Search for and Analysis of Single Nucleotide Polymorphisms (SNPs) in Rice (Oryza sativa, Oryza rufipogon) and Establishment of SNP Markers

Shinobu Nasu, Junko Suzuki, Rieko Ohta, Kana Hasegawa, Rika Yui, Noriyuki Kitazawa, Lisa Monna,* and Yuzo Minobe

Plant Genome Center, 1-25-2 Kannondai, Tsukuba, Ibaraki 305-0856, Japan

(Received 20 July 2002; revised 18 September 2002)

Abstract

We searched for SNPs in 417 regions distributed throughout the genome of three Oryza sativa ssp. japonica cultivars, two indica cultivars, and a wild rice (O. rufipogon). We found 2800 SNPs in approximately 250,000 aligned bases for an average of one SNP every 89 bp, or one SNP every 232 bp between two randomly selected strains. Graphic representation of the frequency of SNPs along each chromosome showed uneven distribution of polymorphism-rich and -poor regions, but little obvious association with the centromere or telomere. The 94 SNPs that we found between the closely related cultivars ‘Nipponbare’ and ‘Koshihikari’ can be converted into molecular markers. Our establishment of 213 co-dominant SNP markers distributed throughout the genome illustrates the immense potential of SNPs as molecular markers not only for genome research, but also for molecular breeding of rice.

Key words: rice (Oryza sativa L., Oryza rufipogon); single nucleotide polymorphism (SNP); molecular marker; genome sequence; polymerase chain reaction (PCR)

1. Introduction

Single nucleotide polymorphisms (SNPs) are the most abundant variations in the genome. They can contribute directly to a phenotype or can be associated with a phenotype as a result of linkage disequilibrium. Most conventional trait markers and molecular markers, such as restriction fragment length polymorphism (RFLP) and cleaved amplified polymorphic sequence (CAPS) markers, are based on SNPs, i.e., nucleotide substitutions or insertions/deletions. Because of their abundance and co-dominance, the use of SNPs as a marker system has the potential for providing the highest map resolution. Typing of SNPs has progressed remarkably over the last several years, making genome-wide linkage analysis and molecular breeding rapid and efficient.

In rice, about 3000 RFLP markers and more than 300 PCR-based genetic markers distributed throughout the genome distinguish the japonica cultivar Nipponbare from the indica cultivar Kasalath. The frequencies of polymorphisms between japonica cultivars detected by RFLP have been published by the Rice Genome Research Program (RGP) (http://rgp.dna.affrc.go.jp/). However, the positions of polymorphic sites on genome sequences are unknown. As the entire genome sequence of rice has now been released, the significance of SNPs in various rice cultivars, strains, and mutants is increasing for both genetic research and breeding.

Recently, we successfully used SNPs as markers in positional cloning of the rice semidwarf gene sd-1. We surveyed SNPs in the candidate region between the japonica cultivar Sasanishiki and the indica cultivar Habataki and converted them to CAPS, derived-CAPS (dCAPS), or SNP markers. In Arabidopsis, more than 50,000 SNPs have been found between ecotypes Columbia and Landsberg and are being analyzed (http://godot.ncgr.org/Cereon/). SNP information has been accelerating positional cloning in Arabidopsis.

We surveyed the frequency of SNPs on all 12 chromosomes in three japonica cultivars, two indica cultivars, and a wild rice (O. rufipogon). We counted the frequency of SNPs in each pair to examine the correlation between genetic distance and SNP frequencies. We also graphed the frequency of SNPs along each chromosome, finding uneven distribution of polymorphism-rich and -poor regions. Finally, we produced 213 co-dominant SNP markers between G4 (indica) and W1943 (O. rufipogon) to illustrate the usefulness of SNPs as a marker system.

Communicated by Michio Oishi

* To whom correspondence should be addressed. Tel. +81-298-39-4823, Fax. +81-298-39-4824, E-mail: monna@pgcdna.co.jp
2. Materials and Methods

2.1. Plant materials

We used three *japonica* cultivars (Nipponbare, Koshi-hikari, and Kitaake), two *indica* cultivars (Kasalath and Guang-lu-ai 4 [G4]), and *O. rufipogon* W1943.

2.2. Detection of SNPs

Genomic DNA was extracted by the CTAB method. We designed PCR primers to amplify fragments of 500 to 1000 bp in the putative intergenic regions, using the published genome sequence and the RiceGAAS. RFLP marker probes or expressed sequence tag (EST) sequences were used for regions for which the sequences were not available. Primer3 software was used to design the primers. Primers were produced at intervals of 5 to 10 cM so that they would be uniformly distributed over all chromosomes. PCR was performed with AmpliTaq Gold (Applied Biosystems, Tokyo, Japan). After the success of amplification was checked by agarose gel electrophoresis, amplified products were treated with ExoSAP-IT (Amersham Biosciences, Tokyo, Japan) to remove excess primers and dNTPs. Nucleotide sequences of amplified fragments were determined by using DYE-namic ET Terminator reagent (Amersham Biosciences) in a MegaBACE 1000 DNA Sequencing System (Amersham Biosciences). Sequencing was performed at least twice for each primer/strain combination to distinguish mistakes of amplification or sequencing from actual SNPs. The sequence results of each strain were aligned using the DNASIS Pro software (Hitachi Software Engineering Co., Ltd., Kanagawa, Japan), and SNPs were searched for and certified by eye.

2.3. Producing SNP markers for G4/W1943

We produced SNP markers for the G4/W1943 combination on the basis of SNPs detected in this study. SNP primers whose 3′ ends lay immediately upstream of the polymorphic site were designed and synthesized. Then a single-nucleotide extension reaction was performed in an AcycloPrime FP SNP Detection Kit (PerkinElmer Life Sciences, Tokyo, Japan) (Fig. 1). The genotype of the target DNA molecule could be determined by excitation of the fluorescent dye in the reaction and measurement on a Wallac 1420 ARVO sx instrument (PerkinElmer Life Sciences) to determine whether a change in fluorescence polarization was observed. The result of genotyping by AcycloPrime FP analysis was that expected from the result of sequencing, the SNP primer was established as an SNP marker.

3. Results and Discussion

3.1. Detection of SNPs

From 607 primer sets covering all chromosomes, we successfully sequenced 417 amplification products. More than 80% of the primers designed from the genome sequence gave successful results, whereas nearly half of the primers produced from partial sequences of ESTs etc. failed in amplification or sequencing (Table 1). Probably the reason for this is that the size of the amplification product was not predictable, because of the absence of introns in EST sequences. The amplification products may have been too long to be amplified. Another possible reason is the multiplicity of homologous genes; some primers gave double or multiple bands on gel electrophoresis.

A total of 2800 SNPs (substitutions, deletions, insertions, and single sequence repeats) were detected among the six strains. The details are shown in Table 2. The rate of polymorphisms was calculated on the assumption that the average length of an amplification product was 600 bp. The total number of base pairs sequenced was thus about 250,000 bp (417 amplifications x 600 bp). The data indicate that rice has an average of one SNP every 89 bp. This level of diversity is lower than maize, but higher than those of human and *Drosophila*.

When the frequencies of polymorphisms were compared for every combination of strains, the highest frequency was between Kasalath and W1943 (0.75%), and the second highest was between G4 and W1943 (0.71%). On the other hand, the frequencies of polymorphisms between *japonica* cultivars and W1943 were about 0.33% to 0.34%. On the basis of these data, it seems that W1943 is genetically more closely related to *japonica* than to...
Table 1. Number of primers produced and analyzed in this study.

| chr | genome | EST | total |
|-----|--------|-----|-------|
| ch1 | 45/49  | -   | 45/49 |
| ch2 | 37/42  | -   | 37/42 |
| ch3 | 36/42  | 8/19| 44/61 |
| ch4 | 23/28  | 7/20| 30/48 |
| ch5 | 31/37  | 6/16| 37/53 |
| ch6 | 32/36  | -   | 32/36 |
| ch7 | 41/46  | -   | 41/46 |
| ch8 | 36/44  | -   | 36/44 |
| ch9 | 1/1    | 14/25| 15/26 |
| ch10| 9/9    | 13/17| 22/26 |
| ch11| 9/11   | 31/81| 40/92 |
| ch12| 9/12   | 29/72| 38/84 |
| total| 309/357| 108/250| 417/607 |

Shown as number of primers successful in PCR amplification/total number of primers tested. Genome: primers were designed from intergenic regions predicted by Rice-GAAS gene prediction software from the publicly available Nipponbare genome sequence. EST: primers were designed from publicly available 5′ and 3′ sequences of rice cDNAs. Primers on ESTs were designed only for regions whose genome sequences were not yet available.

Table 2. Number of SNPs detected between 2 strains.

|     | Nipponbare (100 loci) | Koshihikari (100 loci) | Kitaake (100 loci) | W1943 (100 loci) | G4 (50 loci) |
|-----|-----------------------|------------------------|--------------------|------------------|-------------|
| KSHH | 94 (0.24)             | 77 (0.20)              | 77 (0.20)          | 847 (0.34)       | 847 (0.34)  |
| W1943| 822 (0.33)            | 822 (0.33)             | 822 (0.33)         | 822 (0.33)       | 822 (0.33)  |
| G4   | 1988 (0.68)           | 1716 (0.68)            | 1707 (0.68)        | 1707 (0.68)      | 1707 (0.68) |
| Kasen| 1739 (0.70)           | 1757 (0.70)            | 1748 (0.70)        | 1748 (0.70)      | 1748 (0.70) |

The number of SNPs includes substitutions, insertions, deletions, and single sequence repeats between 2 varieties. On average, 10%–20% of SNPs were insertions/deletions. For example, among 94 SNPs of Nipponbare/Koshihikari, 11 are insertion/deletion polymorphisms. Numbers in parentheses show percentage of polymorphisms (100 x number of polymorphisms/sequenced nucleotides.)

*indica.* Passport data and morphological trait data of W1943 are publicly available on the Wild Rice Database (http://www.pgcdna.co.jp/cgi-bin/wrdb/content.cgi).

The frequencies of polymorphisms between *japonica* cultivars were 0.03% to 0.05%, and between *japonica* cultivars, we found 94 SNPs between Nipponbare and Koshihikari, and 135 between Nipponbare and Kitaake. These results show that DNA marker production is possible even between very closely related cultivars, in which it has been difficult to find polymorphisms by conventional methods such as RFLP. Fig. 3 shows the locations of 64 fragments having SNPs among *japonica* cultivars on the framework of the Nipponbare–Kasalath genetic map. Although 43 fragments had only 1 SNP, the other fragments had multiple SNPs (up to 12) (data not shown).

RGP has produced RFLP markers between *japonica* cultivars. The results (http://rgp.dna.affrc.go.jp/) show that 42 RFLP markers were produced between Nipponbare and Koshihikari by using eight restriction enzymes and 400 probes, screened from a total of 2950 probes. In comparison, we found 94 SNPs (in 50 loci) by using 417 primers. The observed difference reflects the higher efficiency of finding SNPs by direct sequencing.

Using data obtained for Nipponbare and Koshihikari, frequencies of SNPs found in putative intergenic regions and in transcribed (EST) regions were compared, but there was no significant difference. Further investigation will be needed to clarify the regional characteristics in polymorphism frequency, focusing on introns, exons, promoters, untranslated regions (UTRs) and intergenic regions.

### 3.3. Producing SNP markers for G4/W1943

SNPs between G4 and W1943 were found in 312 fragments. SNP primers for these fragments were designed for SNP-typing by AcycloPrime FP reaction, which is based on the single-nucleotide extension method. Fig. 4 shows a sample fluorescence polarization result. As templates we used genomic DNA of each strain and a mixture of both strains as artificial heterozygotes. Thus, only when three clusters (homozygotes of each parent and heterozygotes) were finely separated was the SNP primer accepted as an SNP marker. Consequently, 213 SNP primers produced successful results, and they were es-
Figure 2. Distribution of SNPs among six strains of rice along each chromosome. The horizontal scale indicates the genetic distance from the distal end of the short arm (on the RGP Nipponbare–Kasalath genetic map). The vertical scale indicates the number of SNPs counted at each fragment (locus) among the six strains. The pale box in each figure shows the location of the centromere.
Figure 3. Distribution of SNPs detected among japonica rice cultivars Nipponbare, Koshihikari, and Kitaake. S0214 (chr. 12) is not shown, because its position is unknown.

Figure 4. SNP marker (S0050) measured by AcycloPrime FP analysis. Changes in fluorescence polarization for DNA samples genotyped by AcycloPrime FP analysis are plotted in mP units. (mP expresses the degree of polarization.) A change of ≥40 mP for a dye terminator is scored as positive. A/A: samples positive for the A allele but negative for the G allele (homozygous A); G/G: samples positive for the G allele but negative for the A allele (homozygous G); G/A: samples positive for both alleles (heterozygous).

established as SNP markers for G4/W1943 (Table 3). The positions of the SNP markers are shown in Fig. 5 on the framework of the Nipponbare–Kasalath genetic map.²

As shown in Fig. 5, the SNP markers established in this study are distributed over all 12 chromosomes. These markers can be used for linkage analysis to construct genetic maps and graphical genotypes. Some intervals between markers are still too long, however, so additional SNP markers in these intervals may be required for further analyses.

Using SNP markers for linkage analysis has three advantages: 1) analysis can be performed in the early growth stage of plants, requiring only a small quantity of DNA; 2) a large number of samples can be processed systematically in 96-well plates from sowing to SNP-typing; and 3) time and labor can be saved, as no electrophoresis is needed. Also, compared with CAPS or dCAPS, SNP markers set researchers free from restriction-site hunting (or creating, in dCAPS) and from purchasing expensive restriction enzymes. Another advantage of SNPs is that they are convertible to CAPS or dCAPS markers for laboratories not equipped for SNP typing.

It is likely that the SNP loci found in this study exist in other rice cultivars or strains used in research. Overall, SNPs and SNP markers are applicable to many research and commercial uses, such as studies of genetic diversity, marker-assisted breeding, positional cloning, and cultivar discrimination for quality control of rice in the marketplace.

Acknowledgements: The authors thank Dr. Z-X. Wang for critical reading of the manuscript.
| SNP Marker | Gene/Genome | Chromosome | Reference Genomes | Type | Genomic Location |
|------------|-------------|------------|------------------|------|------------------|
| 168 Rice SNPs and SNP Markers | [Vol. 9, 168] | 1 | 1 | 1 |

Table 3. SNP markers established in this study.
| Chrm | bin 1 | bin 2 | GeneName | Start | End | Direction | Type | mod | Value 1 | Value 2 | Value 3 | Value 4 | Value 5 | GeneName | Score |
|------|-------|-------|----------|-------|-----|-----------|------|-----|---------|---------|---------|---------|---------|----------|-------|
| 22   | 17    | 18    | ACHB    | 687000 | 187600 | Forward  | Gene |     |         |         |         |         |         | ACHB     | 6580.51 |
| 22   | 16    | 17    | ACHD    | 695000 | 187600 | Forward  | Gene |     |         |         |         |         |         | ACHD     | 6580.51 |
| 22   | 17    | 18    | ACHE    | 687000 | 187600 | Forward  | Gene |     |         |         |         |         |         | ACHE     | 6580.51 |
| 22   | 16    | 17    | ACHF    | 695000 | 187600 | Forward  | Gene |     |         |         |         |         |         | ACHF     | 6580.51 |
| 22   | 17    | 18    | ACHG    | 687000 | 187600 | Forward  | Gene |     |         |         |         |         |         | ACHG     | 6580.51 |
| 22   | 16    | 17    | ACHI    | 695000 | 187600 | Forward  | Gene |     |         |         |         |         |         | ACHI     | 6580.51 |

**Table 3. Continued.**
Figure 5. Distribution of SNP markers established for G4 and W1943. This figure is based on the RGP's Nipponbare–Kasalath linkage map and 'Current Genome Sequencing Status' as at April 2001. Markers were distributed on all chromosomes, but some regions had very few markers (for example, an interval of about 30 cM between S0330 and S0036 on chr. 3).

References

1. Risch, N. and Merikangas, K. 1996, The future of genetic studies of complex human diseases, Science, 273, 1516–17.

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, 479–94.

3. Monna, L., Kitaizawa, N., Yoshino, R. et al. 2002, Positional cloning of rice semi-dwarfing gene, sd-1: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis, DNA Res., 9, 11–17.

4. Lukowits, W., Gillmor, C. S., and Scheible, W. R. 2000, Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you, Plant Physiol., 123, 795–805.

5. Murray, M. G. and Thompson W. F. 1980, Rapid isolation of high molecular weight plant DNA, Nucleic Acids Res., 8, 4321–25.

6. Sakata, K., Nagamura, Y., Numa, H. et al. 2002, Rice-GAAS: an automated annotation system and database for rice genome sequence, Nucleic Acids Res., 30, 98–102.

7. Wu, J., Maehara, T., Shimokawa, T. et al. 2002, A Comprehensive Rice Transcript Map Containing 6591 Expressed Sequence Tag Sites, Plant Cell, 14, 525–35.

8. Rozen, S. and Skaltsky, H. 2000, Primer3 on the WWW for general users and for biologist programmers, Methods Mol Biol, 132, 365–86.

9. Chen, X., Levine, L., and Kwok, P. Y. 1999, Fluorescence Polarization in Homogeneous Nucleic Acid Analysis, Genome Res., 9, 492–98.

10. Tenaillon, M. I., Sawkins, M. C., Long, A. D., Gaut, R. L., Doebley, J. F., and Gaut, B. S. 2001, Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.), Proc. Natl. Acad. Sci. USA, 98, 9161–66.

11. Kurata, N., Nagamura, Y., Yamamato, K. et al. 1994, A 300 kilobase interval genetic map of rice including 883 expressed sequences, Nature Genetics, 8, 365–372.