Mol Gen Genet (1996) 252:597–607

ORIGINAL PAPER

O. Panaud · X. Chen · S. R. McCouch

Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.)

Received: 19 January 1996 / Accepted: 3 May 1996

Abstract Microsatellite markers containing simple sequence repeats (SSR) are a valuable tool for genetic analysis. Our objective is to augment the existing RFLP map of rice with simple sequence length polymorphisms (SSLP). In this study, we describe 20 new microsatellite markers that have been assigned to positions along the rice chromosomes, characterized for their allelic diversity in cultivated and wild rice, and tested for amplification in distantly related species. Our results indicate that the genomic distribution of microsatellites in rice appears to be random, with no obvious bias for, or clustering in particular regions, that mapping results are identical in intersubspecific and interspecific populations, and that amplification in wild relatives of *Oryza sativa* is reliable in species most closely related to cultivated rice but becomes less successful as the genetic distance increases. Sequence analysis of SSLP alleles in three related *indica* varieties demonstrated the clustering of complex arrays of SSR motifs in a single 300-bp region with independent variation in each. Two microsatellite markers amplified multiple loci that were mapped onto independent rice chromosomes, suggesting the presence of duplicated regions within the rice genome. The availability of increasing numbers of mapped SSLP markers can be expected to increase the power and resolution of genome analysis in rice.

Key words Microsatellite · Simple sequence repeats (SSR) · Molecular map · Rice (*Oryza sativa*) · Allelic diversity

Communicated by H. Saedler

O. Panaud[1] · X. Chen · S. R. McCouch (✉)
Department of Plant Breeding, 252 Emerson Hall, Cornell University, Ithaca, NY 14853-1901, USA

Present Address:
[1] GEPC, Laboratoire d'évolution et systematique végétale, Bat. 362, Université Paris XI, 91405 Orsay Cédex, France

Introduction

Preliminary studies suggest that simple sequence length polymorphisms (SSLP) are abundant and widely distributed in the rice genome (Wu and Tanksley 1993; Wang and Tanksley 1994; Panaud et al. 1995). This class of co-dominant DNA marker detects higher levels of allelic variation than do RFLP or RAPD markers, is easily and economically assayed by the polymerase chain reaction (PCR), and can be efficiently distributed throughout the world by publication of the sequences of the PCR primers used to amplify the markers. Dense SSLP maps already exist for both human and mouse, with 4,000 and 2,000 markers respectively, mapped to these species (Dietrich et al. 1994; Gyapay et al. 1994). In plants, microsatellite markers are being successfully used as tools in fingerprinting and variety identification, studies of population dynamics and gene diagnostics (Rongwen et al. 1995; Devos et al. 1995; Yang et al. 1994; Blair and McCouch 1995).

Two highly saturated rice RFLP maps are available (Causse et al. 1994; Kurata et al. 1994) and provide the foundation for molecular genetic analysis of almost any trait of interest. Several agronomically important traits, including pest and disease resistances, tolerance to abiotic stress, wide compatibility, plant architecture and yield components, have been located via linkage to molecular markers based on interspecific or intersubspecific crosses (Ronald et al. 1992; Mohan et al. 1993; McCouch and Doerge 1995; Redoña and Mackill 1996; Yanagihara et al. 1995; Cho et al. 1994; Xiao et al. 1995). However, there is a need for markers that reveal a higher level of allelic diversity so that closely related germplasms can be distinguished, and crosses involving members of the same subspecies can be analyzed. The availability of PCR-based markers would also make genome analysis considerably more efficient and cost effective. Microsatellite markers are highly informative and they can be rapidly and reliably visualized using silver staining without the use of radioisotopes. Though

they are expensive to develop, once made available for distribution, they are technically easy and inexpensive to use. The usefulness of microsatellite markers in crop improvement can be expected to increase dramatically as saturated genetic SSLP maps become available.

The aim of our project is to develop an SSLP map of rice for use in rice genome analysis. In this report we present and discuss the strategy we are using to develop our markers, present primer sequences and amplified simple sequence repeat (SSR) motifs, map locations, and an estimate of the polymorphism information content of 20 new markers. Characterization of individual loci provides evidence of complex SSR motifs and multilocus primer pairs.

Materials and methods

Library construction

Rice DNA (cv. IR36) fragments of 300–600 bp in length were obtained by mechanical shearing through a hypodermic needle followed by partial digestion with DNase I. The blunt- ended DNA fragments were ligated with *Eco*RI adaptors (5′-AATTCGGCACGAG-3′, 5′-CTCGTGCCG-3′) into a Lambda-Zap II/*Eco*RI vector. The library was amplified in the *E. coli* host strain XL1-Blue MRF′, selected using X-Gal/IPTG, and titered (Stratagene, La Jolla, Calif., USA).

Isolation of clones containing microsatellite repeats

The small-insert library was screened by plaque and colony hybridization. Plaque lifts were made using Hybond N + nylon filters (Amersham, Arlington Heights, Ill.). Filters were hybridized in hybridization buffer [0.5 M sodium phosphate pH 7.2, 7% SDS (sodium dodecyl sulfate), 1% BSA (albumin)], and probed with four different oligonucleotides that contained microsatellite motifs representing a range of melting temperatures (T_m) as described in Panaud et al (1996). Poly$(CGG)_9$ and poly$(GA)_{13}$ (Pharmacia) were labeled with [α-^{32}P] dCTP using the random hexamer priming method (Feinberg and Vogelstein 1983) and poly$(ATC)_9$ and poly$(ATT)_9$ were labeled with [γ-^{32}P] dATP using the end-labelling method (Sambrook et al., 1989). Filters were washed twice in 2 × SSC, 0.5% SDS, for 15 min each at the appropriate melting temperature as described by Panaud et al. (1995). Positive clones were identified by autoradiography. The process was followed by two more rounds of purification, after which all clones showed a positive hybridization signal. pBluescript phagemids were recovered from the ZapII phage clones by in vivo excision using the Exassist/SOLR system (Stratagene), and plasmid DNA was prepared using an alkali lysing procedure, followed by a resin purification step. The resin used was Celite 545 (Fluka, catalog number 22140) prepared in guanidine hydrochloride pH 5.5, according to the Merlin Miniprep protocol (Gopher Molecular Biology News Network).

Prescreening of microsatellite-containing clones

After we had encountered cases of clones that contained a microsatellite repeat too near to one of the cloning sites to permit the design of flanking PCR primers, a preliminary screening step was introduced to save the cost of sequencing clones where this problem would occur. This prescreening was a modification of that reported in Dietrich et al. (1994) and was used in this study only for clones containing a poly$(GA)_n$ motif, allowing us efficiently to target the most abundant class of microsatellite in rice (Panaud et al. 1996). The procedure was based on four primer combinations: (1) M13 Forward (5′-TGTAAAACGACGGCCAGT-3) and "Internal Forward" (5′-CCCGGATCC$(GA)_9$-3′); (2) M13 Forward and "Internal Reverse" (5′-CCCGGATCC$(CT)_9$-3′), (3) M13 Reverse (5′-CAGGAAACAGCTATGACC-3′) and "Internal Forward", and (4) M13 Reverse and "Internal Reverse". PCR was performed in 50 µl reactions containing 0.2 µM of each primer, 200 µM deoxyribonucleotides, 50 mM KCl, 10 mM TRIS-Cl pH 8.3, 1.5 mM $MgCl_2$, 0.01% gelatin, and 1 unit of *Taq* polymerase. The PCR profile was: 94° for 5 min (denaturation), followed by 35 cycles of 94° for 1 min, 55° for 1 min, 72° for 2 min, and finally by 5 min at 72° for final extension. The PCR reaction was performed in PTC100 96U thermocycler (MJ Research, Watertown, Mass). The products of the four prescreening reactions were run on 2% agarose gels in order to eliminate clones in which the microsatellite motif was located less than 100 bp from either cloning site and, of those remaining, to determine the best orientation for sequencing. Clones that generate the shortest amplification products (but > 100 bp) in reaction (1) or (2) were sequenced using the T7 primer (corresponding to M13 Forward), and those for which the shortest amplification product appeared in reactions (3) or (4) were sequenced with the T3 primer (corresponding to M13 Reverse).

Sequencing

Minipreps were sequenced by the Cornell Sequencing Facility using an Applied Biosystem 373A machine. Based on the results of prescreening, a single reaction using either the T3 or T7 primer was initially obtained. Sequencing from both ends was required for approximately 50% of clones, due to the inability of the sequencer reliably to read through the microsatellite motif.

Primer design and evaluation

PCR primer sequences consisting of 18–24 bp on either side of the microsatellite repeats were selected using the PRIMER program (Eric Lander, Whitehead Institute, Cambridge, Mass). Primers were synthesized by Research Genetics (2130 Memoral Parkway SW, Huntsville, AL 35801; e-mail address: krushing@resgen.com) and can be obtained by requesting them directly from Research Genetics. All primer pairs were tested for PCR amplification using *O. sativa*, cv IR36 DNA (100 ng) as template. PCR conditions were identical to those used for the prescreening step.Of the 25 primer pairs used in germplasm analysis in this study, 20 are reported here for the first time, and five primer pairs were selected from among the nine previously published by Wu and Tanksley (1993).

Mapping of SSLP

Primers showing amplification of a single fragment on a 2% agarose gel were evaluated for polymorphism on denaturing polyacrylamide gels using DNA from the parents of two mapping populations. The first population was a backcross population of 113 individuals derived from an interspecific cross between *O. sativa* (cv. BS 125) and *O. longistaminata* (acc. WL02) (hereafter this population will be referred to as the "SL pop"), which has been used for the development of a high density RFLP map (Causse et al. 1994). The parental survey for this population was made with DNA from BS125 and its F1 hybrid (BS125/WL02). The second population was a doubled haploid population of 135 individuals derived from an intersubspecific cross between the rice varieties IR64 (*indica*) and Azucena

(*tropical japonica*) (hereafter referred to as the "DH pop") (Guiderdoni et al. 1992; Huang et al. 1994). If both parental pairs showed polymorphism, markers were preferentially mapped onto the backcross population, for which the map is most highly saturated. All the individuals of the mapping population were genotyped using the conditions described for the parental survey. Up to five polymorphic markers were multiplexed on a single polyacylamide gel for scoring of segregation: each marker was loaded separately, starting with the one with the lowest molecular weight and run 5 min to allow DNA to penetrate into the gel before subsequent loading.

Non-radioactive detection

PCR products were run on 4% polyacrylamide gels containing 7 M urea, using a sequencing gel apparatus (Owl Scientific, Woburn, Mass). Prior to casting the gel, one glass plate was treated with the binding reagent bind silene (Promega, Madison, Wis.) The other plate was treated with a repelling agent, sigmacote (Sigma Chemical, St. Louis, Mo). The gel was prerun for 1 h to warm it up to 40–60°. Before loading the sample, the PCR products were first denatured at 70–90° for 2 min. After electrophoresis, both plates were separated. Bands were revealed using a silver staining procedure (Promega). The plate on which the gel was bound was treated for 20 min with 10% acetic acid with gentle shaking, then washed three times for 2 min each with distilled water. After incubating for 30 min in staining solution (0.2% w/v silver nitrate, 0.05% formaldehyde) the plate was immersed in distilled water for 10 s and transferred to developing solution (6% w/v sodium carbonate, 0.05% formaldehyde, 0.0002% sodium thiosulfate).

Rice germplasm and pedigree analysis

Seed from 20 rice cultivars and four wild rice species (Table 1) was provided by the International Rice Germplasm Collection at the International Rice Research Institute (Los Banos, Philippines, IRRI) and the USDA Small Grains Germplasm Collection at Aberdeen, Idaho. DNA was extracted from young leaves of plants grown in the Guterman Greenhouse at Cornell University, and used as template for amplification of 25 microsatellite markers. The lines selected for this work included landrace material used as sources of genetic diversity in the early variety development programs (1940–50) in Indonesia, the Philippines, Taiwan, and later at the IRRI, as well as improved varieties from both *indica* and *japonica* subgroups. PCR products were run on a sequencing gel using the procedure described above and alleles were scored according to their migration distance on the gel, using the molecular weight markers V and VIII (Boehringer Mannheim, Indianapolis, Ind). In order to compare the allelic diversity of SSLP and RFLP markers, RFLP data consisting of scores at 87 loci on 13 of the germplasm accessions indicated by a asterisks in Table 1 were extracted from a dataset generated at IRRI by G. Second [CENARGEN, Brasilia, Brazil, personal communication; and available over the RiceGenes database (anonymous FTP: probe.nalusda.gov directory/pub, or WWW: HTTP://probe.nalusda.gov:8000)]. For both microsatellite and RFLP markers, the Polymorphism Information Content (PIC) value was computed according to Anderson et al. (1993):

$$PIC_i = 1 - \sum_{j=1}^{n} P_{ij}^2$$

where P_{ij} is the frequency of the jth pattern for marker i and the summation extends over n patterns. This computation takes into account both the number of alleles detected by a marker at a given locus and the respective frequency of each of the alleles in the subset of germplasm tested. It is therefore a measurement of how informative the markers are.

Table 1 Species evaluated for SSLP using 25 microsatellite primer pairs

Species	Variety	Accession
Oryza officinalis	–	IRGC 101150
Oryza rufipogon	–	IRGC W 1185
Oryza nivara	–	IRGC 103838
Oryza glaberrima	–	IRGC 100127
Oryza sativa	Rexoro	PI 389995
Oryza sativa	Sinawpagh[a]	CI 5094
Oryza sativa	Bluerose[a]	IRGC 151
Oryza sativa	Supreme Bluerose	CI 5793
Oryza sativa	Texas Patna[a]	CI 8321
Oryza sativa	Kitchii-samba[a]	IRGC 49796
Oryza sativa	GEB 24[a]	IRGC 4898
Oryza sativa	Latisail[a]	IRGC 8340
Oryza sativa	Cina[a]	PI 220755
Oryza sativa	Slo-17[a]	IRGC 637
Oryza sativa	CP 231	CI 8993
Oryza sativa	Bluebonnet	CI 8322
Oryza sativa	Benong[a]	IRGC 13530
Oryza sativa	DGWG[a]	IRGC 123
Oryza sativa	Peta[a]	PI 233289
Oryza sativa	Sigadis	IRGC 611
Oryza sativa	CP-SLO[a]	CI 9535
Oryza sativa	IR 8[a]	IRGC 10320
Oryza sativa	IR 127	IRGC 11374
Oryza sativa	IR 24	IRGC 19907
Oryza sativa	IR 64	–
Oryza sativa	Azucena	IRGC 328

[a] Varieties for which RFLP data was retrieved from the RiceGenes database and used to estimate PIC values

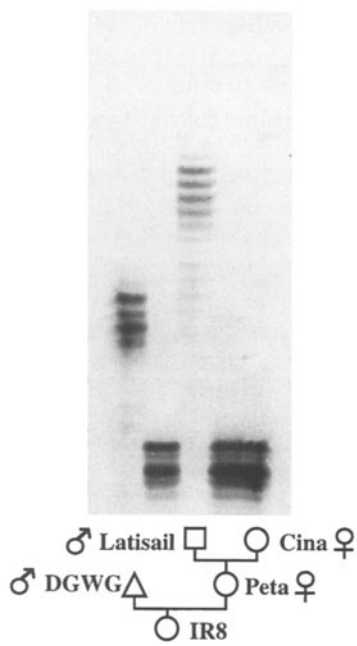

Fig. 1 Allelic diversity at RM21 in the pedigree of rice cultivar IR8

The stability of microsatellite marker alleles was assayed by following their inheritance within the pedigree of rice variety IR8. This pedigree is shown in Fig. 1. The first cross in the pedigree was made in the early 1930s (Partharsarathy 1972) and IR8 was released in

1966 (IRRI 1966).Seed of IR8 has been amplified regularly in the hybridization nursery and as a check in yield trials (IRRI 1994), providing a test of the genetic stability of microsatellite loci that spans a 50–60 yr period.

Sequence homology

To identify sequence homology of the newly identified rice microsatellites with previously reported sequences, a database search was conducted using the Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990). Sixty or more bp of sequence flanking the SSR in the 20 clones identified in this study was used to search the PDB, GenBank and EMBL databases using default settings.

Results

Development of SSLP markers

After three rounds of plaque screening with a labeled poly(GA)$_{15}$ oligonucleotide, 358 positive poly(GA)-containing clones were excised and minipreped. One hundred and forty-four of these clones were eliminated during the prescreening phase and DNA from the other 214 was sequenced. Primers were designed for 144/214 (67%) of these SSR. Forty-five (31%) of these primer pairs gave a PCR amplification product in the expected MW range and 18 poly(GA)-containing microsatellite markers (Table 2) have currently been mapped and characterized for allelic diversity on the 26 *Oryza* accessions listed in Table 1. The remaining primer pairs gave high levels of background, observed as multiple amplification products in the high MW range on polyacrylamide gels.

A smaller number of poly(ATT), poly(ATC), and poly(CGG)-containing clones was screened in this study, and based on that effort, the efficiency of marker development from sequenced clones was between 0–20%. From the 20 sequenced poly(ATT)-containing clones described in Panaud et al. (1995), 11 primer pairs were designed. Seven of these gave high background or no amplification under the PCR conditions tested. Of the remaining four, two were redundant (identical to RM20), and two were monomorphic on all germplasm tested; they could not be mapped and therefore are not reported here. The poly(ATC) marker (RM19) was identified in a minisurvey in which only five poly(ATC)-containing clones were obtained. Primers were designed for three of them, but only RM19 gave reliable amplification. Of the nine poly(CGG)-containing clones identified in Panaud et al. (1995), primers were designed for only two because highly degenerate CG-rich motifs flanking the SSR prohibited reliable primer design. Neither of the two primer pairs bracketing poly(CGG) microsatellites gave reliable amplification and therefore are not reported here.

The number of SSR repeat motifs in the 20 markers characterized in this study varied from 10 to 27, with an average of 17.4. Most of the sequences also revealed degenerate di- or trinucleotide repeats near the major SSR and variation in these adjunct repeats may contribute to the allelic variation detected at these loci.

Allelic diversity

Table 3 summarizes the number of alleles detected in cultivated and wild *Oryza* germplasm for each of the 20 new microsatellite primer pairs and five previously reported by Wu and Tanksley (1993). All 25 markers amplified clearly on 20 varieties of *O. sativa* (Asian cultivated rice) listed in Table 1, on *O. glaberrima* (African cultivated rice), and on the wild species most closely related to *O. sativa* (*O. rufipogon*, and *O. nivara*, both having the AA genome), while all but four of the markers also amplified on the more distantly related *O. officinalis* (a perennial CC genome species). Of the 21 primer pairs that amplified fragments in *O. officinalis*, RM3 and RM22 produced fainter amplification products than the other primers, though both the level of intensity and the molecular weights of those fragments were consistent in multiple experiments. The number of alleles detected in 20 Asian cultivars (*indica* and *japonica* subspecies) varied from two to nine, with an average of five. The three wild species and the African cultivar (*O. glaberrima*) contributed an average of 2.7 new alleles per locus to the diversity represented by Asian rice cultivars, raising the average number of alleles detected from five to eight at these microsatellite loci. When new alleles were detected in species outside of the *indica-japonica* germplasm group, they were detected at 65% of the loci in the AA genome species, *O. rufipogon*, *O. glaberrima*, and *O. nivara*, and at 81% of the loci in *O. officinalis* (Table 3). Therefore, crosses between any of these AA genome species with *indica* or *japonica* varieties appear equally likely to transmit new alleles, and as expected, crosses with more genetically distant material are likely to contribute a greater amount of allelic diversity. No correlation was observed between the number of different alleles detected in this set of germplasm and the number of SSR repeats in the original IR36 clone.

To understand the molecular basis of allelic diversity at a single microsatellite locus, amplification products of RM16 were cloned and sequenced from the three *indica* varieties IR24, DGWG, and IR127. The complete sequence information is presented in Fig. 2. The RM16 clone was originally selected because it contained a poly(GA)n motif, but the sequence information revealed the presence of a poly(TCG)n, a poly(ACG)n, and a poly(CG)n repeat in addition to a poly(GA)n in the region amplified by the RM16 primer pairs for all three genotypes. The sequence information also revealed that allelic variation at the RM16 locus was the

Table 2 Microsatellite information

Locus	Clone	Chromo-some	Size (bp) in IR36	Primer sequence (forward)	Primer sequence (reverse)	Simple sequence repeat motif
RM1	GA12	1	113	GCGAAAACACAATGCAAAAA	GCGTTGGTTGGACCTGAC	ACAA(AG)26CCAC
RM2	GA122	7	150	ACGTGTCACCGCTTCCTC	ATGTCCGGGATCTCATCG	GAGG(GA)2A(GA) 13ATGG
RM3	GA21	6	145	ACACTGTAGCGGCCACTG	CCTCCACTGCTCCACATCTT	TGCC(GA)2GG(GA) 25GGGA
RM4A	GA25	12	159	TTGACGAGGTCAGCACTGAC	AGGGTGTATCCGACTCATCG	TCTC(AG)16TTTG
RM4B	"	11	"	"	"	"
RM5	GA273	1	113	TGCAACTTCTAGCTGCTCGA	GCATCCGATCTTGATGGG	TTAA(GA)14GGCT
RM6	GA285	2	163	GTCCCCTCCACCCAATTC	TCGTCTACTGTTGGCTGCAC	AGGA(AG)16CACA
RM7	GA304	3	180	TTCGCCATGAAGTCTCTCG	CCTCCCATCATTTCGT TGTT	CCTT(GA)19CCGA
RM8	GA317	2	252	CACGTGGCGTAAATACACGT	GGCCAAACCCTAACCCTG	TGAA(GA)14C(GA) 2CGCC
RM9	GA330	1	136	GGTGCCATTGTCG TCCTC	ACGGCCCTCATCACCTTC	ATGG(GA)15GT(GA) 2GGAG
RM10	GA337	7	159	TTGTCAAGAGGAGGCATCG	CAGAATGGGAAATGGGTCC	GAGC(GA)15GGAG
RM11	GA397	7	140	TCTCCTCTTCCCCC GATC	ATAGCGGGCGAGGCTTAG	GACA(GA)17GAAA
RM13	GA478	5	141	TCCAACATGGCAAGAGAGAG	GGTGGCATTCGATTCCAG	ATTA(GA)6TA(GA) TA(GA)16TTGG
RM14	GA506	1	191	CCGAGGAGAGGAGTTCGAC	GTGCCAATTTCCTCGAAAAA	CGAC(GA)18AAGA
RM16	GA53	3	181	CGCTAGGGCAGCATCTAAAA	AACACAGCAGGTACGCGC	AATG(TCG)5(GA) 16ATTA
RM17	GA56	12	184	TGCCCTGTTATTTTC TTCTCTC	GGTGATCCTTTCCCATTTCA	CACA(GA)21AGAA
RM18	GA97	7	157	TTCCCTCTCATGAGCTCCAT	GAGTGCCTGGCGCTGTAC	ACGT(GA)4AA(GA) (AG)16TGAA
RM19	ATC3	12	226	CAAAAACAGAGCAGATGAC	CTCAAGATGGACGCCAAGA	CTGT(ATC)10GTCA
RM20A	ATT59	12	234	ATCTTGTCCCTGCAGGTCAT	GAAACAGAGGCACATTTCATTG	CCTT(ATT)14ATAA
RM20B	"	11	"	"	"	"
RM21	GA275	11	157	ACAGTATTCCGTAGGCACGG	GCTCCATGAGGGTGGTAGAG	GGTA(GA)18GGAG
RM22	GA580	3	194	GGTTTGGGAGCCCATAATCT	CTGGGCTTCTTTCAC TCGTC	GAGC(GA)22GTCC

Table 3 Allelic variation in cultivated and wild relatives of rice

Marker locus	Number of alleles in 20 cultivars[a]	PIC value	Amplification in wild species[b]				Total number of alleles
			O. rufipogon	*O. glaberrima*	*O. nivara*	*O. officinalis*	
RM 1	7	0.75	A*(b)	B*	C*	D*	11
RM 2	4	0.66	A	B*	A*	D*	6
RM 3	5	0.76	A*	B*	C*	D*	9
RM 4A&B	7	0.57	A*	B*	C*	D*	11
RM 5	6	0.77	A	B	C*	D*	8
RM 6	4	0.66	A*	B*	C*	D*	8
RM 7	5	0.77	A	B*	C*	D	7
RM 8	3	0.6	A*	B*	C	D*	6
RM 9	9	0.85	A*	B*	C*	D*	13
RM 10	5	0.75	A*	B	C*	D*	8
RM 11	5	0.75	A*	B*	C*	D*	9
RM 13	6	0.76	A*	B	C*	D*	9
RM 14	5	0.74	A	B*	C	D*	7
RM 16	5	0.66	A*	B	C*	–[c]	7
RM 17	2	0.38	A*	B*	C	D*	5
RM 18	4	0.63	A*	B*	C	D*	7
RM 19	9	0.83	A	B*	C	–[c]	10
RM 20A	7	0.75	A*	B*	C*	D*	11
RM 20B	4	0.62	A*	B*	C*	D*	8
RM 21	8	0.85	A	B*	C	–[c]	9
RM 22	6	0.65	A*	B*	B*	D*	7
RM 122	6	0.65	A	B	A	D*	7
RM 148	3	0.56	A*	A	C*	–[c]	4
RM 164	4	0.73	A*	A	C	D*	6
RM 167	6	0.55	A*	B	C*	D*	9
RM 168	5	0.76	A	B*	C*	D*	8
Average	5.27 ± 0.33	0.692 ± 0.02					7.96 ± 1.95

[a] Cultivars are listed in Table 2
[b] Asterisks indicate that the wild species allele was not detected in cultivars
[c] Dashes indicate no amplification

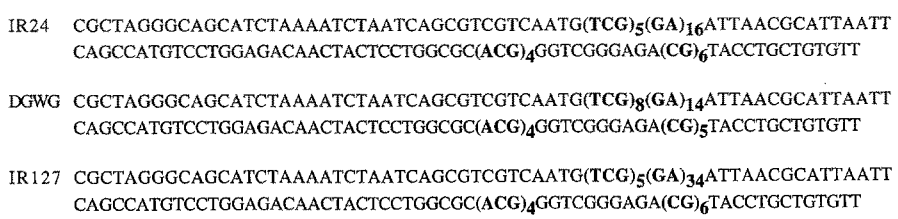

Fig. 2 Sequence of RM16 amplification products obtained from the three rice varieties IR24, DGWG, and IR127

result of independent changes in copy number for three of the four microsatellite motifs (no variation was observed for the poly(ACG)n repeat in these three *indica* varieties). When SSLPs for IR24 and DGWG were compared, the DGWG allele was 3 bp longer than the IR24 allele. From the sequence information, this was determined to be the result of 3 additional poly(TCG) repeats, in combination with 2 fewer poly(GA) and 1 fewer poly(CG) repeat in DGWG compared with IR24. The sequence information also demonstrated that while rice variety IR127 had the same number of poly(TCG) and poly(CG) repeats as IR24, it had more than twice the number of poly(GA) repeats (34 as compared to 16).

A PIC value of 0.692 was calculated for the microsatellite markers based on data from 20 cultivars (Table 3). To compare the information content of microsatellite and RFLP markers, data for 150 RFLP markers was retrieved from the RiceGenes database. RFLP data was available for a subset of 13 of the 20 cultivars evaluated with microsatellites (Table 1), and only probes that were polymorphic for these cultivars were included in this analysis (87/150 polymorphic probes). The data available in the database was based on information derived from a single restriction enzyme per probe as is customary for large germplasm surveys. Using RFLPs, two to four alleles were detected in the 13 varieties, with an average of 2.24 alleles/locus. When PIC values for 25 microsatellite and 87 RFLP markers were compared on the same 13 rice varieties, the value for microsatellites was 0.692 (the same as when 20 cultivars were analyzed), while that computed

for RFLPs was 0.397. This clearly demonstrates the greater power of resolution offered by microsatellite markers and underscores their usefulness for rice genome analysis.

Mapping of SSLP markers

The 20 polymorphic markers identified in this study revealed 22 independent loci. Sixteen loci were mapped using the interspecific backcross population and six were mapped onto the IR64/Azucena doubled-haploid population. To confirm the map positions and to reveal the existence of possible chromosomal rearrangements of these highly variable loci in the interspecific cross, fifteen markers initially mapped on the interspecific backcross population were remapped onto the IR64/Azucena doubled-haploid population. All fifteen markers mapped to identical positions on both populations. The 22 mapped SSLPs are located on eight of the 12 rice chromosomes and show good distribution throughout the genome (Fig. 3).

Five markers (RM3, RM7, RM9, RM20A, and RM21) appeared to be monomorphic when BS125 (the cultivated *indica* parent) was surveyed with its interspecific hybrid. Because we were using a BC population, it was not possible to determine whether this apparent monomorphism was the result of amplification of fragments of identical mw in these *O. sativa* and *O. longistaminata* accessions, or due to a lack of primer amplification in the genetically divergent *O. longistaminata* parent (WL02). To clarify this issue the primers were used to amplify DNA from WL02, which is a highly heterozygous outcrossing species (Causse and Ghesquiere 1991). Amplification was observed in all cases (data not shown) for the five markers appearing to be monomorphic, and a band that was identical in molecular weight to that of BS125 appears to have been transmitted to the progeny. As illustrated in Fig. 4, multiple alleles, suggesting heterozygosity, were often observed in WL02 and not all of these were polymorphic with respect to BS125.

The markers RM4 and RM20, containing a poly $(GA)_{15}$ and a poly$(ATT)_{14}$ motif, respectively, each amplified multiple loci with a single pair of primers. When RM20 primers were used to amplify DNA, four strong bands were visible in inbred rice varieties and two of the bands cosegregated at each locus (Fig. 4). RM20B segregated in both the SL and the DH mapping populations, while RM20A bands were polymorphic only in the DH population. RM20B was mapped to the distal portion of chromosome 11, and RM20A was mapped distally on chromosome 12. A similar situation was observed for RM4, except that both loci could be mapped on the SL as well as the DH population. As shown in Fig. 3, RM4B and RM20B and RM4A and RM20A are closely linked on chromosomes 11 and 12, respectively. Allele frequencies were estimated for each of the loci detected by RM4 and RM20 using a Chi-square test. Only RM4A deviated significantly from the expected 1:1 frequency ($\chi^2 = 8.16$, $P < 0.005$), with an overrepresentation of the *japonica* (Azucena) allele at this locus.

Pedigree analysis

Using the IR8 pedigree (Fig. 1), it was possible to unequivocably trace the ancestral origin of IR8 alleles to a single ancestor at 15 of the 26 microsatellite loci evaluated. As summarized in Table 4, microsatellite alleles appearing in IR8 were arbitrarily assigned an allele designation of "1". Parental lines received a designation of either "1" (if they had the same MW allele as IR8) or "2" (if the MW of their allele differed from IR8). The grandparent generation received designations of either 1, 2, or 3 by the same rule. The immediate parental origin (Peta or DGWG) of the allele carried by IR8 could be determined for 21 of the 26 loci presented in Table 4. For five loci, Peta and DGWG alleles were identical, suggesting that these two *indica* parents may carry alleles that are identical by descent. RM20B was unique among all markers evaluated in this study in that no polymorphism was detected across the entire pedigree. At one locus (RM122), a nonparental allele was detected in Peta, suggesting either that the individual genotype used for DNA extraction from this accession was not identical to that used in the cross from which IR8 was derived, or that a mutation had occurred at this locus in the seed stocks for this variety. Analysis of other individuals from this accession or other accessions with the same name is required to determine the cause of such anomalies.

Homology

The BLAST search detected almost perfect sequence homology (61/62 bp) between the unique sequence flanking the RM11 microsatellite and a rice cDNA clone (DDBJ Accession No. D40450) from 8-day-old etiolated shoot tissue of the *japonica* rice variety Nipponbare, reported by the Japanese Rice Genome Project. A poly(GA) motif consisting of nine repeat units occurred at the 5′ end of the partial cDNA sequence and the location of this microsatellite motif was exactly the same as in the clone obtained from our sheared IR36 library.

Discussion

The 20 SSLP markers developed in this study are technically straightforward and inexpensive to use, and

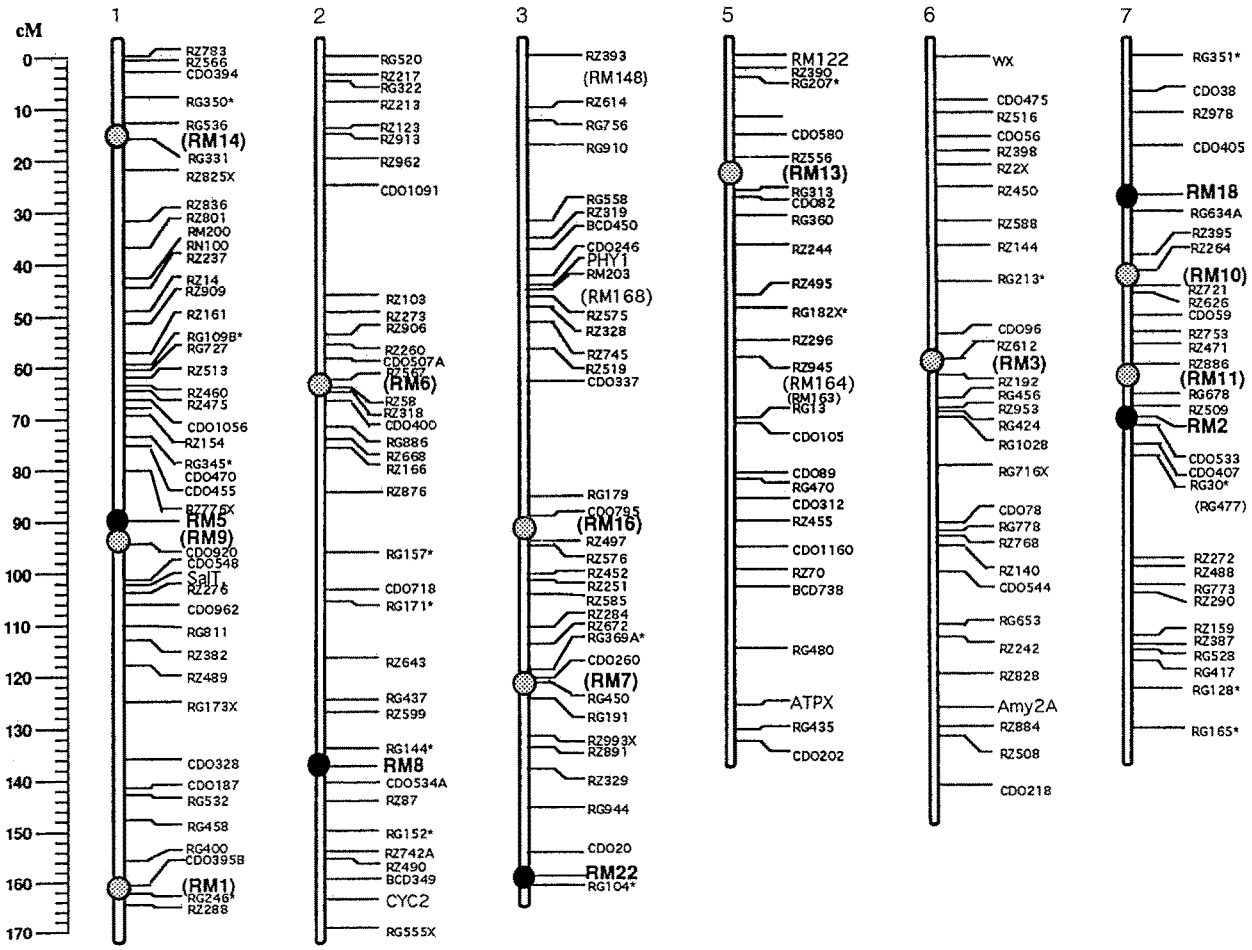

Fig. 3 Framework RFLP map (Causse et al. 1994) of eight rice chromosomes showing the map positions of 20 new microsatellite loci (shown in *bold*). Chromosomal positions designated with *stippled* circles were determined using a doubled-haploid (DH) population (Huang et al. 1994). Chromosomal positions designated with *solid circles* were determined using the interspecific backcross (SL) population (Causse et al. 1994). Markers placed on the SL map based on linkage to markers mapped in the DH population are shown in *parentheses*

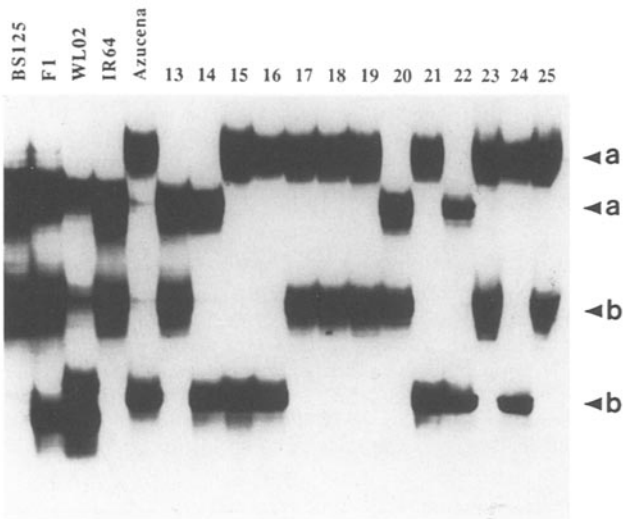

Fig. 4 Multiple loci detected by microsatellite marker RM20, illustrating banding patterns observed in mapping parents from both the SL (BS125, F1, WL02) and DH (IR64, Azucena) populations, and independent segregation of loci (A, B) in a subset of doubled-haploid lines derived from the DH population (Guiderdoni et al. 1992). RM20A maps to chromosome 12, and RM20B maps to chromosome 11

Table 4 Distribution of alleles at 25 microsatellite loci in the pedigree of IR8

Marker[a]	Varieties				
	IR8	Peta (♀)	DGWG (♂)	Cina (♀)	Latisail (♂)
RM 1	1	1	2	1	3
RM 2	1	1	2	1	1
RM 3	1	2	1	2	3
RM 4A&B	1	1	2	1	2
RM 5	1	2	1	3	2
RM 6	1	2	1	1	2
RM 7	1	2	1	3	2
RM 8	1	1	2	1	1
RM 9	1	2	1	3	2
RM 10	1	1	2	2	1
RM 11	1	1	2	3	1
RM 13	1	2	1	2	3
RM 14	1	1	1	2	1
RM 16	1	1	2	1	1
RM 17	1	1	1	1	2
RM 18	1	1	2	3	1
RM 19	1	1	2	1	3
RM 20A	1	1	2	3	1
RM 20B	1	1	1	1	1
RM 21	1	1	2	1	3
RM 22	1	1	1	2	1
RM 122	1	2	1	1	3
RM 148	1	2	1	2	1
RM 164	1	1	2	1	3
RM 167	1	1	1	1	2
RM 168	1	1	2	2	1

[a] Alleles are assigned a designation of "1" if they are identical in MW to alleles in IR8. Designations of "2" and "3" indicate different MW alleles and are assigned sequentially according to generation, with the female parent listed first

are highly informative, as demonstrated by a PIC value almost twice that of RFLP markers. These markers detect discrete loci, are co-dominant and segregate in a Mendelian fashion, making them ideal as genetic markers. The silver staining method for the detection of SSLP markers is a valuable alternative to the radioactive procedure. It is rapid and sensitive and can be used routinely to detect a variety of PCR-based amplification products. In addition, multiplexing of SSLP markers allows us to take advantage of the large size and the high resolution of sequencing gels and increases the efficiency of SSLP analysis when many plant samples are being analyzed with multiple markers.

Despite the value of SSLP markers to users, our results illustrate that microsatellite marker development is a relatively inefficient process. Large numbers of SSR-containing clones must be screened to produce a small number of reliable SSLP markers. Similar observations have been made in wheat (Roeder et al. 1995) and in *Arabidopsis* (Bell and Ecker 1994). Even when the quality of sequence information for microsatellite-containing clones from this sheared library was very high, we experienced a 70% failure rate in obtaining clean PCR-amplified products. Our primer design technique was similar to that being used in other labs (M. Roeder and M. Ganal, Institut fur Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany, personal communication; J. Miller, Whitehead Institute for Biomedical Research, Cambridge, Mass, personal communication), but our success rate was comparatively lower. This may be due to the quality of the library used in this study, and may also be partly due to the use of a single set of PCR conditions for amplification of all microsatellites.

The strategy of using a small-insert library for isolating microsatellite-containing clones made it possible to obtain most of the necessary sequence information in a single sequencing run and avoided the necessity of subcloning. Initially, we chose a sheared library in an effort to avoid any bias in clone availability due to the nonrandom nature of enzyme sites. However, libraries obtained by mechanical shearing or DNase treatment of total genomic DNA may be problematic because blunt-end fragments can readily cross-ligate and generate chimeric clones. Because chimeras cannot be recognized from sequence information, it is virtually impossible to screen against them, but primer pairs designed from chimeric clones will not amplify the corresponding SSR in genomic DNA. It is also possible that clones obtained by shearing may more frequently originate from regions of the genome that do not amplify well by genomic PCR than clones obtained by restriction digestion.

For the next phase of our project, we are developing a size-selected library using enzyme-digested DNA and

sticky-end ligation. We will compare the efficiency of recovery of functional microsatellite markers from the two libraries. Ongoing efforts to map and characterize SSLP markers for rice are aimed at increasing the number of markers available and minimizing the effort required to produce them.

Mapping of SSLP markers indicated that they were well distributed throughout the rice genome. Interestingly, the two primer pairs that each generated multiple bands (RM4 and RM20) show linkage in both locations and map to distal regions of chromosomes 11 and 12. RFLP data (viz. CDO 127 A, B; Causse et al. 1994) previously provided evidence of sequence homology in these two regions of the rice genome, and data supporting this conclusion were also reported by Shimano et al. (1995). The fact that SSLP markers could be mapped to the same locations in both an intra- and an interspecific population, demonstrates that they lie in conserved, evolutionarily stable regions of the genome and may be used with confidence for analysis of germplasm as distantly related as *O. longistaminata.*

The allelic stability of SSLP markers tested in a three-generation rice pedigree suggests that these markers are a powerful tool for tracing the flow of genes in a breeding program. Pedigree breeding has been the method of choice in many national and international rice improvement programs for over 50 years. Complex pedigree relationships are well documented and are available for most modern varieties. In addition, seed stocks of landraces and improved varieties have been preserved and are maintained in several rice germplasm centers around the world. Our data demonstrates that, in comparison to RFLP markers, SSLP markers generate enough allelic diversity to differenciate cultivars within a subspecies, making it possible to analyze the inheritance of specific chromosome segments in germplasm commonly used in rice breeding programs. It should therefore be possible to exploit this information to trace the flow of genes or quantitative trait loci of interest in rice pedigrees and to make predictions about crossing and selection that will increase the efficiency of variety development. In addition, microsatellite marker analysis can be automated (Reed et al. 1994) and this feature is attractive for large-scale, marker-assisted selection programs.

The fact that microsatellite-containing clones have been demonstrated to harbor multiple SSR motifs in a single 300-bp interval in rice contributes to our understanding of allelic variation at these loci. The information suggests that rice varieties may differ in the number of repeats at individual SSR but where multiple SSR motifs occur in a cluster, increases in the number of repeats at one motif may be cancelled out by decreases in other motifs, leading to unexpected similarities in mw of alleles in some cases. These results may help explain the lack of correlation between the number of repeats at a single SSR motif and the number of different alleles detected by these clones in the germplasm evaluated here. The relative frequency of clones containing complex arrays of SSRs may be a function of the species being investigated, the specific SSR being targeted, or the method used to generate the library. Future work will focus on understanding more about the distribution, interspersion patterns, and allelic diversity of microsatellite repeats in the rice genome.

Acknowledgements We would like to thank J. Olufowote, P. Soares and S. Harrington for DNA extractions, N. Huang for mapping data from the IR64/Azucena DH population, and C. Morehouse for formatting the manuscript. We are also grateful for financial support from CNPq in Brasil for a fellowship supporting P.Soares of EPAMIG; the USDA/NRI/NRS Competitive Grants Program on Plant Genome (Grant No. 93-37300-8703), the Rockefeller Foundation, and the International Rice Research Institute.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. J Mol Biol 215:403–410

Anderson JA, Churchill GA, Sutrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis.* Genomics 19:137–144

Blair M, McCouch SR (1995) The development of diagnostic markers for the bacterial blight resistance gene, *xa-5.* Rice Genetics III. International Rice Research Institute (IRRI), Los Banos, Philippines, in press

Causse M, Ghesquiere A (1991) Prospective use of *Oryza longistaminata* for rice breeding. Rice Genetics II (Proceedings of the Second International Rice Genetics Symposium, 14–18 May 1990). International Rice Research Institute, PO Box 933, Manila 1099, Philippines, pp 81–89

Causse M, Fulton TM, Cho YG, Ahn SN, Wu K, Xiao J, Chunwongse J, Yu Z, Ronald PC, Harrington SB, Second GA, McCouch SR, Tanksley SD (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251–1274

Cho YG, Eun MY, McCouch SR, Chae YA (1994) The semidwarf gene, *sd-1,* of rice (*Oryza sativa* L.) II. Molecular mapping and marker-assisted selection. Theor Appl Genet 89:54–59

Devos KM, Bryan GJ, Collins AJ, Stephenson P, Gale MD (1995) Application of two microsatellite sequences in wheat storage proteins as molecular markers. Theor Appl Genet 90:247–252

Dietrich WF, Miller JC, Steen RG, Merchant M, Damron D, Nahf R, Gross A, Joyce DC, Wessel M, Dredge RD, Marquis A, Stein LD, Goodman N, Page DC, Lander ES (1994) A genetic map of mouse with 4,006 simple sequence length polymorphisms. Nature Genet 7:220–246

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Guiderdoni E, Galinato E, Luistro J, Vergara G (1992) Anther culture of *tropical japonica* × *indica* hybrids of rice (*Oryza sativa* L.). Euphytica 62:219–224

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J (1994) The 1993–94 Généthon human genetic linkage map. Nature Genet 7:247–249

Huang N, McCouch SR, Mew MT, Parco A, Guiderdoni E (1994) Development of an RFLP map from a double haploid population in rice. RGN 11:134–137

IRRI (1966) 1966 Annual Report. International Rice Research Institute, PO Box 933, Manila 1099, Philippines

IRRI (1994) 1994 Program Report. International Rice Research Institute, PO Box 933, Manila 1099, Philippines

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang ZX, Momma Y, Umehara Y, Yano M, Sasaki T, Minobe Y (1994) A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genet 8:365–375

McCouch SR, Doerge RW (1995) QTL mapping in rice. Trends Genet 11:482–487

Mohan M, Nair S, Bentur JS, Rao UP, Bennett J (1994) RFLP and RAPD mapping of the rice Gm^2 gene that confers resistance to biotype 1 of gall midge (*Orseolia oryzae*). Theor Appl. Genet. 87:782–788

Panaud O, Chen X, McCouch SR (1995) Frequency of microsatellite sequences in rice (*Oryza sativa* L.). Genome 38:1170–1176

Partharsarathy N (1972) Rice breeding in tropical Asia up to 1960. Rice breeding. International Rice Research Institute, PO Box 933, Manila 1099, Philippines, pp 5–29

Reed PW, Davies JL, Copeman JB, Benett ST, Palmer SM, Pritchard LE, Gough SCL, Kawaguchi Y, Cordell HJ, Balfour KM, Jenkins SC, Powell EE, Vignal A, Todd JA (1994) Chromosome-specific microsatellite sets for fluorescence-based, semi-automated genome mapping. Nature Genet 7:390–395

Roeder M, Plaschke J, Koenig SU, Boerner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327–333

Ronald PC, Albano B, Tabien R, Abenes L, Wu K, McCouch S, Tanksley SD (1992) Genetic and physical analysis of the rice bacterial blight disease resistance locus, *Xa21*. Mol Gen Genet 236:113–120

Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90:43–48

Shimano T, Inoue T, Antonio BA, Kajiya I, Shomura A, Lin SY, Kuboki Y, Nagamura Y, Kurata N, Yano M, Sasaki Y (1995) Extensive conservation in linkage alignment of RFLP markers between rice chromosome 11 and 12. USDA Plant Genome III Meeting, Jan. 15–19, 1995, San Diego, Calif. Poster No. 212

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. a laboratory manual (2nd edn) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Wang Z, Weber JL, Zhong G, and Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1–6.

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225–235

Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745–754

Yanagihara S, McCouch SR, Ishikawa K, Ogi Y, Maruyama K, Ikehashi H (1995) Molecular analysis of the inheritance of the *S-5* locus, conferring wide compatibility in *Indica/Japonica* hybrids of rice (*O. sativa* L.). Theor Appl Genet 90:182–188

Yang GP, Saghai Maroof MA, Xu CG, Zhang Q, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 245:187–194