Genetic structure revealed by a whole-genome single-nucleotide polymorphism survey of diverse accessions of cultivated Asian rice (*Oryza sativa* L.)

Kaworu Ebana1, Jun-ichi Yonemaru1, Shuichi Fukuoka1, Hiroyoshi Iwata2,4, Hiroyuki Kanamori3, Nobukazu Namiki1,6, Hideki Nagasaki1,6 and Masahiro Yano*4

1) QTL Genomics Research Center, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan
2) Data Mining and Grid Research Team, National Agricultural Research Center, 3-1-1 Kannondai, Tsukuba, Ibaraki 305-8666, Japan
3) Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, Tsukuba, Ibaraki 305-0854, Japan
4) Present address: Laboratory of Biometry and Bioinformatics, Graduate School of Agriculture and Life Sciences, University of Tokyo, 1-1-1 Yajo, Bunkyo, Tokyo 113-8657, Japan
5) Present address: Mitsubishi Space Software Inc., 1-6-1 Takezono, Tsukuba, Ibaraki 305-0032, Japan
6) Present address: National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan

To reveal the sequence diversity and population structure of Asian rice (*Oryza sativa* L.) cultivars, we surveyed genome-wide single-nucleotide polymorphisms (SNPs) in 140 diverse accessions. We identified 4357 SNPs distributed on the 12 chromosomes by sequencing PCR amplicons from the exons and introns of anonymous rice genes. We detected 4.87 SNPs per 1 kb genome-wide. By classifying the 140 accessions on the basis of these SNPs, we identified seven cultivar groups that reflected the geographical distribution of the accessions. Three cultivar groups were defined from *tropical japonica* that corresponded to previous categories, and three *indica* cultivar groups were also defined within *indica*. The linkage disequilibrium (LD) distance between SNPs was approximately 250 kb, except for a longer LD detected in the Indica I cultivar group (corresponding to the previously identified *aus* group). The allele frequency of the SNPs varied among cultivar groups, reflecting the level of genetic diversity in each group. These SNPs for the diverse accessions enhance our understanding of natural variation in rice.

Key Words: *Oryza sativa*, single-nucleotide polymorphisms, allele frequency, cultivar group, core collection.

Introduction

A wide range of naturally occurring phenotypic variations have been observed among rice (*Oryza sativa* L.) accessions, in both cultivars and their wild relatives (Jawahar and Panwar 1970, Matsuo 1952, McKenzie et al. 1994, Mochida et al. 2009, Quijano-Guerta et al. 2002, Ueno et al. 2009, Uga et al. 2009). Analyses of molecular markers have also detected a considerable number of sequence variations in diverse cultivars (Sun et al. 2002, Wang et al. 1992, Yan et al. 2007, Zeng et al. 2007). Recent progress in rice genomics has enhanced our understanding of these natural variations, which are potential sources of useful genes for improving rice cultivars (Yamamoto et al. 2009). To assess and explore such variations effectively using genetic analyses, a collection of diverse germplasms consisting of a relatively small number of accessions representative of the variation in the natural rice population should be assembled. For this purpose, the United States Department of Agriculture has developed a core rice germplasm collection (Yan et al. 2007). Moreover, the National Institute of Agrobiological Sciences (NIAS) has established two mini-core collections, each consisting of a small number of accessions that together cover more than 90% of the phenotypic and genotypic variation in rice accessions (Ebana et al. 2008, Kojima et al. 2005). These accessions have been used effectively for genetic analysis of quantitative traits and association mapping in rice (Iwata et al. 2009, Ueno et al. 2009, Uga et al. 2009).

In the last decade, the whole genomes of two rice cultivars, Nipponbare and 93-11, have been sequenced (International Rice Genome Sequencing Project 2005, Yu et al. 2002). The tremendous amount of available sequence information made available by these projects has allowed us to identify single nucleotide polymorphisms (SNPs) between these two cultivars (Feltus et al. 2004, Shen et al. 2004). This information has already made further genetic analysis possible, such as quantitative trait locus mapping and positional cloning of genes of interest. However, too few rice accessions have been completely sequenced to cover the full range of rice diversity. To facilitate functional analysis of rice genes, genome-wide sequence variations among diverse
accessions should be analyzed. The International Rice Functional Genomics Consortium has recently re-sequenced 20 cultivars (McNally et al. 2009), providing insights into nucleotide polymorphisms in a diverse germplasm set, but these cultivars are too few to represent the full diversity of existing rice accessions. To reveal sequence variations and rice population structure, the sequence polymorphisms of a more diverse set of accessions must be analyzed.

In this study, to clarify the population structure and to understand the nature of sequence variations in Asian rice germplasms, we performed a genome-wide discovery of SNPs in 140 diverse accessions, including accessions from both a world core collection and a Japanese core collection (Ebana et al. 2008, Kojima et al. 2005). We defined seven cultivar groups by using the genome-wide SNPs detected by comparing all these accessions. We then analyzed the allele frequencies of the detected SNPs and the genome-wide linkage disequilibrium in the defined cultivar groups. We expect this information to be useful for analyses of the genetic diversity of Asian rice accessions.

Materials and Methods

Plant materials

The accessions were selected from a rice diversity research set (Kojima et al. 2005) and a mini-core collection of Japanese rice accessions (Ebana et al. 2008). We then added other cultivars to include elite cultivars, parental cultivars of experimental lines such as chromosome segment-substitution lines, and other important accessions to allow comparison of our results with those of previous studies of genetic diversity in rice (Supplemental Table 1). Among these 140 accessions, which originated mainly from different regions of Asia (except for seven accessions from other areas, see Supplemental Table 1), 63 belong to the japonica cultivar group A of Kojima et al. (2005), and 52 and 25 belong to indica cultivar groups B and C, respectively (Supplemental Table 1 and Supplemental Fig. 1). We chose these numbers of accessions in each cultivar group to increase the probability of detecting specific minor alleles with a frequency of less than 10% within each group. For such an analysis, at least 10 accessions in each group would typically be necessary. Together, these accessions represent diverse geographical origins and exhibit considerable variation in morphological and agronomical characteristics (Kojima et al. 2005).

Sequence analysis

To obtain a large amount of total DNA from each accession, bulk total DNA was extracted from young leaves of 10 plants in each accession by the Cetyltrimethyl ammonium bromide method (Murray and Thompson 1980). Target sites for SNP discovery were positioned at predicted genes, randomly selected from Nipponbare sequences in the Rice Annotation Project Database (RAP-DB, http://rapdb.dna.affrc.go.jp/), to cover the entire genome at intervals of 100 to 300 kb. By using the Nipponbare reference sequence and Primer3 software (Rozen and Skaletsky 2000), primers were designed at two exons flanking an intron region to produce a PCR product of 500 to 700 bp. The specificity of the sequences of a designed primer pair was verified with the BLASTN program (Altschul et al. 1990) by comparing the forward and reverse primer sequences with sequences from the whole genome. The amplification of a single fragment by each primer pair was also confirmed by agarose gel electrophoresis of the PCR amplicon.

PCR amplification was performed in a 20-µl reaction volume with a reaction mixture containing 20 to 50 ng genomic DNA, 2 μl 10× PCR buffer, 2 μl dNTP mix (2 mM), 1 μl of each primer (5 μM), and 0.1 μl Ex Taq DNA polymerase (Takara Bio Inc., Otsu, Japan). The PCR program was 3 min at 95.0°C; 30 cycles of 30 s at 94.0°C, 30 s at 60.0°C, and 1 min at 72.0°C; and 5 min at 72.0°C. Amplified products were purified with shrimp alkaline phosphatase and exonuclease I (GE Life Sciences, Little Chalfont, UK) according to the manufacturer’s specifications. After purification, amplicons were sequenced with a Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA) and analyzed with ABI3730xl capillary sequencers (Applied Biosystems). Phred Standard chromatogram files of quality-trimmed sequences (http://www.phrap.org/phredphrapconsed.html) (Ewing et al. 1998) were used for base calling and quality trimming of the ABI chromatograms. Only sequences > 50 bp in length and with Ph/Pr > 16 over 100 continuous base pairs were used in the sequence analyses.

SNP detection

The sequences were aligned by using version 5 of the MAFFT software (http://mafft.cbrc.jp/alignment/software/) with the L-INS-i option (Katoh et al. 2005). After alignment, the sequences were validated by manual inspection of the trace files and corrected as necessary. BLASTN was used to align Nipponbare amplicon sequences with the IRGSP Build 4 pseudomolecule assembly (http://rgp.dna.affrc.go.jp/E/IRGSP/Build4/build4.html). After the Nipponbare sequences were clearly aligned to the reference sequence, the PCR products were mapped to the target genomic regions. Sequences not aligned to the Nipponbare pseudomolecules were excluded from the analysis. SNPs and indels were detected based on the comparison between sequences of the PCR products and those of predicted genes in the Nipponbare genome.

Linkage disequilibrium estimation

We calculated Δ² of each detected SNP to determine the chromosome-wide linkage disequilibrium (LD). Δ² is thought to be equivalent to r² (Hill and Weir 1994), and we calculated its value following Devlin and Risch (1995) using a Javascript program on the Linux platform. We performed this analysis only for regions for which sequence data were obtained from at least 80 rice accessions.
Analysis of the population structure

To limit the effect of linkages between SNPs, five randomly selected SNP sets, each comprising 120 SNPs, were used for the structure analysis. After a burn-in of 5000 iterations, 10 cycles of 100,000 iterations each were performed with version 2.03 of the STRUCTURE software (Pritchard et al. 2000) using admixture models with no linkage. Accessions were classified according to the average result of the 10 trials. We decided the inferred $K$ value from estimated value of log $p(x/k)$. The lower value of log $p(x/k)$, obtained using a simplified model in which most individuals did not have mixed ancestry, was adopted for the solution of the population structure.

A dendrogram was generated with version 11.5 of the DARwin software (http://darwin.cirad.fr/darwin) from a pairwise-distance matrix using the neighbor-joining method (Saitou and Nei 1987) to reveal the genetic relationships among the accessions. The pairwise fixation index ($Fst$) values among the cultivar groups were calculated with version 3.11 of the Arlequin software (Excoffier et al. 2005).

Results

SNP discovery

The 140 accessions used in this study included 59 accessions from world rice collections (Kojima et al. 2005) and 50 accessions from a Japanese rice collection (Ebana et al. 2008) (Supplemental Table 1 and Supplemental Fig. 1). In addition, 31 accessions were newly selected, mainly from the japonica cultivar group, in order to cover rice variations more comprehensively. The 140 accessions originated from different regions of Asia (133), from the United States (2), from Brazil (2), from Peru (1) and from Africa (2) (Supplemental Fig. 1).

Among the 3103 tested primers based on predicted genes, no PCR product was obtained with 1104 primers, and the products obtained with 421 primers could not be successfully sequenced. We successfully detected SNPs in the remaining 1578 amplicons from the 140 accessions (Table 1). We examined alignments over a mean of 567 bp per amplicon (range of 220 to 762 bp), representing all 125.3 Mb of the rice genome. By aligning the amplicon sequences for the 140 rice accessions, we were able to discern a total of 4357 SNPs among the accessions. We detected an average of 3.21 SNPs per sequenced site (4.87 SNPs per 1 kb). We observed five chromosomal regions with relatively few SNPs (Fig. 1): the long arm of chromosome 4, and the short arms of chromosomes 7, 9, 11, and 12. In contrast, we detected relatively large numbers of SNPs on the short arm of chromosome 6.

The allele frequency of each SNP in these 140 accessions ranged from 0.7 to 45.0% and the average was 17.5% (Supplemental Fig. 2). About 18% of all SNPs were only detected in one accession (data not shown). A large proportion of the SNPs (33%) belonged to the 0 to 5% allele frequency class (Supplemental Fig. 2).

Information about these SNPs and their neighboring sequences has been compiled as the National Institute of Agrobiological Sciences Oryza SNP Database (http://oryza-
Single-nucleotide polymorphisms and population structures of Asian rice

snp.dna.affrc.go.jp/en/index_en.html). Users can search this database to get information about the SNPs found in this study by inputting either the physical position of the target or choosing from the available cultivars (Nipponbare is the reference cultivar) via the interactive Web interface. Results can be output in table form, and users can use a genome viewer provided by the Web site to access the sequences flanking the SNPs.

Population structure of rice

To optimize the number of populations \(k\), we compared log \(p(x/k)\) values obtained by the structure analysis for values of \(k\) from \(k=2\) to \(k=10\). The result showed that the rice genetic structure had two solutions, at \(k=2\) and \(k=7\) (Supplemental Fig. 3). For \(k=2\), the two populations corresponded to the two major cultivar groups, indica and japonica, identified in previous classification systems. For \(k=7\), we defined seven cultivar groups: Temperate Japonica (TeJ; 29 accessions), Tropical Japonica I (TrJ_I; 19), Tropical Japonica II (TrJ_II; 3), Tropical Japonica III (TrJ_III; 9), Indica I (Ind_I; 23), Indica II (Ind_II; 16), and Indica III (Ind_III; 24), and some admixture accessions (Fig. 2 and Fig. 3). Analysis of pairwise \(Fst\) values showed that this grouping was statistically significant (Supplemental Table 2). Hereafter, we refer to the Ind_I, Ind_II, and Ind_III cultivar groups as the indica group and to the TeJ, TrJ_I, TrJ_II, and TrJ_III cultivar groups as the japonica group.

Among the seven cultivar groups, some had specific geographical distributions. TeJ accessions mainly originated from Japan and China, whereas TrJ_II accessions were from mountainous areas of Laos, Vietnam, and India. TrJ_III accessions were from the geographic center of rice diversity, namely from northern Laos and Vietnam to southwestern or southern China (Chang 1976). Ind_I accessions were from India, Nepal, and Bangladesh, and Ind_II accessions were from China. Accessions in cultivar groups TrJ_I and Ind_III were from various regions of the world.

Allele frequency in the cultivar groups

The allele frequency distribution can reveal intragroup variation in diversity in the defined cultivar groups. We calculated the allele frequencies, using an allele frequency-class size of 5%, for all SNPs in six cultivar groups, excluding the TrJ_II group because of the small sample number (\(N=3\)) (Fig. 4). In the six groups, the frequency of each allele frequency class decreased from the 0 to 5% allele-frequency class to the 45 to 50% class. However, the pattern of decrease was dissimilar between the three indica groups (Ind_I, Ind_II, and Ind_III) and the three japonica groups (TeJ, TrJ_I, and TrJ_III). A monomorphic SNP frequency (0%) was detected more frequently in japonica than in indica groups. Moreover, in the three indica groups, the

<table>
<thead>
<tr>
<th>Chr</th>
<th>Number of primers</th>
<th>Number of amplicons</th>
<th>Sequenced sites</th>
<th>Number of SNPs</th>
<th>Number of Indels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>318</td>
<td>225</td>
<td>202</td>
<td>601 (5.25)</td>
<td>64 (0.56)</td>
</tr>
<tr>
<td>2</td>
<td>221</td>
<td>180</td>
<td>171</td>
<td>453 (4.67)</td>
<td>35 (0.36)</td>
</tr>
<tr>
<td>3</td>
<td>296</td>
<td>231</td>
<td>173</td>
<td>369 (3.76)</td>
<td>64 (0.65)</td>
</tr>
<tr>
<td>4</td>
<td>319</td>
<td>184</td>
<td>149</td>
<td>301 (3.56)</td>
<td>55 (0.65)</td>
</tr>
<tr>
<td>5</td>
<td>189</td>
<td>117</td>
<td>97</td>
<td>197 (3.58)</td>
<td>21 (0.38)</td>
</tr>
<tr>
<td>6</td>
<td>233</td>
<td>171</td>
<td>121</td>
<td>786 (11.46)</td>
<td>85 (1.24)</td>
</tr>
<tr>
<td>7</td>
<td>271</td>
<td>202</td>
<td>156</td>
<td>400 (4.52)</td>
<td>37 (0.42)</td>
</tr>
<tr>
<td>8</td>
<td>294</td>
<td>144</td>
<td>118</td>
<td>281 (4.20)</td>
<td>22 (0.33)</td>
</tr>
<tr>
<td>9</td>
<td>253</td>
<td>123</td>
<td>89</td>
<td>244 (4.84)</td>
<td>28 (0.55)</td>
</tr>
<tr>
<td>10</td>
<td>199</td>
<td>119</td>
<td>87</td>
<td>243 (4.93)</td>
<td>43 (0.87)</td>
</tr>
<tr>
<td>11</td>
<td>269</td>
<td>154</td>
<td>121</td>
<td>217 (3.16)</td>
<td>33 (0.55)</td>
</tr>
<tr>
<td>12</td>
<td>241</td>
<td>149</td>
<td>94</td>
<td>265 (4.97)</td>
<td>58 (1.09)</td>
</tr>
<tr>
<td>Total</td>
<td>3103</td>
<td>1999</td>
<td>1578</td>
<td>4357 (4.87)</td>
<td>545 (0.61)</td>
</tr>
</tbody>
</table>

a A single PCR product, verified by agarose electrophoresis, was counted as one amplicon.

b The number in parenthesis indicate the number of SNPs per 1 kb.

c The number in parenthesis indicate the number of Indels per 1 kb.

Fig. 2. Population structure of diverse rice accessions. Seven cultivar groups (TeJ, TrJ_I, TrJ_II, TrJ_III, Ind_I, Ind_II, Ind_III) were detected using the STRUCTURE program. The x-axis shows the accessions in each of the seven cultivar groups.
frequency decrease was more gradual than in the three japonica groups. Among the indica groups, Ind_III showed more variation in the frequency of allele frequency class than Ind_I or Ind_II.

**Discussion**

We sequenced the amplicons of 140 diverse accessions to detect genome-wide SNPs. We detected a total of 4357 SNPs in the 140 accessions, for a SNP detection rate of 4.87 SNPs per 1 kb. In general, SNP frequency varied depending on the genomic region sequenced and the cultivar. Generally, intergenic regions had relatively many SNPs; for example, the intergenic regions of the indica cultivars GLA4 and Kasalath had 6.30 and 7.33 SNPs per 1 kb, respectively, compared with the reference Nipponbare japonica cultivar (Monna et al. 2006). In contrast, when both intergenic regions and genes were included, the detection rate was relatively low; for example, indica cultivar 93-11 had 1.7 SNPs per 1 kb compared with Nipponbare (Feltus et al. 2004). McNally et al. (2009) reported detecting 1.7 SNPs per 1 kb in 20 rice cultivars by means of genome-wide SNP

**Linkage disequilibrium**

Genome-wide LD based on SNP data was calculated for all accessions. No clear LD decay was found when all accessions were analyzed (Supplemental Fig. 4A). To detect the level of LD decay in the studied populations more clearly, the number of SNP combinations showing absolute LD ($\Delta^2 = 1$) was investigated with reference to the physical distance between two SNPs in five of the cultivar groups (excluding TrJ_II and TrJ_III because of their small sample size, $N < 10$). The number of SNP combinations with absolute LD peaked at a separation distance of approximately 250 kb and gradually decreased with larger separations in four cultivar groups (Ind_II, Ind_III, TeJ, and TrJ_I), whereas the number peaked at a separation distance of approximately 300 kb in Ind_I. This result indicates that in Ind_I, LD was larger than in the other groups (Supplemental Fig. 4B).
discovery. Although SNP frequency is expected to be lower in genic regions than in intergenic regions, we observed higher SNP frequencies than in either of these studies. The large number of collections that we used in this study may explain the higher SNP frequency in genic regions.

Recently, as a result of advances in sequencing technology, a large number of SNPs have been discovered in rice (McNally et al. 2009, Yamamoto et al. 2010). In those studies, the allele frequency of SNPs detected among diverse accessions needs to be analyzed to let researchers use this data effectively in diversity analysis of rice panels. In this study, we determined the allele frequency of each SNP among 140 diverse accessions (Supplemental Table 1 and Fig. 4). Our data suggest that 67% of total SNPs can be considered informative (excluding rare SNPs with an allele-frequency class <5%) for diversity analysis and genetic mapping.

Many studies have classified rice cultivars on the basis of various criteria, such as morphological characteristics, hybrid sterility and cross-compatibility, and molecular markers, including isozymes, restriction-fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs), and SNPs (Fig. 5). All studies differentiated two main cultivar groups, namely indica and japonica (Garris et al. 2005, Glaszmann 1987, Kato et al. 1928, Matsuo 1952, Morinaga 1954, 1968, Nakagahra 1978, Oka 1958). Before the introduction of molecular markers for genetic classification, rice was generally classified into either two or three groups (Matsuo 1952, Morinaga 1968, Oka 1958), although Morinaga (1968) classified rice into six groups (aus, aman, boro, tjereh, bulu, and japonica), on the basis of hybrid sterility. Glaszmann (1987) used isozyme patterns to define six cultivar groups: the two major groups, indica and japonica, and small groups with different geographic origins. More recently, Garris et al. (2005) used SSR markers to classify rice cultivars into five groups: temperate japonica, tropical japonica, aromatic, aus, and indica. In particular, molecular markers such as SSR markers revealed higher diversity in indica cultivars, and indica cultivars have been classified into several cultivar groups (Garris et al. 2005, Glaszmann 1987, Nakagahra 1978). In our previous study based on RFLPs (Kojima et al. 2005), we identified three cultivar groups (A, B, and C) corresponding to japonica, aus, and indica cultivars, respectively. Among the seven cultivar groups defined in this study, TeJ, TrJ_I, TrJ_II, and TrJ_III belong to group A, corresponding to japonica; Ind_I corresponds to group B and is composed mainly of aus accessions; and Ind_II and Ind_III correspond to group C, which corresponds to indica accessions (Supplemental Table 1). Three cultivar groups (TrJ_I, TrJ_II, and TrJ_III) may correspond to the tropical japonica group of Garris et al. (2005). Ind_I, Ind_II, and Ind_III in this study also correspond to previously defined cultivar groups. For example, Ind_I clearly corresponds to the aus group, and Ind_II and Ind_III to the indica group, of Garris et al. (2005). These results strongly indicate that with a large number of SNPs, a fine-resolution structural analysis of rice populations is possible. Pairwise Fst values (Supplemental Table 2) were high between TeJ and Ind_I, Ind_II, and Ind_III, thus clearly differentiating between the two large cultivar groups (indica and japonica). However, the Fst values were smaller in comparisons among the tropical japonica groups (TrJ_I, TrJ_II, TrJ_III) and between groups Ind_II and Ind_III. The Ind_III

![Fig. 5. Relationships between the populations identified in the present study (solid rectangles) and previously proposed rice cultivar groups (rectangles with dashed lines).](image-url)
and TrJ_I accessions originated from various countries around the world, whereas the TrJ_II accessions were from Vietnam, Laos and India only. These results suggest a possible bottleneck effect that produced TrJ_II from the more diverse group, TrJ_III.

Garris et al. (2005) classified aromatic rice cultivars into the independent cluster aroma. Although we did not characterize the accessions used in our study based on their fragrance, only two accessions (Surjumkhi and Kabashiko) would be categorized as aromatic rice based on data from the literature and based on the NIAS Genebank database (http://www.gene.afrc.go.jp/databases_en.php). Therefore, it was difficult to determine the relationship between the previous classification of aromatic rice and our seven cultivar groups. A gluttonous phenotype is widely observed in the Asian indica and japonica cultivars, and gluttonous rice has been selected as a preferred phenotype during rice domestication in Asia (Konishi et al. 2008, Olsen and Purugganan 2002). Olsen and Purugganan (2002) suggested a single evolutionary origin of the waxy gene associated with the gluttonous phenotype based on the observation of a common splice donor site mutation in all waxy alleles. In the accessions used in this study, the gluttonous phenotype was only observed in the four japonica groups. However, we found no difference in the frequency of gluttonous cultivars among the four japonica cultivar groups.

Genome-wide LD is one important feature of any genome-wide association study (Yu and Buckler 2006). In this study, we estimated LD decay in each cultivar group. Because of the small number of SNPs used, LD decay was not clear when all accessions were considered together (Supplemental Fig. 4A). Therefore, we compared the number of SNP combinations with absolute LD ($D^{2} = 1$) among Ind_I, Ind_II, Ind_III, and TrJ_1. The results suggested that the LD level observed in Ind_I was higher than those in the other three groups (Supplemental Fig. 4B). The dendrogram of the seven cultivar groups showed that Ind_I had a narrower range of variation than the other indica groups (Fig. 3).

The SNPs detected in this study can be a powerful tool for genetic analysis, diversity analysis, genome-wide association studies, and breeding studies. In fact, Yamamoto et al. (2010) have successfully used SNPs to define pedigree haplotypes among Japanese accessions and their ancestral lines. It would be possible to associate this haplotype diversity with phenotypic effects. The same kind of study can be performed on the diverse cultivar groups revealed in this study. For this purpose, it will be necessary to increase SNP sources to provide a more comprehensive SNP panel.

Recent progress in sequencing technology has already provided a powerful means of increasing SNP sources. Recently developed array technology has made it possible to determine the allele of several thousand SNPs simultaneously (Appleby et al. 2009). It is usually difficult to cover the entire diversity of a population because a limited number of accessions can be analyzed simultaneously. In particular, it is difficult to use DNA markers such as SSR to reveal the genetic diversity in a large number of accessions. However, by using genotyping array technology, it will be possible to simultaneously analyze genome-wide SNPs in several thousand accessions. Once such genotype information is available, we will be able to focus on genetic diversity in particular chromosomal regions as well as in the entire genome. To enhance this potential, we have recently embarked on a genotyping study of a large number of accessions (more than 6000) in the NIAS Gene Bank based on the SNP information obtained in this study.

Acknowledgments

We thank Ms. N. Fujitsuka in the Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, and Mrs. N. Aikawa, Mrs. E. Amano, Ms. C. Kanehara, and Mrs. S. Shimizu, in the NIAS, for technical assistance. We are also grateful to the Field Management Division of NIAS for rice field management. This work was supported by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan [Integrated Research Project for Plant, Insect and Animal Using Genome Technology QT1001 (2005–2007) and Genomics for Agricultural Innovation, NVR -0002 NVR0002 (2008–2010)].

Literature Cited


