdivisions (Zhang et al, 2007; Jin et al, 2010; Zhang et al, 2013). Zhang et al (2007) reported seven subpopulations of rice landraces from Guizhou, China. Jin et al (2010) reported seven subpopulations among 416 rice landraces and breeding lines collected mostly in China. Zhang et al (2013) reported eight subdivisions from two major clusters of *indica* and *japonica* rice in the study of lowland and upland landraces from China. In the current study, three clustering analysis consistently classified the 167 rice accessions into two major clusters corresponding to *japonica* and *indica* subspecies. Distance-based UPGMA analysis detected a total of eight subgroups similar to that reported by Zhang et al (2013). However, the numbers of the internal subdivisions within each major cluster were different. There were two *japonica* subgroups and six *indica* subgroups detected in this study, while there were three *japonica* subgroups and five *indica* subgroups in the report by Zhang et al (2013). Our UPGMA and PCA analysis showed two subdivisions *temperate japonica* and *tropical japonica* within *japonica* subspecies. This might be due to the difference in the markers and the germplasm. Regarding to rice global classification, the reference *aus* and *aromatic* rice were not comprised in this study. However, all the seven fragrance landraces were grouped together

within the largest *indica* UPGMA-clustered G2.2. Understanding the genetic diversity and genetic relatedness among germplasm genotypes will be useful for selecting parents to broaden genetic basis of breeding materials and to increase heterosis in inter-subspecies hybridization programs.

Clustering results from UPGMA, PCA and population structure analyses indicated the efficacy of this set of co-dominant SSR markers to primarily discriminate the *indica* and *japonica* subspecies. *japonica* and *indica* allelic frequencies of each germplasm based on several reference varieties were primarily calculated and compared. Nipponbare, Azucena and Dawk Pah Yawm were firstly selected as the references for *japonica* allelic frequency estimation. However, the density histograms of the estimated *japonica* allelic frequencies based on these reference varieties showed strong asymmetric right-skewed distributions. This might be due to the fact that *japonica* subspecies were the minor genotypes in this rice collection. The *indica* allelic frequency estimation was then performed because approximately 88% of rice germplasm in this collection are *indica* subspecies. Estimation of the *indica* allelic frequencies was attempted based on several preliminary reference *indica* rice varieties. Of them, the density histograms of the estimated *indica* allelic frequencies based on Non Rai as a reference presented the greatest symmetric normal distribution. Therefore, Non Rai, classified in the UPGMA-clustered largest *indica* subgroup G2.2, was designated as a reference variety for *indica* allelic frequency estimation in the current study.

Estimation of *indica* allele frequency

Selection of a particular parent pair was one of the key steps in inter-subspecies hybridization programs because of inversion between heterosis and fertility of the hybrid offspring. A strategic design for breeding inter-subspecies hybrid rice to balance between a higher degree of heterosis and increased reproductive isolation had been proposed by researchers in China (Danetal, 2014). They suggested that moderate GDI (approximately 0.37 in their case) is optimal in balancing between the great heterosis and fertility (Dan et al, 2014). Information obtained from an estimation of *indica* allele frequency might be used to select the effective parent pairs with moderated GDI to balance between a greater degree of heterosis and fertility (Lu et al, 2009; Wang et al, 2010; Dan et al, 2014; Liu et al, 2015). Understanding genetic divergence between the breeding materials in hand would be helpful for efficient selection of parent combinations and for proficient utilization of rice germplasm resources.

In this study, the SSR markers were used to assess the genetic relatedness within 167 rice accessions mainly Thai upland non-colored and colored landraces, and some Thai lowland and exotic rice varieties having various special traits. These rice accessions were grouped into two major groups with additional subgroups. The two major groups could be noticeably classified as *indica* and *japonica* subspecies. Within the *japonica* group, *temperate japonica* and *tropical japonica* subgroups could be clearly distinguished. *indica* allelic frequencies based on fingerprint of SSR markers were also calculated. This rice collection will be useful as sources for favorable alleles in future breeding programs. The information obtained from this study might be useful for selecting parental pairs for future rice improvement programs to overcome the obstacle in strong sterility of the hybrid offspring and to broaden genetic base of breeding materials.

ACKNOWLEDGEMENTS

The authors thankfully acknowledge the Rice department, Department of Agriculture, National Outstanding Farmer and International Rice Research Institute for providing the rice germplasm, and grateful to Dr. Peera JARUAMPORNPAN for her critical reading and comments on the manuscript. This study was supported by the Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Education (AG-BIO/PERDO-CHE); Agricultural Research Development Agency (ARDA); and National Science and Technology Development Agency in Thailand.

SUPPLEMENTAL DATA

The following materials are available in the online version of this article at http://www.sciencedirect.com/science/journal/16726308; http://www.ricescience.org.

Supplemental Table 1. Characteristics of rice accessions used in the study.

Supplemental Table 2. Information of thirteen simple sequence repeat markers.

Supplemental Table 3. Frequency of *indica* allele (F_i)

calculated based on fingerprint profiles of 13 SSR markers.

REFERENCES

Babu B K, Meena V, Agarwal V, Agrawal P K. 2014. Population structure and genetic diversity analysis of Indian and exotic rice (*Oryza sativa* L.) accessions using SSR markers. *Mol Biol Rep*, **41**(7): 4329–4339.

Bassam B J, Caetano-Anolles G, Gresshoff P M. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem*, **196**(1): 80–83.

Bhattacharya K R, Ali S Z. 2015. An Introduction to Rice-Grain Technology. New Delhi, India: Woodhead Publishing India Pvt. Ltd.

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**(3): 314–331.

Chakhonkaen S, Pitnjam K, Saisuk W, Ukoskit K, Muangprom A. 2012. Genetic structure of Thai rice and rice accessions obtained from the International Rice Research Institute. *Rice*, **5**: 19–31.

Chang T T. 1976. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. *Euphytica*, **25**(1): 425–441.

Chen C, Chen H, Shan J X, Zhu M Z, Shi M, Gao J P, Lin H X. 2013. Genetic and physiological analysis of a novel type of interspecific hybrid weakness in rice. *Mol Plant*, **6**(3): 716–728.

Dan Z W, Liu P, Huang W C, Zhou W, Yao G X, Hu J, Zhu R S, Lu B R, Zhu Y G. 2014. Balance between a higher degree of heterosis and increased reproductive isolation: A strategic design for breeding inter-subspecific hybrid rice. *PLoS One*, **9**(3): e93122.

Earl D A, vonHoldt B M. 2012. Structure Harvester: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*, **4**(2): 359–361.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol*, **14**(8): 2611–2620.

Garris A J, Tai T H, Coburn J, Kresovich S, McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics*, **169**(3): 1631–1638.

Giarrocco L E, Marassi M A, Salerno G L. 2007. Assessment of the genetic diversity in argentine rice cultivars with SSR markers. *Crop Sci*, **47**(2): 853–858.

Glaszmann J C. 1987. Isozymes and classification of Asian rice varieties. *Theor Appl Genet*, **74**(1): 21–30.

Global Rice Science Partnership. 2013. Rice Almanac. Los Baños, the Philippines: International Rice Research Institute.

Guo X, Elston R C. 1999. Linkage information content of polymorphic genetic markers. *Hum Hered*, **49**(2): 112–118.

Heal G, Walker B, Levin S, Arrow K, Dasgupta P, Daily G, Ehrlich P, Maler K G, Kautsky N, Lubchenco J, Schneider S, Starrett D. 2004. Genetic diversity and interdependent crop choices in agriculture. *Resour Energy Econ*, **26**(2): 175–184.

Hori K, Yamamoto T, Yano M. 2017. Genetic dissection of agronomically important traits in closely related temperate *japonica* rice cultivars. *Breeding Sci*, **67**(5): 427–434.

Hoshino A A, Bravo J P, Nobile P M, Morelli K A. 2012. Microsatellites as tools for genetic diversity analysis, genetic diversity in microorganisms. InTechOpen: 382.

Jin L, Lu Y, Xiao P, Sun M, Corke H, Bao J S. 2010. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor Appl Genet*, **121**(3): 475–487.

Jones M P, Dingkuhn M, Alukosnm G K, Semon M. 2004. Interspecific *Oryza sativa* L. X O. *Glaberrima Steud*. progenies in upland rice improvement. *Euphytica*, **94**(2): 237–246.

Kanawapee N, Sanitchon J, Srihaban P, Theerakulpisut P. 2011. Genetic diversity analysis of rice cultivars (*Oryza sativa* L.) differing in salinity tolerance based on RAPD and SSR markers. *Electron J Biotechnol*, **14**(6): 1–17.

Khush G S. 1995. Breaking the yield frontier of rice. *Geo Journal*, **35**(3): 329–332.

Khush G S. 1997. Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol*, **35**: 25–34.

Khush G S. 2001. Green revolution: The way forward. *Nat Rev Genet*, **2**(10): 815–822.

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

Liu K J, Muse S V. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics*, **21**(9): 2128–2129.

Liu P, Dan Z W, Wang Z, Li S Q, Li N W, Yan H X, Cai X X, Lu B R. 2015. Predicting hybrid fertility from maker-based genetic divergence index of parental varieties: Implications for utilizing inter-subspecies heterosis in hybrid rice breeding. *Euphytica*, **203**(1): 47–57.

Lu B R, Cai X X, Xin J. 2009. Efficient *indica* and *japonica* rice identification based on the InDel molecular method: Its implication in rice breeding and evolutionary research. *Prog Nat Sci*, **19**(10): 1241–1252.

Ma L, Zhu F G, Li Z W, Zhang J F, Li X, Dong J L, Wang T. 2015. TALEN-based mutagenesis of lipoxygenase LOX3 enhances the storage tolerance of rice (*Oryza sativa*) seeds. *PLoS One*, **10**(12): e0143877.

Mackill D J. 1995. Classifying *japonica* rice cultivars with RAPD markers. *Crop Sci*, **35**(3): 889–894.

Maclean J L. 2002. Rice Almanac. England: CABI.

Mondini L, Noorani A, Pagnotta M A. 2009. Assessing plant genetic diversity by molecular tools. *Diversity*, **1**(1): 19–35.

Murray M G, Thompson W F. 1980. Rapid isolation of high molecular weight plant DNA. *Nucl Acids Res*, **8**(19): 4321–4325.

Nachimuthu V V, Muthurajan R, Duraialaguraja S, Sivakami R, Pandian B A, Ponniah G, Gunasekaran K, Swaminathan M, Suji K K, Sabariappan R. 2015. 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*, **8**(1): 30.

Nayar N M. 2014. *Oryza* species and their interrelationships. *In*: Origin and Phylogeny of Rices. San Diego: Academic Press: 59–115.

Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA*, **70**(12): 3321–3323.

Ni J J, Colowit P M, Mackill D J. 2002. Evaluation of genetic

diversity in rice subspecies using microsatellite markers. *Crop Sci*, **42**(2): 601–607.

Oka H I. 1957. Genic analysis for the sterility of hybrids between distantly related varieties of cultivated rice. *J Genet*, **55**(3): 397–409.

Oka H I. 1958. Intervarietal variation and classification of cultivated rice. *Ind J Genet Plant Breeding*, **18**: 79–89.

Oka H I. 1988. Origin of Cultivated Rice. Amsterdam: Elsevier.

Panaud O, Chen X, McCouch S R. 1996. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet*, **252**(5): 597–607.

Peng S, Laza R C, Visperas R M, Khush G S, Virk P, Zhu D. 2004. Rice: Progress in breaking the yield ceiling. New directions for a diverse planet. *In*: Proceedings of the 4th International Crop Science Congress. Australia: Brisbane.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breeding*, **2**(3): 225–238.

Pritchard J K, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics*, **155**(2): 945–959.

Ramanatha Rao V, Hodgkin T. 2002. Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell Tissue Organ Cult*, **68**(1): 1–19.

Rohlf F J. 1998. NTSYSpc numerical taxonomy and multivariate analysis system version 2.0 user guide. Setauket, New York: Applied Biostatistics Inc.

Sakamoto T, Matsuoka M. 2004. Generating high-yielding varieties by genetic manipulation of plant architecture. *Curr Opin Biotechnol*, **15**(2): 144–147.

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

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. 2016. Genetic diversity trend in Indian rice varieties: An analysis using SSR markers. *BMC Genet*, **17**(1): 127.

Sneath P H A, Sokal R R. 1973. Numerical Taxonomy. San Francisco, California, USA: W.H.: Freeman and Company.

Subudhi P K, Sasaki T, Khush G S. 2006. Rice, Cereals and Millets. Berlin, Heidelberg: Springer.

Sun C Q, Wang X K, Li Z C, Yoshimura A, Iwata N. 2001. Comparison of the genetic diversity of common wild rice (*Oryza rufipogon* Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. *Theor Appl Genet*, **102**(1): 157–162.

Sutoro, Lestari P, Kurniawan H. 2015. Genetic diversity of upland rice landraces from java island as revealed by SSR markers. *Indones J Agric Sci*, **16**(1): 1–10.

Tang S, Zhang Y, Zeng L, Luo L, Zhong Y, Geng Y. 2010. Assessment of genetic diversity and relationships of upland rice accessions from southwest China using microsatellite markers. *Plant Biosyst*, **144**(1): 85–92.

Thatsanabanjong K, Phoasavadi P. 2017. The music of rice in Amphawa. *J Silpakorn Univ: Veridian E*, **10**(5): 492–505.

Toro M A, Caballero A. 2005. Characterization and conservation of genetic diversity in subdivided populations. *Philos Trans R Soc Lond B Biol Sci*, **360**: 1367–1378.

Travis A J, Norton G J, Datta S, Sarma R, Dasgupta T, Savio F L, Macaulay M, Hedley P E, McNally K L, Sumon M H, Islam M R, Price A H. 2015. Assessing the genetic diversity of rice originating from Bangladesh, Assam and West Bengal. *Rice*, **8**(1): 35.

Virmani S S, Aquino R C, Khush G S. 1982. Heterosis breeding in rice (*Oryza sativa* L.). *Theor Appl Genet*, **63**(4): 373–380.

Wang C H, Zheng X M, Xu Q, Yuan X P, Huang L, Zhou H F, Wei X H, Ge S. 2014. Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm. *Heredity*, **112**(5): 489–496.

Wang M M, Zhu Z F, Tan L B, Liu F X, Fu Y C, Sun C Q, Cai H W. 2013. Complexity of *indica-japonica* varietal differentiation in Bangladesh rice landraces revealed by microsatellite markers. *Breeding Sci*, **63**(2): 227–232.

Wang Y R, Qiu F L, Hua Z T, Dai G J. 2010. Relationship of parental *indica-japonica* indexes with yield and grain quality traits of *japonica* hybrid rice in Northern China. *Rice Sci*, **17**(3): 199–205.

Wuthiyano C. 2000. Thai Rice Landraces. Pathum Thani, Thailand: Department of Agriculture.

Xu Q, Yuan X P, Wang S, Feng Y, Yu H Y, Wang Y P, Yang Y L, Wei X H, Li X M. 2016. The genetic diversity and structure of *indica* rice in China as detected by single nucleotide polymorphism analysis. *BMC Genet*, **17**: 53.

Yonemaru J I, Choi S H, Sakai H, Ando T, Shomura A, Yano M, Wu J, Fukuoka S. 2015. Genome-wide indel markers shared by diverse Asian rice cultivars compared to Japanese rice cultivar 'Koshihikari'. *Breeding Sci*, **65**(3): 249–256.

Zhang D L, Zhang H L, Wei X H, Qi Y W, Wang M X, Sun J L, Ding L, Tang S X, Qiu Z E, Cao Y S, Wang X K, Li Z C. 2007. Genetic structure and diversity of *Oryza sativa* L. in Guizhou, China. *Chin Sci Bull*, **52**(3): 343–351.

Zhang L N, Cao G L, Han L Z. 2013. Genetic diversity of rice landraces from lowland and upland accessions of China. *Rice Sci*, **20**(4): 259–266.

Zhang P, Li J Q, Li X L, Liu X D, Zhao X J, Lu Y G. 2011. Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS One*, **6**(12): e27565.

(Managing Editor: FANG Hongmin)