Theor Appl Genet (2000) 100:697–712

S. Temnykh · W.D. Park · N. Ayres · S. Cartinhour
N. Hauck · L. Lipovich · Y.G. Cho · T. Ishii
S.R. McCouch

Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.)

Received: 5 May 1999 / Accepted: 16 August 1999

Abstract In order to enhance the resolution of an existing genetic map of rice, and to obtain a comprehensive picture of marker utility and genomic distribution of microsatellites in this important grain species, rice DNA sequences containing simple sequence repeats (SSRs) were extracted from several small-insert genomic libraries and from the database. One hundred and eighty eight new microsatellite markers were developed and evaluated for allelic diversity. The new simple sequence length polymorphisms (SSLPs) were incorporated into the existing map previously containing 124 SSR loci. The 312 microsatellite markers reported here provide whole-genome coverage with an average density of one SSLP per 6 cM. In this study, 26 SSLP markers were identified in published sequences of known genes, 65 were developed based on partial cDNA sequences available in GenBank, and 97 were isolated from genomic libraries. Microsatellite markers with different SSR motifs are relatively uniformly distributed along rice chromosomes regardless of whether they were derived from genomic clones or cDNA sequences. However, the distribution of polymorphism detected by these markers varies between different regions of the genome.

Key words Microsatellite markers · Genetic map · Allelic diversity · Genome organization · Rice (*Oryza sativa* L.)

Communicated by F. Salamini

S. Temnykh · N. Hauck · L. Lipovich · Y.G. Cho · T. Ishii
S.R. McCouch (✉)
Department of Plant Breeding, 252 Emerson Hall,
Cornell University, Ithaca, NY 14853-1901, USA
e-mail: stm4@cornell.edu
Tel.: +1 607 255 0420

William D. Park · Nicola Ayres · Sam Cartinhour
Department of Biochemistry and Biophysics,
Texas A&M University, College Station, TX 77845, USA

Present addresses:
N. Hauck, Department of Horticulture/Plant Breeding and Genetics Program, 326 Plant and Soil Science Building, University of Michigan, East Lansing, MI 48825, USA

L. Lipovich, Department of Molecular Biotechnology, University of Washington, 1705 NE Pacific St., Box 357730, Seattle, WA 98195-7730, USA

Y.G. Cho, Department of Agronomy,
Chungbuk National University, Chongju 361–763, Korea;
Laboratory of Plant Breeding

T. Ishii, Faculty of Agriculture,
Kobe University, Nada-ku,Kobe 657–8501, Japan

S. Cartinhour, USDA/ARS & Dept. of Plant Breeding,
252 Emerson Hall, Cornell University, Ithaca, NY 14853–1901, USA

Introduction

Microsatellite markers based on the variation in the number of simple sequence DNA repeats (SSRs) have become the markers of choice for a wide spectrum of genetic, population, and evolutionary studies (Jarne and Lagoda 1996; Powell et al. 1996). Significant progress has been made in the development of second-generation genetic maps based on these abundant and highly polymorphic markers for many different species, including human (Dib et al. 1996), mouse (Dietrich et al. 1996), rat (Serikawa et al. 1992; Jacob et al. 1995), dog (Mellersh et al. 1997) chicken (Groenen et al. 1998), and plants such as wheat (Bryan et al. 1997; Röder et al. 1998), maize (Chin et al. 1996; Taramino and Tingey 1996), potato (Milbourne et al. 1998), and soybean (Akkaya et al. 1995; Cregan et al. 1999).

Studies in various organisms provide evidence that the number of microsatellite sequences in a genome, their length, composition, mutation rate and chromosomal distribution can vary drastically among taxa. This has implications for SSLP marker development. For instance, CA/GT short-sequence repeats, which are the most abundant and variable class of microsatellites in mammalian genomes, are generally less frequent and less variable in plant genomes (Powell et al. 1996). In *Arabidopsis*, CA/GT microsatellites are poorly represented (Depeiges et al. 1995) and show a very low level of variability (Bell and Ecker 1994). In sugar beet, GT-

containing microsatellite sequences are part of a more complex repeating element which is present in multiple copies near centromeres and thus have limited potential for mapping purposes (Schmidt and Heslop-Harrison 1996). However, successful development of many informative SSLP markers based on this GT/CA SSR motif for maize (Chin et al. 1996; Taramino and Tingey 1996), barley (Liu et al. 1996), wheat (Bryan et al. 1997; Röder et al. 1998) and white pine (Echt et al. 1996) suggests that it can be a valuable source of microsatellite markers in some other plant species in addition to the most frequently exploited GA motif.

In rice (*Oryza sativa* L.), early studies (reviewed by McCouch et al. 1997) demonstrated that microsatellite markers are distributed relatively uniformly throughout the genome and detect a high level of allelic diversity in cultivated varieties and distantly related species. A map consisting of 121 microsatellite loci and providing genome-wide coverage in rice has been recently published (Chen et al. 1997). These simple sequence repeats (SSRs) were predominantly poly(GA) motifs isolated from two genomic libraries (Panaud et al. 1996; Chen et al. 1997), with a smaller number of SSLP markers with tri- and other di-nucleotide motifs developed from microsatellite-containing sequences from GenBank (Wu and Tanksley 1993; Akagi et al. 1996). There are an estimated 5700–10 000 microsatellite sequences with different di-, tri-, and tetra-nucleotide repeat units in the rice genome that can be potentially used to construct a genetic map based solely on microsatellite markers (McCouch et al. 1997). The relative frequencies of 13 different di-, tri- and tetra-nucleotide repeats in the rice genome have been estimated in hybridization experiments by Panaud et al. (1995) and several markers containing GT, AT, TCT and ATT repeats have been mapped (Wu and Tanksley 1993; Akagi et al. 1996; Panaud et al. 1996). However, the limited number of loci characterized for each motif restricted the evaluation of these less-abundant classes of SSR sequences for length variation and genome distribution.

In this study we have developed microsatellite markers in rice with different di- and tri-nucleotide repeats based on the screening of a small-insert *Tsp*509-digested genomic library and a search of public databases. In addition, we have compared the efficiency of marker development between different SSR motifs and between SSR-containing sequences obtained from library screening versus those extracted from the GenBank database. A total of 312 microsatellite markers including 124 previously reported and 188 newly developed (97 isolated from genomic libraries and 91 derived from sequences extracted from GenBank) have been mapped to construct a microsatellite map for rice with a density sufficient for both basic genetic studies and breeding applications. In a companion paper by Cho et al. (1999), the markers have been evaluated for genetic variability using a panel of 14 rice varieties representing diverse germplasm. The level of variability at SSLP loci was then mapped onto the rice genome and used to assess the organization of microsatellite sequences in rice.

Materials and methods

Isolation of clones containing microsatellite sequences

A previously constructed *Tsp*509-digested small-insert library in a Zap II/*Eco*RI phage cloning vector (Stratagene, LaJolla, Calif.), as described by Chen et al. (1997) was screened for the presence of microsatellite sequences by plaque hybridization with ^{32}P-labeled synthetic oligonucleotides according to the protocol described by Panaud et al. (1995). For the first round of screening the library was plated using a *Escherichia coli* 'XL1-Blue MRF' strain with a density of about 5000 plaques per 130-mm plate. Duplicate membranes were lifted from each plate. One set was hybridized with a labeled di- or tri-oligo-nucleotide probe and the other set was hybridized with a labeled poly(GA) probe which served as a control for the estimation of the relative frequency of clones with different motifs. Filters were hybridized at 65°C and washed at 65°C for GA, 62°C for GT and 56°C for CTT and CAT motifs, as recommended by Panaud et al. (1995).

Putatively positive clones isolated after the second round of purification and containing inserts of the expected size range (300–1200 bp) based on PCR-based prescreening, as described by Chen et al. (1997), were selected and sequenced by the Cornell sequencing facility using standard dideoxy dye-terminator chemistry on an Applied Biosystems 377 machine.

Six microsatellite markers were isolated from an enriched genomic library as described by Bligh et al. (1999).

Identification of simple sequence repeats in GenBank database

A total of 12 532 rice DNA sequences were obtained from GenBank (Release 96, August 1996) (see Table 1). Perfect tandem di- and tri-nucleotide repeats with more than five repeat units were extracted from sequences in a FASTA format using Perl scripts and regular expression matching.

BLAST search and redundancy search

Unique flanking regions of microsatellite-containing sequences were submitted to BLAST (Altschul et al. 1990) and subjected to a redundancy search against previously isolated SSRs of the same motif in order to eliminate redundant SSRs from the new set of markers.

Primer design and evaluation of polymorphism

PCR primers flanking microsatellite repeat sequences were selected using the Primer 0.5 program (S. Lincoln, M. Daly, and E. Lander, Cambridge, Mass.) and synthesized by Research Genetics (Huntsville, Ala.). Newly synthesized primers were tested for amplification and polymorphism using DNA from the parents of mapping populations. DNA was extracted from fresh leaves by the potassium acetate method (Dellaporta et al. 1983). PCR was performed in a PTC100 96V thermocycler (MJ Research Inc., Watertown, Mass.) as described by Chen et al. (1997) with the exception that 15 µl of reaction mixture was used instead of 25 µl, and 20 ng of DNA, 1 pmol of each primer and 0.5 units of *Taq* were added per reaction. The basic profile was: 5 min at 94°C, 35 cycles of 1 min at 94°C, 1 min at 55°C, 2 min at 72°C, and 5 min at 72°C for final extension. Two different annealing temperatures, 61°C and 67°C, were used to amplify specific microsatellite primer sets (see Table 2). PCR products were separated on 4% polyacrylamide denaturing gels and marker bands were revealed using the silver-staining protocol as described by Panaud et al. (1996).

Mapping of SSLPs

Four mapping populations were used as the basis for placing microsatellite markers onto rice chromosomes: DH1 from the Inter-

Table 1 Size distribution of microsatellite motifs observed in 12 532 rice sequences in the GenBank database

SSR motif	Number of repeat units															
	5	6	7	8	9	10	11	12	13	14	15	16–20	21–25	26–30	>30	Total
AG	173	91	69	40	27	15	13	9	9	3	3	6	1	1	2	463
AC	41	11	8	9	1					1		1				72
CG	51	8	1													60
AT	24	12	2	1			1					3	3	2	1	49
															Total	644
CCG	312	157	72	24	6	3	2	1								577
ACG	198	57	17	10	1											283
AGG	135	84	40	8	2											269
ACC	82	17	10	4		1	1									115
AAG	63	23	13	1	9				1				1			111
ACT	12	4	2				1									19
AAC	9	3	1	1												14
AAT	3					2			1							6
															Total	1394
AAGA	2		1													3
ACCT	2															2
ATAA	8															8
ATAG	1															1
ATCG	2															2
ATCT	0	0	0	1												1
CCTT	2															2
CTCG	2	3														5
GCAG		1														1
TCCA	1															1
TGTA	1															1
TTAA	4	1														5
															Total	32

national Rice Research Institute (Los Baños, Philippines), RIL1 from the Korean Rice Genome Project (National Agricultural Science and Technology Institute, Suweon, Korea), RIL2 from Texas A&M University (College Station, Tex.), and SL from Cornell University (Ithaca, N.Y.) (see Table 1 in a companion paper by Cho et al. (1999). For the DH1 and SL populations, markers were placed using a randomly selected subset of 96 individuals from the original mapping populations and the RFLP data sets described in Huang et al. (1994) and Causse et al. (1994), respectively. For the RIL1 and RIL2 populations, markers were placed using the complete sets of recombinant inbred lines. PCR and microsatellite detection were as described in Chen et al. (1997). Segregation was scored and markers were integrated into the existing RFLP framework maps for each population using MAPMAKER 2.0 (Lander et al. 1987) on a Macintosh computer. The "ripple" test was used to confirm marker order as determined by multipoint analysis. Markers with a ripple of LOD >2.0 were integrated into the framework maps, and those mapping with LOD <2.0 were assigned to the most-likely intervals.

The distribution of polymorphism as a function of map position was observed using the GeneFlow software designed by E. Paul (epaul@idsonline.com, Alexandria, Va.). The program supports several conceptual display frameworks, including a genome diagram that allows users to manipulate and display information about polymorphism. In this study, we evaluated the genotypes of a panel of rice varieties at a set of microsatellite loci distributed throughout the genome and compared the allelic diversity of the loci as a function of their map position.

Results

Microsatellite markers derived from the GenBank database

Approximately half of the microsatellite markers developed in this study were derived from rice sequences extracted from the GenBank database. Screening of 12 532 entries tagged as originating from rice identified 644 sequences with dinucleotide motifs, 1394 sequences with different trinucleotide motifs and 32 sequences with tetranucleotide motifs, with the GA and CCG motifs being the most frequent among di- and tri-nucleotide SSRs, respectively (Table 1). The rice sequences in the database contained mostly unannotated ESTs derived from cv Nipponbare (*O. sativa japonica*) (Sasaki et al. 1994; Yamamoto and Sasaki 1997) as well as a smaller number of ESTs from other sources (http://bioserver.myongji.ac.kr/ricemac.html.), and the complete genomic sequences of several rice genes. Although a large number of SSR-containing sequences were found, only 222 primer pairs were designed for sequences with the longest repeat motifs. Of these, 142 gave amplification products of the expected size. A total of 88 of the 142 primer pairs produced informative polymorphic markers which could be incorporated into the genetic linkage map. In general, microsatellite sequences located in cDNA

clones are short, with less than ten repeat units in a run (Table 1). This is especially true for GC-rich trinucleotide motifs, such as CCG, ACG, AGG and ACC. By contrarst, GA- and AT-polydinucleotides as well as AT-rich polytrinucleotites (AAT and AAG motifs, specifically) have a tendency to contain longer tracts of perfect repeats. The positions of 27 microsatellites within known or putative rice genes are described by Cho et al. (1999).

Isolation of microsatellite markers from genomic libraries

Another set of SSLP markers has been developed based on sequences isolated from genomic libraries. Most were derived from a *Tsp*509-digested small-insert genomic library previously used for the development of a set of GA-containing microsatellite markers reported by Chen et al. (1997). In addition to the poly (GA) SSR motif, poly (GT), (CAT) and (CTT) motifs were targeted. These simple-sequence repeats were less abundant in the rice genome than the GA motif, with 2–6-times fewer positive clones isolated from the same number of plaques during the first round of screening. There were 169 putative positive clones for (GT)*n* versus 385 for (GA)*n*, 70 (CAT)*n* versus 165 (GA)*n* for CAT, and 57 (CTT)*n* versus 356 (GA)*n* for the CTT motif.

After sequencing 197 clones, 170 high-quality sequences were obtained: 59 with the GT motif, 54 with the GA motif, 34 with CAT, and 23 with the CTT motif. All sequences containing the same microsatellite motif were subjected to the redundancy search using local BLAST. Nonredundant sequences were then used for primer design. In this screening of the *Tsp*509-digested library, the proportion of clones isolated more than once was higher among CAT and CTT clones (17% and 12%, respectively) than among GA and GT clones (about 4%). This suggests that only a limited number of trinucleotide microsatellite sequences is present in the *Tsp*509-digested library, perhaps due to the genome-coverage bias inherent in any enzyme-digested genomic library. Another factor influencing the efficiency of marker development was the isolation of a relatively high proportion of sequences (about 30%) with short tracts of non-interrupted repeats (less than five repeat units) for CAT and GT motifs. Many of these sequences contained degenerate microsatellite-like motifs, frequently adjoined by other types of simple repeats (Table 2). GT repeats were found in association with poly(AT)*n* motifs or with AT-rich tri- and tetra-nucleotides in 26% and 22% of the cases, respectively. Most of the sequences with fewer than five perfect repeat units produced monomorphic markers and therefore could not be genetically mapped. Several primer pairs produced complex patterns of segregating bands with dominant or codominant inheritance and they were mapped as multiple loci. The highest efficiency was achieved for the GA and CTT clones, which tended to contain sequences with long tracts of perfect repeats that could be easily converted into highly polymorphic genetic markers.

Eleven more markers were obtained by redesigning primers for SSR-containing sequences previously isolated from genomic libraries that failed to amplify (Panaud et al. 1996; Chen et al. 1997); of these, five contained the GA motif (clones GA264, GA588, CT109, CT210 and CT483) and six had trinucleotide motifs (ATT and TCT clones from the physically sheared library: ATT20, ATT35, TCT114, TCT116, TCT117, TCT121). The strategy was to select new primers closer to the target microsatellite sequence. This procedure minimized the probability that if clones were chimeric, primer sequences would reside in different segments of chimeric inserts. When primers were redesigned to be closer to the SSR motif, it increased the frequency of PCR amplification giving a product of the predicted size with genomic DNA as a template, while previously designed primers amplified only with purified DNA from the corresponding clone.

Marker information and nomenclature

Information related to the 188 microsatellite markers developed in this study, 91 from the sequence databases and 97 from genomic libraries, is summarized in Table 2. It includes locus designation, chromosome location, primer sequence information, description of microsatellite motif, and the size of PCR product amplified in reference lines IR36 or Nipponbare (predicted based on the sequence used for primer design). Accession numbers for the GenBank-derived microsatellites and clone names for markers isolated from genomic libraries are also included. All mapped markers were assigned RM locus names according to the nomenclature guidelines presented in earlier studies (Panaud et al. 1996; Chen et al. 1997): RM1–100 numbers indicate markers from the sheared library; RM101–199 numbers indicate GenBank-derived markers; RM201–345 are markers from the *Tsp*509-digested library and RM345–351 are from other genomic libraries. Markers that mapped to more than one locus were given a suffix (A, B, C) following the RM designation. Markers identified in this study that showed sequence similarity to those reported by Akagi et al. (1996) were included in Table 2 with the primer sequences designed in our labs and RM locus designations along with previously reported OSR names.

Information related to the genetic variability of the microsatellite markers reported in this study was included in Table 2 but is discussed in a companion paper by Cho et al. (1999).

Map construction

One hundred and eighty eight new microsatellite markers were integrated into an existing map consisting of 121 microsatellite markers previously reported by Chen

Table 2 Microsatellite marker information

Locus name	GenBank accession number	Map	Repeat type and length	No. of alleles	PIC	PCR product in refer. line	Size range (bp)	Forward primer	Reverse primer	Annealing temp.
RM101 = OSR2[a]	D17586[b]	12	$(CT)_{37}$	6	0.59	324	258–334	gtgaatggtcaagtgacttaggtggc	acacaacatgttccctcccatgc	55
RM102	D17586[b]	1	$(GGC)_7(CG)_6$	3	0.66	311	428–444	aactttcccaccaccaccgcgg	agcagcagcaagccagcaagcg	61
RM103	D16221[b]	6	$(GAA)_5$	3	0.70	336	334–340	cttccaattcaggccggctggc	cgccacagctgaccatgcatgc	55
RM104	D24755	1	$(GA)_9$	3	0.62	222	222–238	ggaagaggagagaaagatgtgtgtcg	tcaacagacacaccgccaccgc	61
RM105	D15582	9	$(CCT)_6$	5	0.76	134	131–140	gtcgtcgacccatcggagccac	tggtcgaggtggggatcgggtc	55
RM106	D15600	2	$(GAA)_5$	2	0.50	297	288–297	cgtcttcatcatcgtcgccccg	ggcccatcccgtcgtggatctc	55
RM107	D15785	9	$(GA)_7$	3	0.62	189	180–189	agatcgaagcatcgcgcccgag	actgcgtcctctgggttcccgg	67
RM108	D15910	9	$(GCT)_{10}$	3	0.60	80	69–80	tctcttgcgcgcacactggcac	cgtgcaccaccaccaccaccac	67
RM109	D15964	2	$(AG)_{16}$	6	0.66	97	89–101	gccgccggagagggagagagag	ccccgacgggatctccatcgtc	67
RM110	D22203	2	$(GA)_{15}$	5	0.73	156	138–156	tcgaagccatccaccaacgaag	tcctgtacgccgacgaggtcgag	55
RM111	D22423	6	$(GA)_9$	3	0.54	124	118–126	cacaacctttgagcaccgggtc	acgcctgcagcttgatcaccgg	55
RM112	D22694	2	$(GAA)_5$	2	0.47	128	124–127	gggaggagaggcaagcggagag	agccggtgcagtggacggtgac	55
RM113	D22801	1	$(CA)_8$	4	0.63	151	146–152	caccattgcccatcagcacaac	tcgccctctgctgcttgatggc	55
RM114	D23054	3	$(GA)_7$	2	0.26	209	167–175	cagggacgaatcgtcgccggag	ttggccccctgaggttgtcgg	55
RM115	D23382	6	$(AG)_7$	2	0.47	190	186–188	ttgccgcagtggccgttaccac	aggaggcggcggaaatggaagg	61
RM117	D24346	12	$(AG)_7$	3	0.70	208	203–207	cgatccattcctgctgctcgcg	cgcccccatgcatgagaagacg	55
RM118	D24674	7	$(GA)_8$	3	0.56	156	156–160	ccaatcggagccaccggagagc	cacatcctccagcgacgccgag	67
RM119	D24703	4	$(GTC)_6$	3	0.56	166	166–172	catccccctgctgctgctgctg	cgccggatgtgtgggactagcg	67
RM121	D38933	6	$(CT)_7$	3	0.6	170	258–264	accgtcgccttccactttcccc	ttcggggttgccggtgatgttg	55
RM124	D39769	4	$(TC)_{10}$	3	0.65	271	265–271	atcgtctgcgttgcggctgctg	catggatcaccgagctcccccc	67
RM125	D39885	7	$(GCT)_8$	3	0.61	127	124–136	atcagcagccatggcagcgacc	aggggatcatgtgccgaaggcc	55
RM126	D39942	8	$(GA)_7$	3	0.64	171	167–171	cgcgtccgcgataaacacaggg	tcgcacaggtgaggccatgtcg	55
RM127	D40108	4	$(AGG)_8$	3	0.57	223	209–226	gtgggatagctgcgtcgcgtcg	aggccagggtgttggcatgctg	55
RM128 = OSR27[a]	D40151	1	$(GAA)_9$	4	0.63	157	148–166	agcttgggtgatttcttggaagcg	acgacgaggagtcgccgtgcag	55
RM129	D40184	1	$(CGG)_8$	3	0.57	205	203–214	tctctccggagccaaggcgagg	cgagccacgacgcgatgtaccc	55
RM130	D40295	3	$(GA)_{10}$	3	0.52	85	80–87	tgttgcttgccctcacgcgaag	ggtcgcgtgcttggtttggttc	55
RM131	D40427	4	$(CT)_9$	4	0.7	215	209–217	tcctccctcccttcgcccactg	cgatgttcgccatggctgctcc	61
RM132	D40636	3	$(CGG)_8$	2	0.5	83	83–86	atcttgttgtttcggcggcggc	catggcgagaatgcccacgtcc	67
RM133	D40650	6	$(CT)_8$	3	0.56	230	226–232	ttggattgttttgctggctcgc	ggaacacggggtcggaagcgac	61
RM134	D40973	7	$(CCA)_7$	2	0.28	93	84–93	acaaggccgcgagaggattccg	gctctccggtggctccgattgg	55
RM135 = OSR31[a]	D41754	3	$(CGG)_{10}$	3	0.62	131	119–131	ctctgtctcctcccccgcgtcg	tcagcttctggccggcctcctc	61
RM136	D46916	6	$(AGG)_7$	3	0.62	101	98–104	gagagctcagctgctgcctctagc	gaggagcgccacggtgtacgcc	55
RM137	D47598	8	$(CT)_7$	3	0.52	218	221–227	gacatcgccaccagcccaccac	cgggtggtccccgaggatcttg	55
RM138 = OSR26[a]	D48106	2	$(GT)_{14}$	2	0.47	233	203–217	agcgcaacaaccaatccatccg	aagaagctgcctttgacgctatgg	55
RM139	D48278	11	$(CT)_5$	3	0.63	386	396–410	gagagggaggaagggaggcggc	ctgccatggcagagaaggggcc	55
RM140	D48363	1	$(CT)_{12}$	3	0.47	261	259–263	tgcctcttccctggctcccctg	ggcatgccgaatgaaatgcatg	55
RM141	D48997	6	$(CT)_{12}$	3	0.54	136	126–145	caccaccaccaccacgcctctc	tcttggaggaggaggaggcgcgg	55
RM142	D49041	4	$(CGG)_7$	2	0.47	240	235–238	ctcgctatcgccatcgccatcg	tcgagccatcgctggatggagg	67
RM143	D78609[a]	3	$(CGG)_7$	3	0.6	207	195–207	gtcccgaaccctagcccgaggg	agaggccctccacatggcgacc	67
RM144	X67711[b]	11	$(ATT)_{11}$	6	0.76	237	214–255	tgccctggcgcaaatttgatcc	gctagaggagatcagatggtagtgcatg	55
RM145 = OSR9A[a]	D16340[b]	2	$(GA)_{31}$	4	0.65	215	179–220	ccggtaggcgccctgcagtttc	caaggaccccatcctcggcgtc	67
RM146	X58877[b]	5	$(CT)_{11}$–$(CT)_7$	3	0.46	345	340–348	ctattattccctaacccccatacctcc	agagccactgcctgcaaggccc	55
RM147	D10397	10	$(TTCC)_5(GGT)_5$	2	0.5	97	94–97	tacggcttcggcggctgattcc	cccccgaatcccatcgaaaccc	55
RM149 = OSR7[a]	Z11920[b]	8	$(AT)_{10}$	4	0.73	253	243–303	gctgaccaacgaacctaggccg	gttggaagcctttcctcgtaacacg	55
RM150A	D14000[b]	1	$(CGT)_6(CGG)_5$	n.a.	n.a.	320	n.a.	cacgacgacgacgagcagcagc	gctcgagggagagcgacctgcc	61
R150B	D14000[b]	4	$(CGT)_6(CGG)_5$	n.a.	n.a.		n.a.	cacgacgacgacgagcagcagc	gctcgagggagagcgacctgcc	61
RM150C	D14000[b]	6	$(CGT)_6(CGG)_5$	n.a.	n.a.		n.a.	cacgacgacgacgagcagcagc	gctcgagggagagcgacctgcc	61

Table 2 (continued)

Locus name	GenBank accession number	Map	Repeat type and length	No. of alleles	PIC	PCR product in refer. line	Size range (bp)	Forward primer	Reverse primer	Annealing temp.
RM151 = OSR2[a]	L37528[b]	1	$(TA)_{23}$	10	0.87	197	205–317	ggctgctcatcagctgcatgcg	tcggcagtggtagagtttgatctgc	55
RM152 = OSR34[a]	D22858	8	$(GGC)_{10}$	4	0.75	151	142–157	gaaaccaccacacctcaccg	ccgtagaccttcttgaagtag	55
RM153 = OSR35[a]	D48916	5	$(GAA)_{9}$	4	0.64	201	189–204	gcctcgagcatcatcatcag	atcaacctgcacttgcctgg	55
RM154 = OSR11[a]	D39059	2	$(GA)_{21}$	5	0.80	183	165–169	accctctccgcctcgcctcctc	ctcctcctcctgcgaccgctcc	61
RM155 = OSR32[a]	X07515[b]	12	$(CTT)_{7}$	3	0.52	255	271–273	gagatggccccctccgtgatgg	tgccctcaatcggccacacctc	55
RM156	D39072	3	$(CGG)_{8}$	2	0.50	160	150–160	gccgcaccctcactccctcctc	tcttgccggagcgcttgaggtg	67
RM157A	D39530	3	$(CT)_{11}(TC)_{10}$	6	0.72	137	112–134	cctcctcctcacgaatcccgcc	gggcttcttctccgccggcttc	55
RM157B	D39530	1		3	0.55	106	104–107	cctcctcctcacgaatcccgcc	gggcttcttctccgccggcttc	55
RM158	U12171[b]	1	$(GGC)_{9}$	2	0.36	231	260–263	atggtgagagttgctgccgccg	gatgacgcagaacggcatcgcc	55
RM159	D48905	5	$(GA)_{19}$	7	0.87	248	238–252	ggggcactggcaagggtgaagg	gcttgtgcttctctctctctctctctc	55
RM160 = OSR29[a]	D40093	9	$(GAA)_{23}$	5	0.66	131	86–130	agctagcagctatagcttgctggagatc	tctcatcgccatgcgaggcctc	55
RM161	D41873	5	$(AG)_{20}$	6	0.64	187	165–189	tgcagatgagaagcggcgcctc	tgtgtcatcagacggcgctccg	61
RM162 = OSR18[a]	D48213	6	$(AC)_{20}$	3	0.54	229	217–239	gccagcaaaaccagggatccgg	caaggtcttgtgcggcttgcgg	61
RM165	U33175[b]	1	$(CT)_{13}$	3	0.59	185	180–184	ccgaacgcctagaagcgcgtcc	cggcgaggtttgctaatggcgg	67
RM166	X54046	2	$(T)_{12}$	2	0.50	321	n.a.	ggtcctgggtcaataattgggttacc	ttgctgcatgatcctaaaccgg	61
RM169	D48964	5	$(GA)_{12}$	6	0.71	167	164–194	tggctggctccgtgggtagctg	tcccgttgccgttcatccctcc	67
RM170	D15716	6	$(CCT_{7}$	5	0.69	121	106–119	tcgcgcttcttcctcgtcgacg	cccgcttgcagaggaagcagcc	55
RM171 = OSR33[a]	D84275	10	$(GATG)_{5}$	5	0.74	328	318–343	aacgcgaggacacgtacttac	acgagatacgtacgcctttg	55
RM172	D39653	7	$(AGG)_{6}$	3	0.56	159	159–165	tgcagctgcgccacagccatag	caaccacgacaccgccgtgttg	55
RM173	D30794[b]	5	$(GA)_{9}$	2	0.43	186	186–188	cctacctcgcgatccccccctc	ccatgaggaggaggcggcgatc	67
RM174	D48756	2	$(AGG)_{7}(GA)_{10}$	7	0.63	208	207–222	agcgacgccaagacaagtcggg	tccacgtcgatcgacacgacgg	67
RM175	D15231	3	$(CCG)_{8}$	2	0.36	95	92–95	cttcggcgccgtcatcaaggtg	cgttgagcagcgcgacgttgac	67
RM176	X64619[b]	6	$(CCG)_{8}$	3	0.54	134	133–139	cggctcccgctacgacgtctcc	agcgatgcgctggaagaggtgc	67
RM177	D25142	4	$(AG)_{8}$	2	0.36	195	192–195	ccctcttagacagaggccagaggg	gtagccgaagatgaggccgccg	61
RM178	D49132	5	$(GA)_{5}(AG)_{8}$	4	0.61	117	117–123	tcgcgtgaaagataagcggcgc	gatcaccgttccctccgcctg	67
RM179	D47661	12	$(TG)_{7}$	2	0.14	190	188–190	ccccattagtccactccaccacc	ccaatcagcctcatgcctcccc	61
RM180	D63901[b]	7	$(ATT)_{10}$	6	0.63	110	107–204	ctacatcggcttaggtgtagcaacacg	acttgctctacttgtggtgagggactg	55
RM181 = OSR1[a]	D78506[b]	11	$(CT)_{13}(AT)_{19}$	2	0.47	240	240–242	acgggagcttctccgacagcgc	tatgcttttgccgtgtgccgcg	67
RM182 = OSR4[a]	L10346[b]	7	$(AT)_{16}$	3	0.64	346	336–346	tgggatgcagagtgcagttggc	cgcaggcacggtgccttgtaag	67
RM183	D48890	2	$(GA)_{8}$	n.a.	n.a.	222	n.s.	ggagcgggagagagagcccacg	tgccgatgaaggactgcgacgc	55
RM184	U40708[b]	10	$(CA)_{7}$	3	0.66	219	207–221	atcccattcgccaaaaccggcc	tgacacttggagagcggtgtgg	55
RM185	D24415	4	$(AGG)_{9}$	2	0.50	197	194–197	agttgttgggagggagaaaggcc	aggaggcgacggcgatgtcctc	61
RM186	D39420	3	$(CGG)_{5}$	3	0.44	124	116–130	tcctccatctcctccgctcccg	gggcgtggtggccttcttcgtc	61
RM187	D25465	11	$(AT)_{29}(GT)_{7}$	7	0.84	146	136–164	ccaagggaaagatgcgacaattg	gtggacgctttatattatggg	55
RM188	DD2459	5	$(CA)_{8}$	2	0.26	210	210–212	tccgcctctcctctcgcttccc	gcaacgcacaaccgaaccgagc	61
RM189	D25487	9	$(AG)_{11}$	n.a.	n.a.	126	n.a.	cgtcttccccaacgctaaaa	cgccgggcttcgcttc	61
RM190 = OSR19[a]	X65183[b]	6	$(CT)_{11}$	7	0.80	124	104–124	ctttgtctatctcaagacac	ttgcagatgttcttcctgatg	55
RM191[c]	D47370	2	$(AG)_{7}$	1	0.0	227	n.a.	cccatcctcaccgatctctctaaac	gtgcgcacggaggaggaaaggg	61
RM192[c]	D24331	7	$(TGG)_{5}$	1	0.0	267	n.a.	gcggcggatcatgaattgcgag	cttgttccccggcgtcgagtcc	61
RM193[c]	D24361	6	$(GCT)_{5}$	1	0.0	189	n.a.	cgcctcttcttcctcgcctccg	cgggtccatccccctctcctc	61
RM194[c]	D23574	5	$(GA)_{6}$	1	0.0	250	n.a.	gccctgcttcttgcccaccacc	tccagggagggcaaggctgagc	55
RM195	D46433	8	$(GA)_{9}$–$(CT)_{8}$	2	0.36	311	308–310	agaaagagaggccgtcggcggc	gggctcacccccaaacctgcag	61
RM197	D38221	6	$(ACC)_{7}$	n.a.	n.a.	106	n.a.	gatccgtttttgctgtgccc	cctcctctccgccgatcctg	60
RM226 = OSR3[a]	M29259[b]	1	$(AT)_{38}$	8	0.82	274	264–342	agctaaggtctgggagaaacc	aagtaggatggggcacaagctc	55

Table 2 (continued)

Locus name	Clone name	Map	Repeat type and length	No. of alleles	PIC	PCR product in refer. line	Size range (bp)	Forward primer	Reverse primer	Annealing temp.
RM264	CT109	8	$(GA)_{27}$	8	0.83	178	148–178	gttgcgtcctactgctacttc	gatccgtgtcgatgattagc	55
RM265	CT210	1	$(GA)_{8}$	3	0.54	106	106–110	cgagttcgtccaagtgagc	catccaccattccaccaatc	55
RM206	CT483	2	$(GA)_{19}$	10	0.87	127	121–137	tagtttaaccaagactctc	ggttgaacccaaatctgca	55
RM267	CT606	5	$(GA)_{21}$	6	0.77	156	137–160	tgcagacatagagaaggaagtg	agcaacagcacaacttgatg	55
RM269	CT611	10	$(GA)_{17}$	5	0.64	182	182–188	gaaagcgatcgaaccagc	gcaaatgcgcctcgtgtc	55
RM270	CT624	12	$(GA)_{13}$	5	0.29	108	104–117	ggccgttggttctaaaatc	tgcgcagtatcatcggcgag	55
RM271	CT625	10	$(GA)_{15}$	4	0.72	101	92–105	tcagatctacaattccatcc	tcggtgagacctagagagcc	55
RM272	CT697	1	$(GA)_{9}$	2	0.35	119	119–121	aattggtagagaggggagag	acatgccattagagtcaggc	55
RM273	CT710	4	$(GA)_{11}$	3	0.34	207	199–207	gaagccgtcgtgaagttacc	gtttcctacctgatcgcgac	55
RM274	CT711	5	$(GA)_{15}$-7-$(CGG)_{5}$	3	0.56	160	146–160	cctcgcttatgagagcttcg	cttctccatcactcccatgg	55
RM275	CT714	6	$(GA)_{15}$	2	0.47	114	110–114	gcattgatgtgccaatcg	cattgcaacatcttcaacatcc	55
RM276	CT715	6	$(AG)_{8}A_{3}(GA)_{33}$	7	0.84	149	85–153	ctcaacgttgacacctcgtg	tcctccatcgagcagtatca	55
RM277	CT720	12	$(GA)_{11}$	3	0.56	124	118–124	cggtcaaatcatcacctgac	caaggcttgcaagggaag	55
RM278	CT738	9	$(GC)_{17}$	5	0.77	141	131–146	gtagtgagcctaacaataatc	tcaactcagcatctctgtcc	55
RM279	CT743	2	$(GA)_{16}$	6	0.73	174	148–174	gcgggagagggatctcct	ggctaggagttaacctcgcg	55
RM280	CT780	4	$(GA)_{16}$	5	0.75	155	148–181	acacgatccactttgcgc	tgtgtcttgagcagccagg	55
RM281	CT782	8	$(GA)_{21}$	5	0.77	138	126–138	accaagcatccagtgaccag	gttcttcatacagtccacatg	55
RM282	CT787	3	$(GA)_{15}$	3	0.62	136	128–136	ctgtgtcgaaaggctgcac	cagtcctgtgttgcagcaag	55
RM283	CT788	1	$(GA)_{18}$	3	0.61	151	147–155	gtctacatgtacccttgttggg	cggcatgagagtctgtgatg	55
RM284	CT794	8	$(GA)_{8}$	4	0.63	141	141–149	atctctgatactccatccatcc	cctgtacgttgatccgaagc	55
RM285	CT797	9	$(GA)_{12}$	5	0.65	205	203–213	ctgtgggcccaatatgtcac	ggcggtgacatggagaaag	55
RM286	CT806	11	$(GA)_{16}$	6	0.72	110	99–128	ggcttcatctttggcgac	ccggattcacgagataaactc	55
RM287	CT838	11	$(GA)_{21}$	7	0.83	118	98–118	ttccctgttaagagagaaatc	gtgtatttggtgaaagcaac	55
RM288	CT843	9	$(GA)_{7}G_{6}(GA)_{7}$	2	0.50	125	121–125	ccggtcagttcaagctctg	acgtacggacgtgacgac	55
RM289	CT875	5	$G_{11}(GA)_{16}$	4	0.63	108	88–180	ttccatggcacacaagcc	ctgtgcacgaacttccaaag	55
RM290	CT876	2	$(GA)_{12}$	5	0.68	142	142–182	acccttattcctgctctcctc	gtgctgtagatggaagggag	55
RM291	GT67	5	$(GAT)_{4}GA(GT)_{4}$-$(GT)_{4}$	2	0.47	245	245–251	gttgcactacgtattctgag	gatccagataaatgaggcac	55
RM292	GT194	1	$(GT)_{10}$-G-$(TGA)_{2}$	3	0.46	159	154–162	actgctgttgcgaaacgc	tgcagcaaatcaagctggaa	55
RM293	GT345	5	$(GT)_{20}$	4	0.70	207	202–211	tcgttgggaggtatggtacc	ctttatctgatccttgggaagg	55
RM294 A	GT52	10	$[(GT)_{3}T_{2}AGGGACA]_{2}$	n.a.	n.a.	173	n.a.	ttggcctagtgcctccaatc	gagggtacaacttaggacgca	55
RM294 B	GT52	1		n.a.	n.a.		n.a.	ttggcctagtgcctccaatc	gagggtacaacttaggacgca	55
RM295	CT382	7	$(GA)_{2}A(GA)_{3}G_{2}(GA)_{9}$	5	0.63	180	176–190	cgagacgagcatcggataag	gatctggtggaggggagg	55
RM296	CT615	9	$(GA)_{10}$	2	0.43	123	117–123	cacatggcaccaacctcc	gccaagtcattcactactctgg	55
RM297	CT807	1	$(GA)_{13}$	6	0.75	148	148–191	tctttggaggcgagctgag	cgaagggtacatctgcttag	55
RM298	GT92	7	$(GT)_{10}$	n.a.	n.a.	135	n.a.	ctgatcactggatcgatcatg	catgccaagatgcaacag	55
RM300	GTT711	2	$(GTT)_{14}$	4	0.70	121	121–163	gcttaaggacttctgcgaacc	caacagcgatccacatcatc	55
RM301	GT1	2	$(GT)_{5}G_{2}(GT)_{8}T_{2}(GT)_{3}$	2	0.47	153	151–153	ttactctttgtgtgtgtgtag	ctacgacacgtcatagatgacc	55
RM302	GT2	1	$(GT)_{30}(AT)_{8}$	6	0.76	156	120–191	tcatgtcatctaccatcacac	atggagaagatggaatacttgc	55
RM303	GT4	4	$(GT)_{7}(ATGT)_{6}$	6	0.78	200	143–205	gcatggccaaatattaaagg	ggttggaaatagaagttcggt	55
RM304	GT25	10	$(GT)_{2}(AT)_{10}(GT)_{33}$	6	0.79	160	128–160	tcaaaccggcacatataagac	gatagggagctgaaggagatg	55
RM305	GT71	5	$(GT)_{4}$+degener	2	0.47	203	203–207	tactgccaaaggcgagcttc	gtgagaggctacagctaacc	55
RM306	GT81	1	$(GT)_{18}(AT)_{8}CT(GT)_{6}$	5	0.64	155	135–175	caaggtcaagaatgcaatgg	gccactttaatcattgcatc	55
RM307	GT97	4	$(AT)_{14}(GT)_{21}$	7	0.83	174	124–176	gtactaccgacctaccgttcac	ctgctatgcatgaactgctc	55
RM308	GT119	8	$(AT)_{4}$-6-$(GT)_{2}T_{2}(GT)_{7}$	2	0.15	132	132-146	ggctgcacacgcacactata	ttacgcatatggtgagtaggc	55
RM309	GT126	12	$(GT)_{13}$	3	0.39	169	165–169	gtagatcacgcacctttctgg	agaaggcctccggtgaag	55
RM310	GT137	8	$(GT)_{19}$	6	0.83	105	85–120	ccaaaacatttaaaatatcatg	gcttgttggtcattaccattc	55
RM311	GT147	10	$(GT)_{3}(GTAT)_{8}(GT)_{5}$	4	0.66	179	164–186	tggtagtataggtactaaacat	tcctatacacatacaaacatac	55

Table 2 (continued)

Locus name	Clone name	Map	Repeat type and length	No. of alleles	PIC	PCR product in refer. line	Size range (bp)	Forward primer	Reverse primer	Annealing temp.
RM312	GT165	1	$(ATTT)_4(GT)_9$	4	0.68	97	97–103	gtatgcatatttgataagag	aagtcaccgagtttaccttc	55
RM313	GT167	12	$(GT)_6CA(CG)_5$-6-$(GT)_8$	4	0.51	111	111–117	tgctacaagtgttcttcaggac	gctcacctttgtgttccac	55
RM314	GT177	6	$(GT)_8(CG)_3(GT)_5$	5	0.72	118	108–118	ctagcaggaactcctttcagg	aacattccacacacacacgc	55
RM315	GT254	1	$(AT)_4(GT)_{10}$	3	0.57	133	133–139	gaggtacttcctccgtttcac	agtcagctcactgtgcagtg	55
RM316	GT264	9	$(GT)_8$-$(TG)_9(TTTG)_4(TG)_4$	5	0.79	192	192–212	ctagttgggcatacgatggc	acgcttatatgttacgtcaac	55
RM317	GT316	4	$(GC)_4(GT)_{18}$	4	0.60	155	146–166	catacttaccagttcaccgcc	ctggagagtgtcagctagttga	55
RM318	GT338	2	$(GT)_{15}$	4	0.37	140	134–154	gtacggaaaacatggtaggaag	tcgagggaaggatctggtc	55
RM319	GT363	1	$(GT)_{10}$	2	0.47	134	132–134	atcaaggtacctagaccaccac	tcctggtgcagctatgtctg	55
RM320	GT372	7	$(AT)_{11}GTAT(GT)_{13}$	8	0.85	167	153–254	caacgtgatcgaggatagatc	ggatttgcttaccacagctc	55
RM321	CAT63	9	$(CAT)_5$	2	0.43	200	200–203	ccaacactgccactctgttc	gaggatggacaccttgatcg	55
RM322	CAT65	2	$(CAT)_7$	2	0.47	122	106–112	caagcgaaaatcccagcag	gatgaaactggcattgcctg	55
RM323	CAT69	1	$(CAT)_5$	2	0.47	244	241–244	caacgagcaaatcaggtcag	gttttgatcctaaggctgctg	55
RM324	CAT73	2	$(CAT)_{21}$	6	0.76	175	135–180	ctgattccacacacttgtgc	gattccacgtcaggatcttc	55
RM325A	CAT78	8	$(CAT)_4TAG(CAT)_5$	n.a.	n.a.	201	n.a.	gacgatgaatcaggagaacg	ggcatgcatctgagtaatgg	55
RM325B	CAT78	7		n.a.	n.a.		n.a.	gacgatgaatcaggagaacg	ggcatgcatctgagtaatgg	55
RM327	CAT99	2	$(CAT)_{11}(CTT)_5$	3	0.64	213	207–216	ctactcctctgtccctcctctc	ccagctagacacaatcgagc	55
RM328	CAT118	9	$(CAT)_5$	2	0.47	172	172–181	catagtggagtatgcagctgc	ccttctcccagtcgtatctg	55
RM329	CAT128	1	$(CAT)_7$	3	0.62	153	144–153	cattcggctgctgctattc	gcttgtcacatcttgcacag	55
RM330A	CAT130	10	$(CAT)_5$	n.a.	n.a.	177	n.a.	caatgaagtggatctcggag	catcaatcagcgaaggtcc	55
RM330B	CAT130	8	$(CAT)_5$	n.a.	n.a.		n.a.	caatgaagtggatctcggag	catcaatcagcgaaggtcc	55
RMRM331	CTT13	8	$[(CTT)_4GTT]_2(CTT)_{11}$	5	0.69	176	149–179	gaaccagaggacaaaaatgc	catcatacatttgcagccag	55
RM332	CTT38	11	$(CTT)_5$-12-$(CTT)_{14}$	5	0.65	183	162–183	gcgaaggcgaaggtgaag	catgagtgatctcactcaccc	55
RM333	CTT39	10	$(TAT)_{19}(CTT)_{19}$	8	0.83	191	164–215	gtacgactacgagtgtcaccaa	gtcttcgcgatcactcgc	55
RM334	CTT48	5	$(CTT)_{20}$	7	0.83	182	146–197	gttcagtgttcagtgccacc	gactttgatctttggtggacg	55
RM335	CTT50	4	$(CTT)_{25}$	8	0.84	104	104–155	gtacacacccacatcgagaag	gctctatgcgagtatccatgg	55
RM336	CTT53	7	$(CTT)_{18}$	6	0.79	154	148–193	cttacagagaaacggcatcg	gctggtttgtttcaggttcg	55
RM337	CTT64	8	$(CTT)_4$-19-$(CTT)_8$	3	0.66	192	156–192	gtaggaaaggaagggcagag	cgatagatagctagatgtggcc	55
RM338	CTT71	3	$(CTT)_6$	2	0.26	183	180–183	cacaggagcaggagaagagc	ggcaaaccgatcactcagtc	55
RM339	CTT85	8	$(CTT)_8CCT(CTT)_5$	4	0.7	148	142–160	gtaatcgatgctgtgggaag	gagtcatgtgatagccgatatg	55
RM340	CTT101	6	$(CTT)_8T_3(CTT)_{14}$	4	0.67	163	119–189	ggtaaatggacaatcctatggc	gacaaatataagggcagtgtgc	55
RM341	CTT77	2	$(CTT)_{20}$	4	0.66	172	136–172	caagaaacctcaatccgagc	ctcctcccgatcccaatc	55
RM342A	CAT83	8	$(CAT)_{12}$	n.a.	n.a.	141	n.a.	ccatcctcctacttcaatgaag	actatgcagtggtgtcaccc	55
RM342B	CAT83	9		n.a.	n.a.		n.a.	ccatcctcctacttcaatgaag	actatgcagtggtgtcaccc	55
RM343	CAT112	6	$(CAT)_5(CAC)_5CAT(CAC)_4$	2	0.5	233	230–233	ccacgaacccttttgcatc	gtgatgatgcgtcggttg	55
RM344	CTT60	8	$(TTC)_2$-5-$(CTT)_3$-$(CTT)_4$	2	0.36	163	160–163	cagagacaatagtccctgcac	gtaggaggagatggatgatgg	55
RM345	CTT97	6	$(CTT)_9$	4	0.56	167	152–167	attggtagctcaatgcaagc	gtgcaacaaccccacatg	55
RM346	M7[a]	7	$(CTT)_{18}$	6	0.75	175	140–175	cgagagagcccataactacg	acaagacgacgaggagggac	55
RM347	M12[a]	3	$(GGC)_5(AT)_7$	5	0.7	207	214–300	cacctcaaacttttaaccgcac	tccggcaagggatacggcgg	55
RM348	M16[a]	4	$(CAG)_7$	2	0.47	136	130–139	ccgctactaatagcagagag	ggagctttgttcttgcgaac	55
RM349	M22[a]	4	$(GA)_{16}$	5	0.7	136	132–146	ttgccattcgcgtggaggcg	gtccatcatccctatggtcg	55
RM350	M26[a]	8	$(CT)_{10}$	4	0.67	208	200–210	tgatcgtcgcgattcccggc	ccccaccctgcgcctctccc	55
RM351	M10[a]	7	$(CCG)_9(CGAAG)_4$	2	0.5	134	129–134	ccatcctccaccgcctctcg	tggaggaaggaaaggggacg	55
RM32	GA264	8	$(TC)_3A(CT)_9(TC)_5$	n.a.	n.a.	168	n.a.	agtctacgtggtgtacacgtgg	tgcggcctgccgtttgtgag	55
RM52	GA558	8	$(AG)_{19}$	n.a.	n.a.	240	n.a.	ctactcgcgcgtggagtt	tgtcttactggtgaagctgg	55
RM71	ATT20	2	$(ATT)_{10}T(ATT)_4$	4	0.63	149	125–149	ctagaggcgaaaacgagatg	gggtgggcgaggtaataatg	55

Table 2 (continued)

Locus name	Clone name	Map	Repeat type and length	No. of alleles	PIC	PCR product in refer. line	Size range (bp)	Forward primer	Reverse primer	Annealing temp.
RM72	ATT35	8	$(TAT)_5C(ATT)_{15}$	7	0.85	166	152–198	ccggcgataaaacaatgag	gcatcggtcctaactaaggg	55
RM85	TCT114	3	$(TGG)_5(TCT)_{12}$	4	0.59	107	85–107	ccaaagatgaaacctggattg	gcacaaggtgagcagtcc	55
RM86	TCT117	1	$(CTT)_{16}$	n.a.	n.a.	160	n.a.	tacacctcatcgatcaatcg	ctttcgaatctgaagatc	55
RM87	TCT121	5	$(CTT)_3T(CTT)_{11}$	5	0.73	151	124–151	cctctccgatacaccgtatg	gcgaaggtacgaaaggaaag	55
RM88	TCT116	8	$(TCT)_{11}$	n.a.	n.a.	180	n.a.	actcatcagcatggccttgctc	taatgctccacctcaccac	55

[a] OSR loci were previously identified by Akagi et al. (1996); clones designated with an M prefix were identified by Blight et al. (1999)
[b] Microsatellites in known or putative genes as described by Cho et al. (1999)
[c] Markers are monomorphic among *O. sativa* varieties but polymorphic in the SL cross

et al. (1997). This map was originally constructed based on the population of doubled-haploid lines (DH1 population) derived from the inter-subspecific cross between IR64 (*indica*) and Azucena (*japonica*) varieties (Huang et al. 1994) and contains 145 RFLP markers that provide anchor points on the high-density molecular genetic map reported by Causse et al. (1994). The current linkage map consists of 312 SSLPs, as shown in Fig. 1. Of the 188 new SSLPs, 141 were mapped directly onto the DH1 (IR64 × Azucena) population and 68 were mapped onto the RIL2 (Lemont × Teqing) population, including 17 markers that were not polymorphic in the DH parents. An additional 19 polymorphisms were mapped onto RIL1 (Milyang23 × Gihobyeo) and ten were mapped onto the SL population, including five SSLP markers that were polymorphic only at the interspecific level.

Microsatellites in known and putative genes

Twenty six SSLP loci mapped in this study have been identified within or adjacent to known rice genes (markers with the letter [b] in Table 2 in this paper and in Table 4 in a companion paper by Cho et al. 1999). For example, RM150 resides in a lipoxigenase gene, and three independently segregating loci were identified by primers for RM150, suggesting the presence of several copies of this gene in the genome. In two cases, microsatellite markers occurred in genes coding for amylases; marker RM176 for the alpha-amylase gene was mapped to the end of the of the long arm of chromosome 6, where Amy 2A had been previously mapped on the DH population (Huang et al. 1994), and RM182 mapped to chromosome 7 and was identified in the beta-amylase gene that was initially reported by Akagi et al. (1996) as the monomorphic marker OSR4.

In addition, three SSLP markers derived from genomic clones showed significant similarity to rice sequences in the GenBank database. Clone GT254, which was mapped as marker RM315 on chromosome 1, showed 91% homology to a 70-bp segment of the rice glycine-rich cell-wall protein gene, Angrp-1 (Acc. #U40708). Interestingly, this sequence itself contained a microsatellite with a GT repeat unit that was mapped as SSLP locus RM184 on chromosome 10. This suggests that there are two different locations in the rice genome where sequences related to the glycine-rich cell-wall protein gene are found. Two other markers, RM324 (CAT73) and RM334 (CTT48), showed significant but quite short (50–75 bp) homology with an *Oryza longistaminata* receptor kinase gene and the *O.sativa* putative ADH-glucose pyrophosphorylase subunit, SH2 gene, respectively.

Genome coverage and distribution of SSRs along rice chromosomes

The 312 markers reported here provide genome-wide coverage with an average density of one SSLP marker

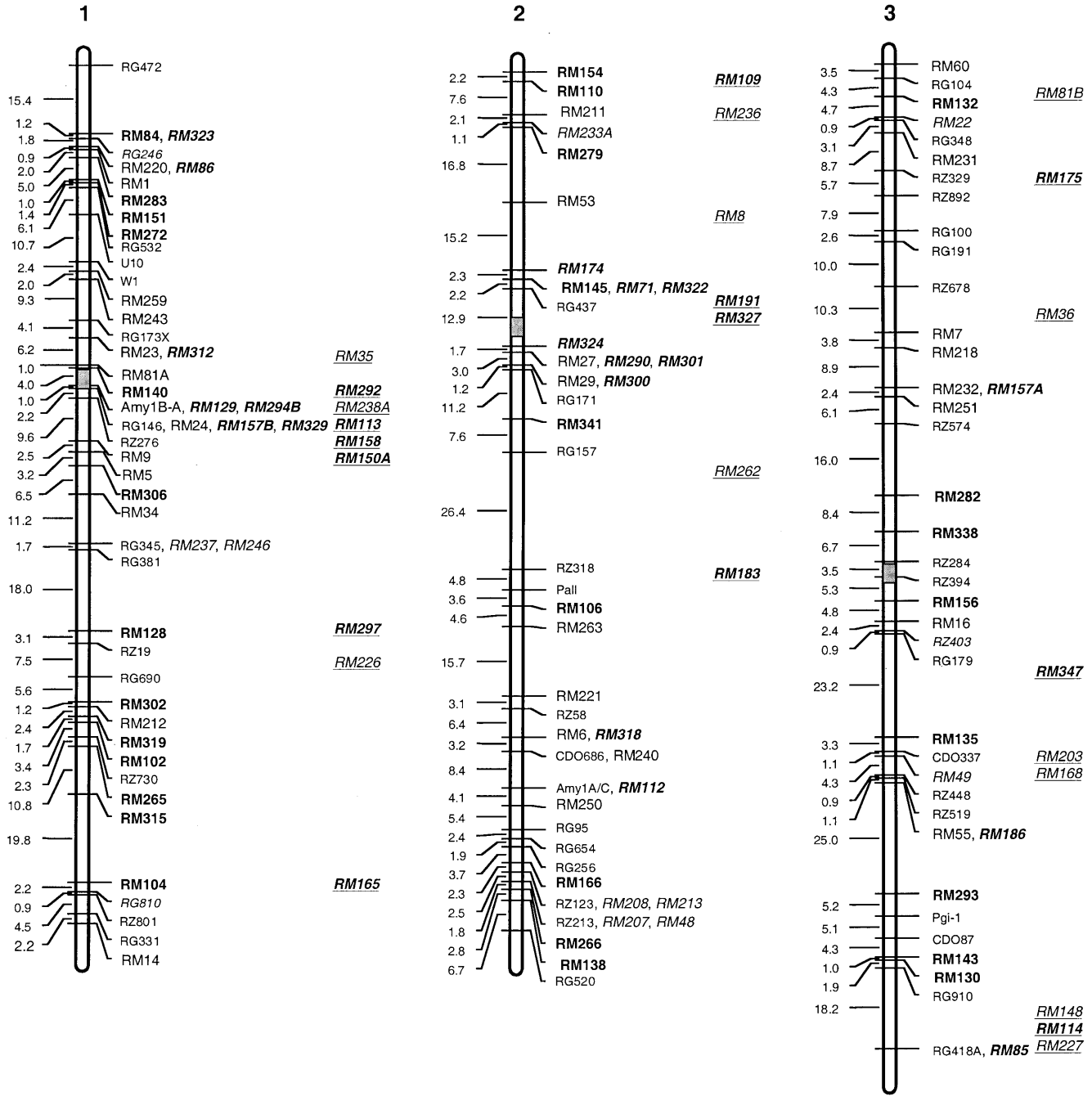

Fig. 1 Molecular genetic map of rice. The framework is based on the IR64/Azucena doubled-haploid population (DH1). Short arms of chromosomes are at the top. Approximate positions of centomeres are indicated by *dark boxes* on the chromosome bars. Framework markers (those ordered at LOD score >2.0) are shown in *regular script* and the remainder are in *italics*. Markers mapped onto other populations and integrated into the DH1 map via anchored RFLP markers are *underlined* and placed to the side of the DH map. Microsatellite loci have the designation "*RM*" for *R*ice *M*icrosatellites. New SSLP loci identified in this study are shown in *boldface*

every 6 cM. In general, SSLP markers are relatively evenly distributed throughout the linkage maps of the 12 rice chromosomes without obvious clustering in centromeric or telomeric regions (Fig. 1). We detected no obvious biases in the localization of microsatellite loci with different motifs or derived from different origins. In most cases new markers were mapped inside the boundaries of the previously reported framework map or very close to the most distal RFLP markers. The addition of a sufficient number of new markers allowed us to populate several regions previously lacking in microsatellite markers. Three SSLP markers, including RM190, which corresponds to the *waxy* gene (Ayres et al. 1997), were mapped onto the end of the short arm of chromosome 6, two markers (RM337 and RM152) were placed at the top of chromosome 8, and two (RM147 and RM333) on the bottom of chromosome 10 (Fig. 1). In all cases, these new distal markers corresponded to regions already defined by RFLP markers on the high-density maps

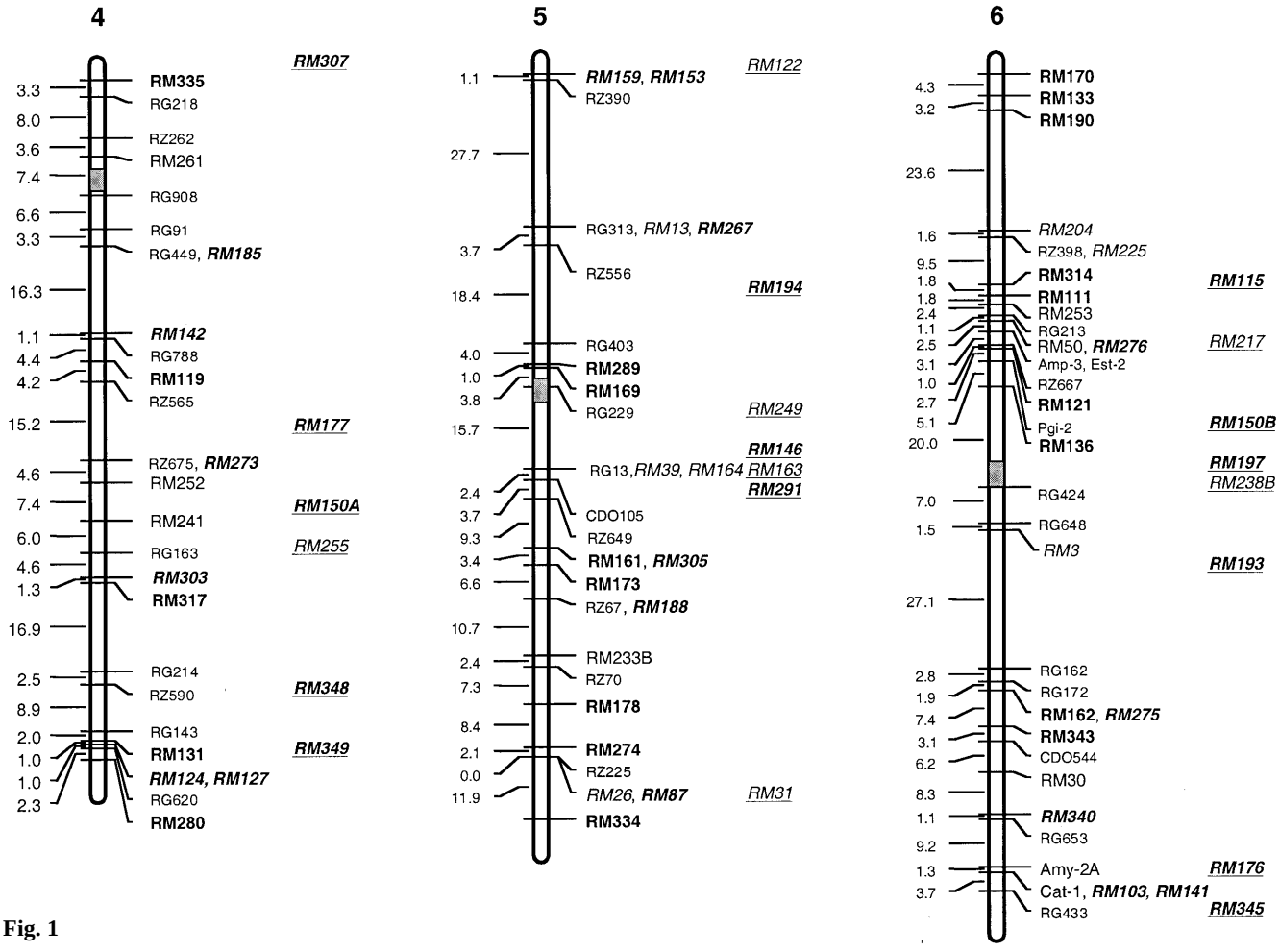

Fig. 1

(Causse et al. 1994; Harushima et al. 1998). The resulting overall map length for the IR64/Azucena DH population is approximately 1822 cM.

Although the number of SSLP markers mapped onto individual chromosomes is roughly proportional to their cytogenetic length (Fukui and Iijima, 1991) with no significant differences between observed and expected number of markers as indicated by Z-score analysis (data not shown), the distribution of SSLP markers across the chromosomes is not uniform. There are several regions with a low density of markers, appearing as large gaps on the map. In some cases, these intervals coincide with comparable regions of the chromosomes on the high-density SL map (Causse et al. 1994). For example, a 26.5 cM interval between markers RG179 and CDO337 on the DH map for chromosome 3 corresponds to a 21.5 cM gap between these loci on the SL map, and a 23.1 cM gap between RG20 and RM25 on chromosome 8 is precisely aligned with the biggest marker interval for this chromosome on the SL map. Interestingly, the gap of 27.7 cM at the top of chromosome 6 between markers comprising the *waxy* region and RFLP marker RZ398, corresponds to a region with a high recombination rate previously described by Causse et al. (1994). In some other cases, extension of the map was observed only for the DH cross. These data suggest that the large distances between many of the markers on the DH map are likely to be the result of a comparatively higher recombination rate in the doubled-haploid lines than in the interspecific cross. Therefore, we have no evidence that these regions with a very low density of SSLP markers (appearing as gaps) correspond to physical segments devoid of microsatellite sequences.

To examine the distribution of microsatellite polymorphism throughout the genome, GeneFlow software was used. This facilitated the detection of several clearly defined regions covered by SSLP markers with uniformly higher or lower levels of polymorphism. For example, the long arms of chromosomes 3 and 6 contain clusters of markers with only 2–4 alleles, while chromosome 11 preferentially contains markers with more than six alleles (Fig. 2). Although this finding is preliminary and deserves further investigation at a higher level of resolution, it suggests that polymorphism may not be randomly distributed in the rice genome.

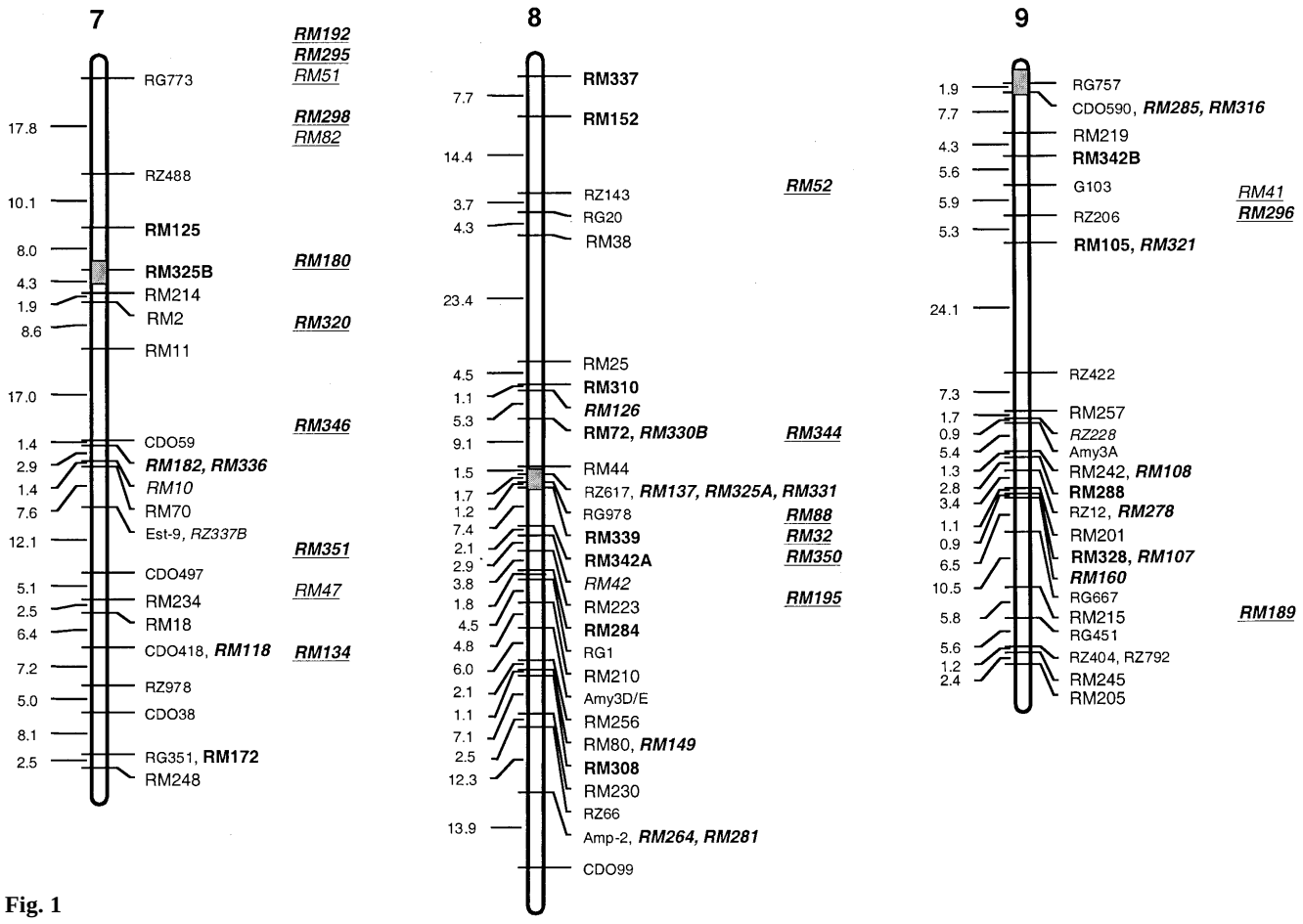

Fig. 1

Discussion

Two major sources of microsatellite-containing sequences were used in this study to develop SSLP markers in rice - the GenBank database and a Tsp509-digested small-insert genomic library. The GenBank database search was focused on the identification of all possible di-, tri- and tetra-nucleotide simple sequence repeats with a number of repeat units greater than five. The set of 12 532 rice DNA sequences consisted of partially sequenced cDNA clones and several known rice genes for which complete sequence data were available. This set of DNA sequences allowed the development of SSLP markers for some genes of known or putative function and provided a good opportunity to estimate the frequency, length and position of microsatellite sequences with different motifs in or near expressed genes in the rice genome.

It was found that 60% of EST-derived microsatellite sequences in rice were represented by the following four trinucleotide motifs: CCG, ACG, AGG and ACC. Similar observations were made in maize, where trinucleotide motifs comprised about 50% of the SSR-containing sequences extracted from sequence databases, with CCG/GGC and CCT/AGG motifs being the most abundant (Chin et al. 1996). Interestingly, there were only 32 tetranucleotide SSRs among the 2070 (1.5%) microsatellite-containing sequences identified for rice in comparison to a much higher proportion of this class of microsatellites found during similar database searches in maize (27%) (Chin et al. 1996) or in rat (25%) (Serikawa et al. 1992). A notable deficiency of AT/TA repeats, which were predicted to be the most frequent class of microsatellite sequences in plant genomes (Mongante and Olivieri 1993; Wang et al. 1994), suggests that they are rare in the portion of the rice genome captured as cDNA. It is known from the detailed compositional analysis of SSR sequences from primates that AT-rich di- and tri-nucleotides occur predominantly in non-coding regions, frequently being associated with repetitive DNA (Jurka and Petiyagoda 1995). In this respect, the database search in this study was limited by the set of rice DNA sequences available in GenBank.

As an alternative, screening of genomic libraries allows the identification of unlimited numbers of clones containing diverse microsatellite motifs from a more random representation of the genome. In this study, a *Tsp*509-digested small-insert library was used to isolate microsatellite sequences containing four different motifs and to evaluate the utility of the resulting markers for mapping. While the frequency of the (GT)*n*-containing sequences was relatively high, a large number of puta-

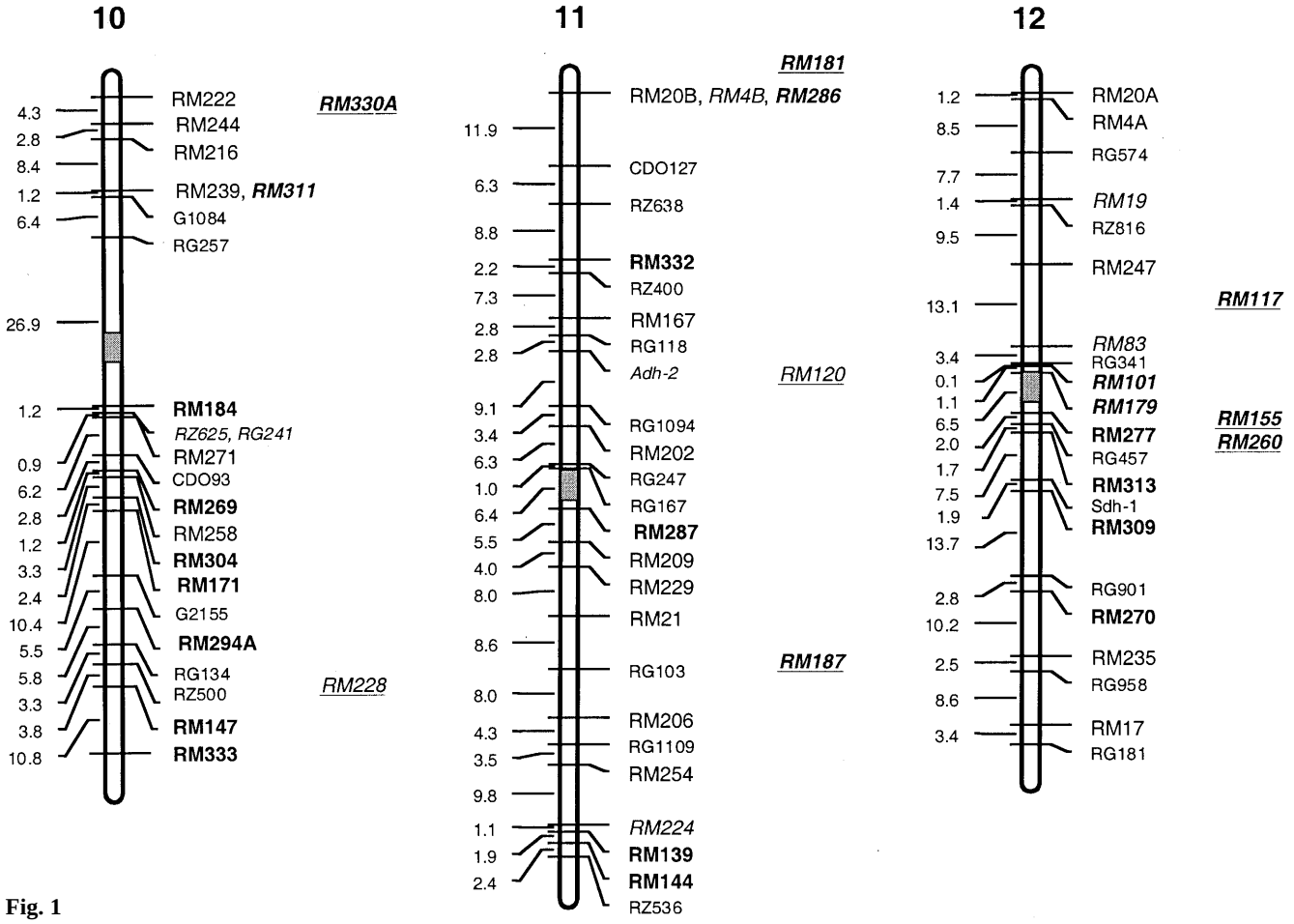

Fig. 1

tive positive GT clones contained only very short tracts of repeat units (less than five). Nevertheless, there exist GT-containing sequences with long tracts of pure GT repeats or with adjoining AT repeats, which have been shown to be highly polymorphic markers. Interestingly, in maize (Taramino and Tigney 1996) and eastern white pine (Echt et al. 1996) stretches of GT repeats were also shorter compared to GA repeats and were frequently associated with the AT repeats.

Among the trinucleotides, poly(CTT)*n* repeats comprise a moderately abundant class of SSR sequences with long tracts of repeat units, which can be considered as a valuable source of informative genetic markers in rice. In contrast to the CTT motif, (CAT)*n* repeats represent a class of relatively short and highly degenerate microsatellite sequences in the rice genome with a low potential for length variation. Comparison of trinucleotide repeats in the human genome showed that (CAT)*n* loci are three times less polymorphic than (ATT)*n* or (CTT)*n* loci (Gastier et al. 1995), which is similar to our observations in rice. Considering the coding equivalency of TAC (opposite reading of CAT) to ATG, which is the initiation codon on the complementary DNA strand, the lack of long microsatellite sequences with this SSR motif of variable length in eukaryotic genomes can be explained by the rarity and constrained placement of these codons.

As described in Cho et al. (1999), GenBank-derived microsatellites had lower variability values (number of repeat units, number of alleles, allele size range and expected genetic diversity) than microsatellites isolated from genomic libraries. This difference reflected the constraints on DNA sequence variation in transcriptionally active portions of the genome. From this point of view, screening of genomic libraries was more efficient than the GenBank search, since more informative markers capable of detecting more genetic differences were developed based on genomic clones. On the other hand, the GenBank-derived markers produced a high proportion of intra-subspecifically conserved microsatellite markers with distinct allele patterns for the *indica* and *japonica* subspecies, and these can be useful for evolutionary studies and applications in breeding programs involving the two different subspecies of *O. sativa*.

In rice, as in many other species for which full-genome SSLP-based maps are available, microsatellite markers are distributed relatively uniformly throughout the genome, and in this case provide good coverage of all 12 chromosomes. There are no obvious biases in terms of chromosomal location for SSLP markers containing different motifs or derived from different origins. Nevertheless, there are some regions on the map with a poor representation of microsatellite markers or else con-

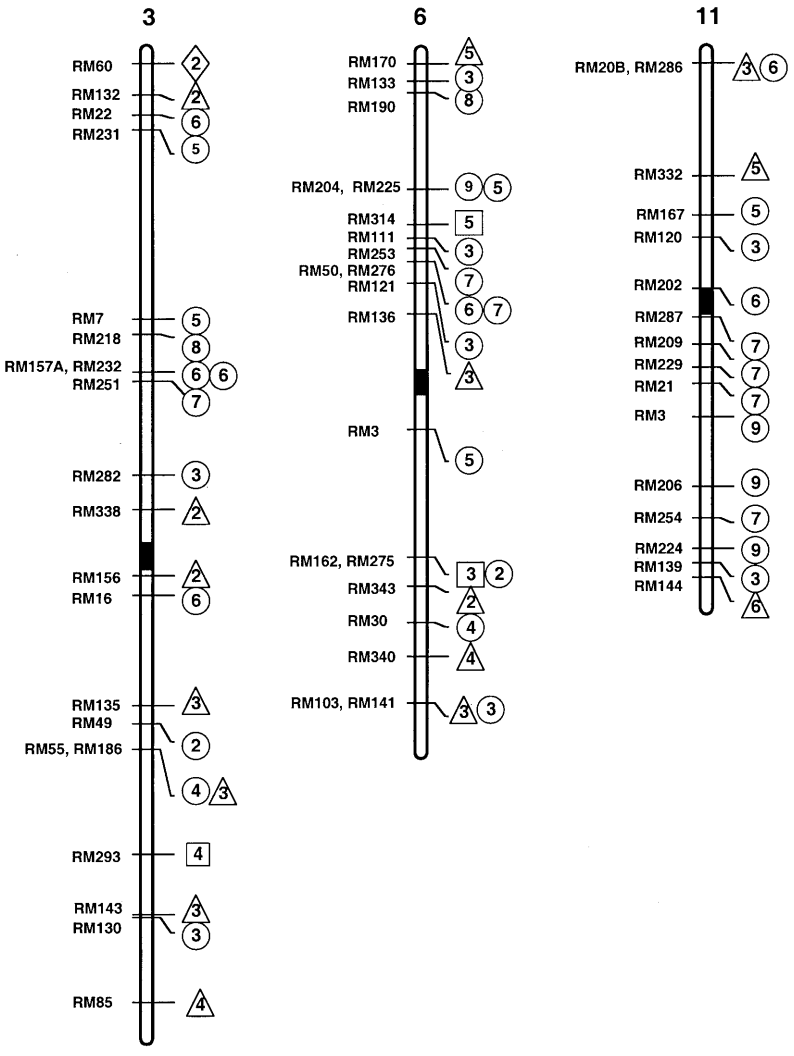

Fig. 2 Maps of three rice chromosomes showing the distribution of polymorphism at microsatellite loci. *Numbers* inside of geometric shapes indicate the number of alleles observed at each locus for the 13 rice cultivars analyzed. SSLP markers with the GA motif appear in *Circles*; GT motifs appear in *Squares*, trinucleotides appear in *Triangles* and other types of SSR motifs appear in *Diamonds*. Orientation of the chromosomes and positions of the markers are as in Fig. 1

taining markers with a low level of genetic variability. The question of whether these regions reflect differences in the density of microsatellite sequences along the physical length of the chromosomes or differences in the rate of recombination is still open. Non-random distribution of microsatellite polymorphism has been detected in the mouse, where chromosomes 10 and X contained fewer SSLP markers and showed substantially lower polymorphism rates than the other chromosomes (Dietrich et al. 1996). In hexaploid wheat, fewer SSLP markers and a lower number of alleles per SSLP locus were detected for the D-genome (Bryan et al. 1997; Röder et al. 1998). A positive correlation between the number of SSLP and RFLP markers developed within the A, B and D genomes suggested that different amounts of DNA polymorphism are present in the three genomes of this allohexaploid species (Röder et al. 1998). In rice, the intensive selection for agronomically important traits during the process of domestication and breeding, which has been accompanied by some population bottlenecks, might have led to a non-random distribution of allelic diversity along chromosomes. It is possible that chromosomal segments with a low level of SSLP diversity could correspond to genomic regions where microsatellite loci are linked to genes of agricultural importance and are affected by linkage drag. Alternatively, the polymorphism distribution may reflect some structural or functional properties of the DNA in the less-variable segments of the genome. In this case, a comparison of cultivated varieties with closely related wild species of rice that have not been subjected to artificial selection, as well as an evaluation of different types of markers, would provide the opportunity to resolve the issue.

Our success in developing microsatellite markers with different motifs and extracted from different sources provides evidence that this approach can be effective for further saturation of the microsatellite map of the rice genome. Advanced technologies in sequencing and the growing pool of published sequence information will provide a major resource for the future development of these PCR-based markers. Further investigation of the distribution and variability of microsatellite sequences can provide new information about the organization of this class of repetitive DNA elements in the rice genome, as well as valuable information for researchers wishing to use microsatellite markers for genetic studies and breeding applications.

Table 2 and additional information about the polymorphism potential of over 300 rice microsatellite markers (number of alleles, polymorphism information content, range of variation) can be found in Cho et al. (1999) and in the RiceGenes database (http://ars-genome.cornell.edu/rice/). Primers for the previously developed 121 markers and 188 reported here are available from Research Genetics (http://www.resgen.com/).

Note: RM260 was previously reported on chromosome II (Chen et al., 1997) but the map position is corrected in this study and RM260 now appears on chromosome 12.

Acknowledgements We gratefully acknowledge Matthew Blair, Xiuli Chen and Julie Ho for a critical review of this manuscript and Carole Morehouse and Lois Swales for formatting assistance. We are grateful for financial support from the Rockefeller Foundation (RF 95001, no. 315), the USDA (NRI Competitive Grant 93–37300–8703), the USDA-ARS, the Texas Advanced Technology Development Experiment Station, and the Institute of Biosciences and Technology at Texas A&M University.

References

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 94: 61–67

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB (1995) Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci 35: 1439–1445

Altschul SF, Gish W, Miller W, Myers EW, DJ Lipman (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94: 773–781

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

Bligh HFJ, Blackhall NW, Edwards KJ, McClung AM (1999) Using amplified fragment length polymorphisms and simple sequence length polymorphisms to identify cultivars of brown and white milled rice. Crop Sci (in press)

Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB, MD Gale (1997) Isolation and charcterization of microsatellites from hexaploid bread wheat. Theor Appl Genet 94: 557–563

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

Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor Appl Genet 95: 53–567

Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39: 866–873

Cho YG, McCouch SR, Kuiper M, Kang M-R, Pot J, Groenen JTM, Eun MY (1998) Integrated map of AFLP, SSLP and FRLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). Theor Appl Genet 97: 370–380

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, Park WD, Ayres N, Cartinhour S, McCouch SR (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated molecular genetic linkage map of soybean genome. Crop Sci 39:1464–1490

Dellaporta SL, Wood T, Hicks TB (1983) A plant DNA mini preparation: version II. Plant Mol Biol Rep 1: 19–21

Depeiges A, Goubely C, Lenoir A, Cocherel S, Picard G, Raynal M, Grellet F., Delseny M (1995) Identification of the most represented repeated motifs in *Arabidopsis thaliana* microsatellite loci. Theor Appl Genet 91: 160–168

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Mirissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5264 microsatellites. Nature 380: 152–154

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ, Evans CA, DeAngelis MM, Levinson DM, Kruglyak L, Goodman N, Copelang NG, Jenkins NA, Hawkins TL, Stein L, Page DC, Lander ES (1996) A comprehensive genetic map of the mouse genome. Nature 380: 149–152

Echt CS, May-Marquardt P, Hseih M, Zahorchak R (1996) Characterization of microsatellite markers in eastern white pine. Genome 39: 1102–1108

Fukui K, Iijima K (1991) Somatic chromosome map of rice by imaging methods. Theor Appl Genet 81: 589–596

Gastier, JM, Pulido JC, Suden S, Bordy T, Buetow KH, Murray JC, Weber JL, Hudson TJ, Sheffeld VC, Duyc GM (1995) Survey of trinucleotide repeats in the human genome: assessment of their utility as genetic markers. Hum Mol Genet 4: 1829–1836

Groenen MAM, Crooijmans RPMA, Veenendaal A, Cheng HH, Siwek M, van der Poel JJ (1998) A comprehensive microsatellite linkage map of the chicken genome. Genomics 49: 265–274

Harushima Y, Jano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T, Lin SY, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice genetic linkage map with 2275 markers using a single F_2 population. Genetics 148: 479–494

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

Jacob HJ, Brown DM, Bunker RK, Daly MJ, Dzau VJ, Goodman A, Koike G, Kren V, Kurtz T, Lernmark A, Levan G, Mao Y, Pettersson A, Pravenec M, Simon JS, Szpirer C, Szpirer J, Trolleit MR, Winer ES, Lander ES (1995) A genetic linkage map of the laboratory rat, *Rattus norvegicus*. Nature Genet 9: 63–69

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11: 424–429

Jurka J, Pethiyagoda C (1995) Simple repetitive DNA sequences from primates: compilation and analysis. J Mol Evol 40: 120–126

Lander ES, Green P, Abrahamson J, Barlow MJ, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174–181

Liu Z-W, Biyashev RM, Saghai Maroof (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93: 869–876

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 35: 89–99

Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, Francisco LV, Gibbs M, Auirre GD, Ostrander EA (1997) A linkage map of the canine genome. Genomics 46: 326–336

Milbourne D, Meyer RC, Collins AJ, Ramsay LD, Gebhardt C, Waugh R (1998) Isolation, characterization and mapping of simple sequence repeat loci in potato. Mol Gen Genet 259: 233–245

Mongante M, Olivieri AM (1993) PCR-amplified microsatellites as markers for plant genetics. Plant J 3: 175–182

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

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

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1: 215–222

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H., Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Sasaki T, Song J, Koga-Ban Y, Matsui E, Fang F, Higo H, Magasaki H, Hori M, Miya M, Murayama-Kayano E, Takiguchi T, Takasuga A, Niki T, Ishimaru K, Ikeda H, Yamamoto Y, Mukai Y, Ohta I, Miyadera N, Havukkala I, min obe Y (1994) Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library. Plant J 6: 615–624

Schmidt T, Heslop-Harrison JS (1996) The physical and genomic organization of microsatellites in sugar beet. Proc Natl Acad Sci USA 95: 8761–8765

Serikawa T, Kuramoto T, Hilbert P, Mori M, Yamada J, Dubay CJ, Lindpainter K, Ganten D, Guenet J-L, Lathrop GM, Beckmann JS (1992) Rat gene mapping using PCR-analyzed microsatellites. Genetics 131: 701–721

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39: 277–287

Wang Z, Weber JL, Zhong G, 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

Yamamoto K, Sasaki T (1997) Large scale EST sequencing in rice. Pl Mol Biol 35: 135–144