Single Nucleotide Polymorphisms in Randomly Selected Genes among *japonica* Rice (*Oryza sativa* L.) Varieties Identified by PCR-RF-SSCP

Kenta Shirasawa, Lisa Monna, Sachie Kishitani, and Takeshi Nishio,

Graduate School of Agricultural Science, Tohoku University, Sendai 981-8555, Japan and Plant Genome Center, Tsukuba, Ibaraki 305-0856, Japan

(Received 26 April 2004; revised 13 July 2004)

Abstract

DNA polymorphism of randomly selected genes in rice cultivars was analyzed by the polymerase chain reaction-restriction fragment-single strand conformation polymorphism (PCR-RF-SSCP) technique. Single DNA fragments were amplified from genomic DNA of the Nipponbare cultivar by 671 primer pairs among the 1000 primer pairs tested. PCR-RF-SSCP analysis using the 671 primer pairs detected polymorphism in 108 DNA fragments between 17 *japonica* paddy-rice cultivars. An average of 36.9 DNA fragments showed polymorphism between any pair of *japonica* paddy-rice cultivars. The nucleotide sequences of the polymorphic DNA fragments were determined for 50 alleles of 45 genes together with Nipponbare alleles. In these genes, 142 SNPs and 32 insertions/deletions were identified. Among these 174 sequence variations, 71 were in exons, 78 in introns, and 25 in unassigned regions. There were 28 alleles which had sequence variations in the exons. One allele had a 1-bp deletion in the exon causing a frame-shift mutation, 15 alleles had missense mutations, and the other 12 alleles had synonymous changes and/or sequence variations in 3′ untranslated regions. The number of genes having sequence variations between the rice cultivars and the functional implications of the identified SNPs are herein discussed.

Key words: DNA polymorphism; marker-aided selection; point mutation

1. Introduction

Most of the phenotypic variations in plants are attributable to nucleotide sequence polymorphism of genes. Such polymorphism can be analyzed by various methods for DNA analysis, such as RFLP,\(^1,2\) (restriction fragment length polymorphism), CAPS\(^3\) (cleaved amplified polymorphic sequence), SSR\(^4,5\) (simple sequence repeat), and AFLP\(^6\) (amplified fragment length polymorphism). However, most of the point mutations of genes, e.g., substitutions, insertions, and deletions of single nucleotides, which are considered to be the major source of the phenotypic variation, are not efficiently identified by these methods. Single nucleotide polymorphism (SNP) analysis is therefore required to identify specific point mutations.

For the screening of SNPs in the genome of various organisms, nucleotide sequences of randomly selected chromosomal regions have been determined.\(^7,8\) Such screening is both costly and time consuming. Between closely related crop cultivars, such as the rice cultivars Nipponbare and Koshihikari, the efficiency of SNP detection is extremely low,\(^9\) indicating that detection of one SNP between two such cultivars is much more labor intensive and costly than that between distantly related cultivars (e.g., *japonica* and *indica* in rice). In conventional plant breeding, crossings of closely related cultivars are common. Information on SNPs in the genes between leading cultivars and the lines used frequently as parental lines is much more helpful in practical plant breeding than those between distantly related cultivars.

In rice, the nucleotide sequences of the genome\(^11,12\) and ca. 28,000 full-length cDNA clones\(^13\) have been determined, and the number of genes in the rice genome has been estimated to be 35,000 to 62,500.\(^11,14\) Genes controlling several agronomically important traits have been identified.\(^15,16\) However, most of the genes assigned by the genome projects are of unknown function. When SNPs in all the putative genes in the genome are finally elucidated between the leading cultivars, the genes responsible for agronomically important traits will be easily identified.

Point mutations in the *Wx* gene of rice have been
detected efficiently by the polymerase chain reaction-restriction fragment-single strand conformation polymorphism (PCR-RF-SSCP) technique. Since this method is simple and cost effective, it is expected to be applicable to the screening of SNPs in genes, especially for closely related cultivars. In the present study, we investigated DNA polymorphism in more than 600 genes of unknown function in 21 rice cultivars, including leading *japonica* paddy-rice cultivars, their parental cultivars, and *indica* cultivars using PCR-RF-SSCP (abbreviated to PRS hereafter).

2. Materials and Methods

**Plant materials and DNA isolation:** Twenty-one cultivars of rice (*Oryza sativa* L.) were used as plant materials. Among them, 17 were *japonica* paddy-rice cultivars: Akihikari, Hatsuboshi, Hatushishiki, Hinohikari, Hitomebore, Itadaki, Koganebare, Koshibikari, Kihou, Kirara 397, Nipponbare, Nourin 1, Nourin 6, Nourin 8, Nourin 22, Rikuu 132, and Sasanishiki. One was an upland variety Senshou, and three were *indica* cultivars (Basmati 370, IR36, and Kasalath). Nipponbare, which is the standard *japonica* cultivar for genome study, as well as Koshihikari, Akihikari, Itadaki, Kasalath, and Senshou were provided by Dr. M. Yano of the National Institute of Agrobiological Science (Tsukuba, Japan), and the other cultivars were provided by Mr. K. Nagano of Miyagi Prefecture Furukawa Agricultural Experiment Station (Furukawa, Japan). Genomic DNA was isolated from a leaf by the CTAB (Cetyl trimethyl ammonium bromide) method.

**Primers for amplification of genomic DNA:** Two hundred and forty-five primer pairs, previously developed for STS and CAPS markers and whose sequences are published on line by the Rice Genome Project (http://rgp.dna.affrc.go.jp/publicdata/caps/index.html), were selected for amplification of gene fragments ranging in size from 550 to 4,500 bp. Using the EST sequences published in a database (http://rgp.dna.affrc.go.jp/publicdata/estmap2001/index.html) to amplify about 2-kb genomic DNA, 755 primer pairs were designed from the genomic DNA sequences (http://RiceBLAST.dna.affrc.go.jp/). Primer3 software (http://www.genome.wi.mit.edu/genome-software/other/primer3.html) was used for designing primer sequences. The analyzed genes were arbitrarily selected, so as to evenly distribute the genes on the rice chromosomes.

**PRS analysis:** One microliter (20 ng) of the sample DNA was mixed with 1.3 µl of 10× ExTaq Buffer, 0.8 µl of 2.5 mM of each dNTP, 0.5 µl of 20 µM primers, 0.1 µl of 5 U/µl ExTaq DNA polymerase (Takara, Japan), and water to give a final volume of 13 µl. DNA was amplified under the following thermal cycling conditions: a 1-min denaturation at 94°C; 40 cycles of a 30-sec denaturation at 94°C, a 30-sec annealing at 62°C, and a 90-sec extension at 72°C; and a 3-min final extension at 72°C. After amplification of DNA, the PCR product was divided into two tubes. Half of the PCR product was digested with one restriction enzyme and the other half was digested with another restriction enzyme; these were mixed before electrophoresis. The restriction enzymes were chosen from *Afa I*, *Alu I*, *Mbo I*, *Msp I*, *Hae III* and *Hha I* since digests with these enzymes yield DNA fragments of 100–500 bp for SSCP analysis. The digested sample was mixed with four volumes of denaturing solution (96% [v/v] formamide, 0.1% [w/v] xylene cyanol, 0.1% [w/v] bromophenol blue and 20 mM EDTA). After heating at 95°C for 5 min, the sample was immediately placed on ice. The sample was applied to native 6% polyacrylamide gel (20 cm × 22 cm) containing 5% glycerol and 0.5 × TBE buffer (44.5 mM Tris-borate, 1 mM EDTA) and run under 300 V or 400 V at 10–18°C. After electrophoresis, the gel was stained with GelStar nucleic acid stain (BioWhittaker Molecular Applications), and DNA bands were detected with molecular imager FX (Bio-Rad).

**DNA sequence analysis:** The DNA fragments showing polymorphism by PRS were sequenced according to the method of Nasu et al. (2002). Nucleotide sequences of DNA fragments amplified by PCR were determined by direct sequencing using DYE-namic ET Terminator reagent (Amersham Biosciences) in a MegeBACE 1000 DNA Sequencing System (Amersham Biosciences). Sequencing was carried out on at least two independent PCR amplifications for each allele to distinguish amplification error artifacts from real sequence polymorphisms. Prediction of membrane-spanning domain, cellular localization, and motifs was carried out using Memsat (http://www.cs.ucl.ac.uk/staff/d.jones/memsat.html), PSORT (http://psort.nibb.ac.jp/), and InterPro (http://www.ebi.ac.uk/interpro/), respectively.

3. Results and Discussion

3.1. DNA polymorphism of genes in rice cultivars detected by PRS

Single DNA fragments were amplified from the genomic DNA of Nipponbare by 671 primer pairs (67.1%) of the 1,000 primer pairs tested. All of the 17 *japonica* paddy-rice cultivars except Itadaki were positive for the 671 primer pairs. Itadaki DNA was not amplified by 1 of the 671 primer pairs. In Kasalath, Basmati 370, IR36 and Senshou, 27, 21, 30 and 8 of the 671 primer pairs, respectively, did not amplify DNA fragments. A total of 622 primer pairs amplified single DNA fragments from all 21 cultivars.

PRS analysis using the 671 primer pairs detected DNA polymorphism of 516 DNA fragments (76.9%) between Nipponbare and Kasalath, but detected DNA polymorphism of only 34 DNA fragments (5.1%) between Nipponbare and a leading Japanese cultivar Koshihikari.
In pairwise comparisons between the 17 *japonica* paddy-rice cultivars, the average number of DNA fragments showing polymorphism was 36.9 (5.5%), ranging from 4 (0.6%) between Hitomebore and Hatabushi to 67 (10.0%) between Kirara 397 and Nourin 6. The average number of DNA fragments showing polymorphism between the three *indica* cultivars was 402.7 (60.0%), ranging from 360 (53.7%) between IR36 and Kasalath to 427 (63.6%) between Basmati 370 and IR36.

The average number of polymorphic DNA fragments between the *japonica* paddy-rice cultivars and the *indica* cultivars were 453.9 (67.6%). Basmati 370 showed remarkably fewer differences from the *japonica* paddy-rice cultivars (355.4 polymorphic DNA fragments) than the other *indica* cultivars (503.2 fragments). The upland cultivar Senshou was not similar to the *japonica* paddy-rice cultivars, showing DNA polymorphism in 181.1 fragments. However, neither was Senshou similar to the *indica* cultivars, with polymorphism being detected in 434.7 DNA fragments.

The primer sequences for the 671 genes and band patterns of all the 21 cultivars are shown in Supplemental Table 1, http://www.dna-res.kazusa.or.jp/11/4/05/supplement/table1.html, where the “A” allele is assigned to the band pattern of Nipponbare. There were 108 genes showing DNA fragment polymorphism in the *japonica* paddy-rice cultivars, and these genes are distributed evenly on the 12 chromosomes. Among the *japonica* paddy-rice cultivars, two alleles were observed in 101 genes. There were three alleles in 5 genes. S10844 and R1906 had four and five alleles, respectively. Four out of these 7 genes having more than 3 alleles were close to each other on chromosome 2. It can be inferred that a chromosomal region harboring these 4 genes is unusually polymorphic in the *japonica* cultivars.

The efficiency of the detection of SNPs by this method has been estimated to be about 70% when one restriction enzyme is used. In the present study, the PCR products were treated separately by two restriction enzymes and mixed just before electrophoresis. Therefore, the efficiency of point mutation detection in the present analysis is expected to be about 90%, 1-(1-0.7)^2. Of the DNA fragments whose sequence variations were detected by the present PRS analysis, the average variation was 5.5%, 60.0%, and 67.6% between the *japonica* paddy-rice cultivars, between the *indica* cultivars, and between the *japonica* paddy-rice cultivars and the *indica* cultivars, respectively. Therefore, the DNA fragments having sequence variation can be estimated to be 6.1%, 66.7%, and 75.1% between the *japonica* paddy-rice cultivars, between the *indica* cultivars, and between the *japonica* paddy-rice cultivars and the *indica* cultivars, respectively.

The average length of rice genes can be estimated to be about 4 kb (ca. 3 kb of the transcribed region plus promoter and terminator). In the present study, DNA fragments of ca. 2 kb in the downstream regions, including a 3′-untranslated region of the genes, were analyzed by PRS. The downstream regions of the genes were chosen because the efficiency of successful PCR amplification using the designed primers was higher in the downstream regions than in the upstream regions of the genes in our preliminary investigation. Assuming that the frequency of the sequence variation in the downstream regions of the genes is not different from that in the upstream re-
278 SNPs in Genes among *japonica* Paddy-Rice Cultivars

---

Figure 1. PCR-RF-SSCP analysis of the 21 rice cultivars using the primers of S13574. Allele names are shown under the picture.

...ca. 4-kb fragments of the genes having the sequence variations between the *japonica* paddy-rice cultivars, between the *indica* cultivars, and between the *japonica* paddy-rice cultivars and the *indica* cultivars can be estimated to be 11.8% (1-(1-0.061)^2), 88.9%, and 93.8%. For the closest cultivar combinations, Hatsuboshi and Hitomebore, 1.3% of genes are considered to have sequence variation. If the number of rice genes is 60,000, sequence variation should be found in 780 genes between these two cultivars.

Both parents of Nourin 22, Hatsunishiki, Koshihikari, Hatsuboshi, Koganebare, Hitomebore, and Hinohikari were included in the 17 *japonica* cultivars analyzed (Fig. 2). In four combinations of both parents and their progeny, the number of different alleles between the parents corresponded to the sum of the number of different alleles between one parent and the progeny and that between the other parent and the progeny. These results indicate that the alleles in the parents were exactly inherited by the progeny. Three combinations of the parents and the progeny did not follow this rule. Although the allele of S0703 on chromosome 5 was the same between Koshihikari and Kihou, the S0703 allele in Hatsuboshi—which is the progeny of Koshihikari and Kihou—was different from that in the parents. The S0703 allele in Koganebare and the alleles of E61986 and R2382 on chromosome 8 in Hatsunishiki were also different from those in their parents, Nipponbare–Kihou and Nourin 1–Nourin 22. Different plants of Koganebare had two different alleles, suggesting the incomplete selection of homozygous lines in the breeding of Koganebare or contamination of other lines during seed propagation. Since F10 populations are released generally as new cultivars in rice breeding in Japan, 1/512 of the genes which are different between parents are considered to be heterozygous at the time of completion of cross breeding. These results suggest that the initial crossing in the breeding of these cultivars may have been performed using parental lines in which these genes were heterozygous. In the process of plant breeding, many agronomically important traits of the breeding lines and their parental lines are investigated. Combining the data on important traits of these lines and accumulated information on the genes in the chromosomal segments derived from each parent will enable the screening of the genes controlling some important traits.

3.2. Identification of SNPs in *japonica* paddy-rice cultivars

Complete nucleotide sequences of the amplified DNA fragments were determined in 41 alleles of 36 genes and partial sequences were determined in 9 alleles of 9 other genes (Table 2). These sequences were compared with the allele “A” sequences in Nipponbare determined in this study (Table 2). In the 50 alleles, 142 SNPs and 32 indels were found (Supplemental Table 2, http://www.dnares.kazusa.or.jp/11/4/05/supplement/table2.html). The complete nucleotide sequences of the 41 DNA fragments (average length 1,758 bp) contained 117 SNPs (1 SNP per 616.1 bp) and 29 indels (1 indel per 2,485.4 bp).

Using the published nucleotide sequences of full-length cDNA clones and the rice genome, positions with the sequence variations, e.g., SNPs and indels, in genes were identified. Among the 174 sequence variations, 71 (40.8%) were in exons, 78 (44.8%) in introns, and 25 (14.4%) in genes with no full-length cDNA sequence (Table 2). The average length of the exons and the introns of the sequenced regions of the 46 alleles having published cDNA sequences was 889.5 bp and 840.7 bp, respectively. The frequency of the sequence variations in the exons (0.0017 per base pair) was comparable to that in the introns (0.0020 per base pair). The sequence variations in the exons included 33 nonsynonymous changes (46.5%), 21 synonymous changes (29.6%), and 17 sequence variations in the 3′ untranslated region (23.9%).

There were 28 alleles having sequence variation in exons, and 18 alleles having the sequence variation only in introns (Table 2). Among the 28 alleles with sequence variation in the exons, the B allele of S13157 had a 1-bp deletion in the exon causing a frame-shift mutation. In addition to this 1-bp deletion, this allele had 3 other...
indels from 3 bp to 28 bp and 11 SNPs, which are not shown in Table 2. No nonsense mutation and no splice junction loss were found. Fifteen alleles had missense mutations, 2.0 missense mutations per allele on average. In the other 12 alleles with sequence variation in the exons, 4 had only synonymous changes, 6 had only sequence variation in the 3' untranslated region, and 2 had both.

In *Arabidopsis thaliana*, about 1/20th of EMS-induced point mutations have been reported to be truncation mutations (frame-shift, nonsense mutation, and splice junction loss). The proportion of intron regions to the transcribed regions in rice genes (1.7 kb/3.0 kb) is about two times larger than that in *A. thaliana* genes (0.64 kb/1.99 kb), and therefore the proportion of truncation mutations in the sequence variation is inferred to be smaller than 1/20th in rice. Furthermore, due to natural and artificial selection, the frequency of deleterious mutations in the natural variation among leading cultivars is expected to be lower than that in the EMS-induced mutations. Corresponding to this, among the 123 SNPs and 26 indels within the alleles with full-length cDNA sequence data (Table 2) in the *japonica* paddy-rice cultivars, only one in S13157 was a truncation mutation.

**Table 2.** SNPs and indels in the genes having DNA polymorphism between the *japonica* paddy-rice cultivars.

| marker name | Chr. | cM | allele | analysis∗ | truncation | missense | synonymous | 3' UTR | intron | unassigned region** |
|-------------|------|----|--------|-----------|------------|----------|------------|--------|--------|-------------------|
| S13157      | 1    | 5.6| B       | F/1489    | Del (1)*** | -        | SNP (2)    | SNP (1) | -      | -                 |
| C60656      | 1    | 10.9| B      | F/1765    | -          | SNP (1)  | SNP (1)    | -       | -      | In (1)            |
| S4655       | 1    | 36.9| B       | F/1644    | -          | -        | SNP (1)    | -       | -      | -                 |
| S3813       | 1    | 45.7| B       | F/1750    | In (1), Del (2), SNP (10) | -        | SNP (1)    | -       | -      | -                 |
| S13781      | 1    | 139.9| B    | F/2157    | -          | SNP (1)  | SNP (1)    | -       | -      | -                 |
| S10925      | 1    | 142.9| B    | F/1786    | -          | -        | SNP (1)    | -       | -      | SNP (1)           |
| S20768      | 1    | 95.7| B       | F/1753    | -          | -        | SNP (1)    | -       | -      | In (1), SNP (6)   |
| R1906       | 2    | 96.6| B       | F/1461    | -          | -        | SNP (1)    | -       | -      | -                 |
| C12187      | 2    | 13.5| B       | F/1814    | -          | -        | SNP (1)    | -       | -      | -                 |
| C12409      | 2    | 57.6| B       | F/1461    | SNP (1)    | SNP (1)  | SNP (1)    | In (1), Del (3), SNP (6) | - | - |
| R1843       | 2    | 96.6| B       | F/1461    | -          | -        | SNP (1)    | -       | -      | -                 |
| C913A       | 10   | 17.9| B       | F/1751    | -          | -        | SNP (1)    | -       | -      | -                 |
| E50590      | 10   | 30.2| B       | F/1941    | -          | -        | SNP (1)    | -       | -      | -                 |
| C913A       | 12   | 61.6| B       | F/1751    | -          | -        | SNP (1)    | -       | -      | -                 |
| E60142      | 12   | 100.9| B    | F/1941    | -          | -        | SNP (1)    | -       | -      | -                 |

∗ Sequenced length/Length of Nipponbare fragment, F: Full length was sequenced.

∗∗ The position of SNP or indel was not assigned because of no full-length cDNA sequence data.

∗∗∗ Del: deletion, In: insertion; Number of different nucleotides was shown in parenthesis.

∗∗∗∗ Deletion with nucleotide change from AAAAATTA to TTTTTTGT.
C1272 (= J023022E12) protein is similar to the extracellular domain, such as the C1272 protein, are Cf-2 and Cf-9 proteins having LRRs without the intracellular kinase domain, such as the receptor protein kinase encoded by C1272. Membrane-spanning receptor proteins having LRRs without the intracellular kinase domain, such as the C1272 protein, are Cf-2 and Cf-9 in \textit{Arabidopsis thaliana} \cite{Gns14}

\textbf{Table 3. Differences of the deduced amino-acid sequences of the alleles.}

| Marker name Chr. cM Clone name Accession No. Genes having highest similarity in BlastX* Allele |
|---------------------------------------------------------------|
| S13157 5.6 J013046E24 AK066041 receptor-like protein kinase - rice (T03027) |
| C60956 10.9 001-045-F12 AK062139 probable AAA-type ATPase - \textit{Arabidopsis thaliana} (F84674) |
| S13781 139.9 J013038F20 AK065721 hypothetical protein F24G2.160 - \textit{Arabidopsis thaliana} (T04051) |
| C112 181.8 J033133D12 AK103593 probable hydrolase - \textit{Arabidopsis thaliana} (E84729) |
| R1906 96.6 J023108N22 AK071805 probable chloroplast protein import component - \textit{Arabidopsis thaliana} (B85042) |
| S10844 98.2 J013145E21 AK072178 probable retroelement pol polyprotein - \textit{Arabidopsis thaliana} (F84811) |
| C12187 2.0 J013066F17 AK066538 CER1-like protein - \textit{Arabidopsis thaliana} (T02536) |
| C12409 2.0 J033050K07 AK101576 oligopeptidase A-like protein - \textit{Arabidopsis thaliana} (T49985) |
| C1272 6.3 J023022E12 AK066041 probable receptor protein kinase F14G9.24 - \textit{Arabidopsis thaliana} (G96602) |
| C1467 7.2 001-036-C07 AK105037 GO52 protein - rice (S1633) |
| R80 8.2 J013058G09 AK066237 beta-1,3-glucanase homolog F10M23.170 - \textit{Arabidopsis thaliana} (H84912) |
| C148 10.9 J023031K22 AK066972 protein integral membrane protein F25P22.12 - \textit{Arabidopsis thaliana} (B96764) |

* PIR ACCESSION was in parenthesis.

The deduced amino-acid sequence of R80 (= J013058G09) is similar to \(\beta\)-1,3-glucanase in \textit{A. thaliana} \cite{Gns2-Gns14}, which are expressed in different tissues and by different treatments, have been classified into 4 subfamilies, namely A-D, \cite{Gns2-Gns14} and R80 belongs to subfamily D. The protein encoded by the B allele had an amino acid substitution between LRRs and the transmembrane domain. This amino acid difference may influence the recognition function of the receptor protein encoded by C1272. Membrane-spanning receptor proteins having LRRs without the intracellular kinase domain, such as the C1272 protein, are Cf-2 and Cf-9 in \textit{tomato} \cite{Gns2-Gns14} and CLAVATA2 (CLV2) in \textit{A. thaliana}. \cite{Gns2-Gns14} Cf-2 and Cf-9 are resistance genes for \textit{Cladosporium fla}vum, and CLV2 is a gene controlling cell differentiation in the meristems.

The deduced amino-acid sequence of R80 (= J013058G09) is similar to \(\beta\)-1,3-glucanase in \textit{A. thaliana}, Acc. No. T04806. Genes of \(\beta\)-1,3-glucanase in rice, \textit{Gns2-Gns14}, which are expressed in different tissues and by different treatments, have been classified into 4 subfamilies, namely A-D, \cite{Gns2-Gns14} and R80 belongs to subfamily D. The protein encoded by the B allele had an amino acid change from arginine to glycine at 345 a.a. in the conserved motif of \(\beta\)-1,3-glucanase. Members of subfamily A are considered to participate in disease resistance, and those of subfamilies B, C, and D function in growth regulation. \cite{Gns2-Gns14} However, overexpression of \textit{Gns1} in subfamily B produced disease-resistant rice plants, but did not result in abnormal plant growth, \cite{Gns2-Gns14} indicating that \textit{Gns1} is responsible for disease resistance. Given the previous evidence about the functional significance of this conserved protein domain, the hypothesis can be made that this polymorphism in the B allele of R80 may change...
Figure 2. Genealogy and graphical genotyping of the *japonica* paddy-rice cultivars. The chromosomal regions with the A allele are shown by red boxes, those with the B allele by yellow, those with the C allele by blue, those with the D allele by light blue, and those with the E allele by green. The number in parenthesis between the cultivars indicates the number of different alleles between them. The map of Itadaki, a progeny derived from Koshihikari, Hatsubosi, Sasanishiki, and other cultivars, is shown.
the function of this gene in rice. However, experimental evidence is needed to test this prediction.

Based on the sequence similarity to the reported sequences, C12187 (= J013066F17), C148 (= J023031K22), R10906 (= J023108N22), and C1467 (= 001-036-C07) are predicted to be the genes of CERI-like protein having 4 transmembrane domains and the His-rich motif of sterol desaturase, protein integral membrane protein having 10 putative transmembrane domains and a multi-antimicrobial extrusion motif, a chloroplast protein import component having 5 transmembrane domains, and GOS2 having one transmembrane domain and a motif of translation initiation factor SU1, respectively (Table 3). The amino acid changes in the B to E alleles of these genes were not considered to be of functional significance. The protein encoded by S10844 (= J013145E21) having 4 alleles in the *japonica* paddy-rice cultivars is similar to a retroelement pol polyprotein in *A. thaliana* (Acc. No. F84811). The other genes, C60656 (= 001-0445-F12), C12409 (= J033050K07), and R1862 (= 001-033-H05), having missense mutations are inferred to be nonfunctional in rice, because the full-length cDNA sequences lack either the N-terminal regions or the conserved motifs.

### 3.3. Use of SNPs in plant breeding

Techniques for the identification of cultivars are required in practical plant breeding and for quality control in food markets. Although the cultivars used in this study were closely related to each other, all cultivars were distinguished by the PRS analysis. Analyses of 5 DNA fragments enabled differentiation of the 21 cultivars (Supplemental Table 1, [http://www.dna-res.kazusa.or.jp/11/4/05/supplement/table1.html](http://www.dna-res.kazusa.or.jp/11/4/05/supplement/table1.html)). SSR analysis is widely used for the identification of cultivars in many crop species because of its high polymorphism. However, mutants with point mutations and their original cultivars are not distinguished from each other by SSR analysis. Backcross progeny are not easily distinguished from the recurrent parents by SSR analysis, either. A large number of SNPs are distributed on the genome, and some SNPs in the genes determine important traits of the cultivars. The SNPs in the genes are highly valuable as markers for identification of the cultivars.

The DNA markers linked to the genes of some agronomically important traits are useful in marker-aided selection for practical plant breeding. In rice, DNA markers linked to disease-resistance have been reported. These DNA markers have been successfully used in the breeding of rice isogenic lines for bacterial blight resistance genes and traits related to the quality of rice for cooking and eating. However, for successful breeding, frequent recombinaton to break the linkage between a favorable gene and an unfavorable gene is required. Selection using the linked DNA markers may sometimes result in the loss of the target genes by crossing over between the markers and the target genes. To avoid this problem, it is necessary to use the target genes themselves as the DNA markers. Simple and cost-effective methods for SNP analysis are especially important for the development of a new breeding technology, named “DNA-selection breeding,” where most of the genes for the agronomically important traits are selected by SNP analysis without investigating the plants in the field, only the final test for productivity and adaptability of the lines being performed in the field so as to reduce the labor, cost, and time required for plant breeding.

**Acknowledgements:** We are grateful to Dr. M. Yano of the National Institute of Agrobiological Science, Japan, and Mr. K. Nagano of Miyagi Prefecture Furukawa Agricultural Experiment Station, Japan, for providing plant materials. We also thank Ms. Y. Sugiyama for her technical assistance. This work was supported in part by the project “Development of DNA Marker-aided Selection Technology for Plants and Animals” of the Ministry of Agriculture, Forestry, and Fisheries of Japan.

### References

1. Bernatzky, R. and Tanksley, S. C. 1986, Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics*, **112**, 887–898.
2. Harushima, Y., Yano, M., Shomura, A. et al. 1998, A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics*, **148**, 479–494.
3. Konieczny, A. and Ausubel, F. M. 1993, A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.*, **4**, 403–410.
4. Maroof, M. A. S., Biyashev, R. M., Yang, G. P., Zhang, Q., and Allard, R. W. 1994, Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. *Proc. Natl. Acad. Sci. USA*, **91**, 5466–5470.
5. Panaud, O., Chen, X., and McCouch, S. R. 1996, Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mal. Gen. Genet.*, **252**, 597–607.
6. Vos, P., Hoger, R., Bleeker, M. et al. 1995, AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.*, **23**, 4407–4414.
7. The International SNP Map Working Group 2001, A map of human genome sequence variation containing 1.42 million single nucleotide polymorphism. *Nature*, **409**, 928–933.
8. Wade, C. M., Kulbokas, III E. J., Kirby, A. W. et al. 2002, The mosaic structure of variation in the laboratory mouse genome. *Nature*, **420**, 574–578.
9. Nasu, S., Suzuki, J., Ohta, R. et al. 2002, Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa, Oryza rufipogon*) and establishment
of SNP markers. *DNA Res.*, 9, 163–171.

10. Zhu, Y. L., Song, Q. J., Hyten, D. L. et al. 2002, Single-nucleotide polymorphism in soybean. *Genetics*, 163, 1123–1134.

11. Goff, S. A., Ricke, D., Lan, T.-H. et al. 2002, A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science*, 296, 92–100.

12. Yu, J., Hu, S., Wang, J. et al. 2002, A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, 296, 79–92.

13. The Rice Full-Length cDNA Consortium 2003, Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science*, 301, 376–379.

14. Sasaki, T., Matsumoto, T., Yamamoto, K. et al. 2002, The genome sequence and structure of rice chromosome 1. *Nature*, 420, 312–316.

15. Wang, Z.-Y., Wu, Z.-L., Xing, Y.-Y. et al. 1999, Nucleotide sequence of rice waxy gene. *Nucleic Acids Res.*, 27, 5898.

16. Song, W.-Y., Wang, G.-L., and Chen, L.-L. 1995, A receptor kinase-like protein encoded by the rice disease resistance gene, *xa21*. *Science*, 270, 1804–1806.

17. Wang, Z.-X., Yano, M., Yamanouchi, U. et al. 1999, The *Pib* gene for rice blast resistance belongs to the nucleotide binding and leucine-rich repeat class of plant disease resistance genes. *Plant J.*, 19, 55–64.

18. Yano, M., Katayose, Y., Ashikari, M. et al. 2000, *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *Plant Cell*, 12, 2473–2483.

19. Monna, L., Kitazawa, N., Yoshino, R. et al. 2002, Positional cloning of rice semidwarfing gene, *sd-1*: rice “green revolution gene” encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res.*, 9, 11–17.

20. Sato, Y. and Nishio, T. 2003, Mutation detection in rice waxy mutants by PCR-RF-SSCP. *Theor. Appl. Genet.*, 107, 560–567.

21. Murray, M. G. and Thompson, W. F. 1980, Rapid isolation of high molecular weight plant DNA. *Nucleic Acid Res.*, 8, 4321–4325.

22. Wu, J., Maehara, T., Shimokawa, T. et al. 2002, A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell*, 14, 525–535.

23. Rozen, S. and Skaletsky, H. 2000, Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, 132, 365–386.

24. Jones, D. T., Taylor, W. R., and Thornton, J. M. 1994, A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry*, 33, 3038–3049.

25. Nakai, K. and Horton, P. 1999, PSORT: a program for subcellular localization. *Trends Biochem. Sci.*, 24, 34–35.

26. Mulder, N. J., Apweiler, R., Attwood, T. K. et al. 2003, The InterPro database, 2003 brings increased coverage and new features. *Nucleic Acids Res.*, 31, 315–318.

27. Greene, E. A., Codomo, C. A., Taylor, N. E. et al. 2003, Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics*, 164, 731–740.

28. Hanks, S. K., Quinn, A. M., and Hunter, T. 1988, The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*, 241, 42–52.

29. Tamura, T., Hara, K., Yamaguchi, Y., Koizumi, N., and Sano, H. 2003, Osmotic stress tolerance of transgenic tobacco expressing a gene encoding a membrane-located receptor-like protein from tobacco plants. *Plant Physiol.*, 131, 454–462.

30. Wang, X., Zafian, P., Choudhary, M., and Lawton, M. 1996, The PR5K receptor kinase from Arabidopsis thaliana is structurally related to a family of plant defense proteins. *Proc. Natl. Acad. Sci. USA*, 93, 2598–2602.

31. Dixon, M. S., Jones, D. A., Keddie, J. S., Thomas, C. M., Harrison, K., and Jones, J. D. G. 1996, The tomato * Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell*, 84, 451–459.

32. Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J., and Jones, J. D. G. 1994, Isolation of the tomato * Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science*, 266, 789–793.

33. Jeong, S., Trotchaud, A. E., and Clark, S. E. 1999, The Arabidopsis *CLAVATA2* gene encodes a receptor-like protein required for the stability of the CLAVATA1 receptor-like kinase. *Plant Cell*, 11, 1925–1933.

34. Romero, G. O., Simmons, C., Yaneshita, M., Doan, M., Thomas, B. R., and Rodriguez, R. L. 1998, Characterization of rice endo-β-glucanase genes (*Gns2*-Gns4) defines a new subgroup within the gene family. *Gene*, 223, 311–320.

35. Nishizawa, Y., Saruta, M., Nakazono, K. et al. 2003, Characterization of transgenic rice plants over-expressing the stress-inducible β-glucanase gene *Gns1*. *Plant Mol. Biol.*, 51, 143–152.

36. Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., and Balyan, H. S. 2003, Transferable EST-SSR markers for the study of polymorphism and genetic diversity in wheat. *Mol. Gen. Genomics.*, 270, 315–323.

37. Yu, B. S., Xu, W. J., Vijayakumar, C. H. M. et al. 2003, Molecular diversity and multilocus organization of the parental lines used in the international rice molecular breeding program. *Theor. Appl. Genet.*, 108, 131–140.

38. Blair, M. W., Garris, A. J., Iyer, A. S., Chapman, B., Kresovich, S., and McCouch, S. 2003, High resolution genetic mapping and candidate gene identification at the *xa5* locus for bacterial blight resistance in rice (*Oryza sativa* L.). *Theor. Appl. Genet.*, 107, 62–73.

39. Conaway-Bormans, C. A., Marchetti, M. A., Johnson, C. W., McCouch, S. A., and Park, W. D. 2003, Molecular markers linked to the blast resistance gene *Pi-z* in rice for use in marker-assisted selection. *Theor. Appl. Genet.*, 107, 1041–1020.

40. Singh, S., Sidhu, J. S., Huang, N. et al. 2001, Pyramiding three bacterial blight resistance genes (*xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. *Theor. Appl. Genet.*, 102, 1011–1015.

41. Zhou, P. H., Tan, Y. F., He, Y. Q., Xu, C. G., and Zhang, Q. 2003, Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. *Theor. Appl. Genet.*, 106, 326–331.