Theor Appl Genet (2002) 105:449–457
DOI 10.1007/s00122-002-0921-5

Matthew W. Blair · Veena Hedetale
Susan R. McCouch

Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.)

Received: 19 June 2001 / Accepted: 25 November 2001 / Published online: 23 May 2002

Abstract Four multiplex panels of fluorescent-labeled rice microsatellite markers were designed to survey allelic diversity at a total of 27 simple sequence repeat loci. Automated fragment detection, size calling and binning were used to identify the microsatellite alleles. The panels were tested on 72 accessions representing the diversity of ecotypes and isozyme groups in cultivated rice (*Oryza sativa* L.). Genetic diversity was evaluated by estimating the distribution and range of allele sizes, as well as the polymorphism information content (PIC), for each microsatellite locus. The multiplex panels were shown to be useful for fingerprinting and for clustering rice varieties, and genetic associations determined in this analysis agreed well with previous isozyme- and RFLP-based studies of these genotypes. Because of the wide cross-section of germplasm represented, we expect that the allele size-ranges reported here are close to the maximum for cultivated rice. In the future, panels of microsatellites from each rice chromosome will be made to facilitate the mapping of segregating populations and the identification of genes and QTLs underlying traits of interest.

Keywords Rice (*Oryza sativa* L.) · Fluorescent-labeled microsatellites · Allelic diversity · Multiplex panels

Communicated by F. Salamini

The first two authors made equal contributions to this manuscript

M.W. Blair · V. Hedetale · S.R. McCouch (✉)
Department of Plant Breeding, 252 Emerson Hall,
Cornell University, Ithaca, NY 14853-1901, USA
e-mail: SRM4@cornell.edu
Tel.: +1-607-255-0420, Fax: +1-607-255-6683

Present addresses:
M.W. Blair, CIAT, International Center for Tropical Agriculture,
A. A. 6713, Cali, Columbia, S.A.

V. Hedetale, CPGL Monsanto Enterprises Ltd. Private,
#60, II Floor, 8th Main, 18th Cross Opposite Post Office,
Bangalore 560 5509, India

Introduction

Microsatellites are highly informative polymerase chain reaction (PCR)-based markers that detect length polymorphisms at loci with simple sequence repeats (Powell et al. 1996). Over 500 microsatellite markers have been developed for rice (*Oryza sativa* L.), and their chromosomal location and level of polymorphism have been determined (reviewed in Temnykh et al. 2001). Rice microsatellites have been found to vary in the polymorphism they detect depending on the length and sequence of the repeat motif they contain and their position in coding versus non-coding segments of the genome (Cho et al. 2000; Temnykh et al. 2000). Microsatellites have been ideal markers for characterizing genetic diversity in cultivated rice at both the inter-varietal (Yang et al. 1994; Xiao et al. 1996) and intra-varietal (Olufowote et al. 1997) levels. In addition, many microsatellite primer pairs reliably amplify loci in a range of closely related non-domesticated rice species (Wu and Tanksley 1993; Panaud et al. 1996; Harrington 2000; Moncada et al. 2001; Ishii et al. 2001) and sometimes in other more-distantly related grass species (Ishii and McCouch 2000; Chen et al. submitted). The extent of allelic diversity detected at rice microsatellite loci depends on the specific loci being assayed, the number and diversity of the genotypes sampled and, to a certain degree, the sensitivity of the technique used to detect small molecular-weight differences.

Fluorescent labeling of markers, using different dye colors, and automated detection on DNA sequencers has been shown to be an inexpensive and efficient method of genotyping (Ziegle et al. 1992; Mansfield et al. 1994). Multiplexing of markers can increase throughput by allowing many polymorphic DNA fragments, representing multiple loci from across the genome, to be visualized per lane. In multiplex panels, markers labeled with the same fluorophore dye must be of different size, while markers that overlap in size must be distinguished by using different colored dyes to label the microsatellites (Ziegle et al. 1992). Several microsatellites can be

pooled post-amplification or amplified together in a common PCR reaction (Mitchell et al. 1997). If a universal multiplex panel is desired, information about the size-range of microsatellite alleles at each locus in diverse germplasm accessions is essential when determining which markers to multiplex and what dye colors to use for the labeling of individual markers.

Although automated fluorescent detection of microsatellites is in common use for human genome studies (Reed et al. 1994; Schwengel et al. 1994) and has been developed for several plant species such as *Arabidopsis* (Ponce et al. 1999), *Brassica* sp. (Kresovich et al. 1995; Mitchell et al. 1997), corn (Smith et al. 1997), soybean (Diwan and Cregan 1997) and tomato (Bredemeijer et al. 1998), this method has not been extensively applied to the analysis of rice varieties. Panels of compatible microsatellites for multiplex fluorescent detection would be useful for studies of cultivar diversity and high-throughput genotyping in this important crop species. The objective of this research was to design and test multiplex panels of fluorescent-labeled rice microsatellites, to determine the range of alleles found in a wide array of cultivated rice varieties and to evaluate the potential of this technique for fingerprinting and clustering of rice genotypes.

Materials and methods

Plant material

A total of 72 rice accessions were evaluated in this study, including two sets of germplasm important to rice researchers. In the first part of the study, we genotyped a set of 13 parental varieties (referred to as germplasm set 1) that have been used to create seven mapping populations studied around the world for genetic linkage analysis of rice (for a review see Cho et al. 2000). In the second part of the study, a set of 59 genotypes (referred to as germplasm set 2) was carefully selected to represent the world-wide-diversity of cultivated rice. The varieties in this second germplasm set included 29 genotypes from the *Indica* subspecies (isozyme group I), 17 from the *Japonica* subspecies (isozyme group VI), and 13 genotypes from less well-characterized ecotypes (isozyme groups II, III and V; Glaszmann 1987). Four cultivars were common to both germplasm sets 1 and 2 so that the repeatability of the genotyping procedure could be assessed. In addition, two varieties that were the source of the clones used to design the microsatellites markers were included as controls in the experiment: IR36 was the source of the library from which 23 of the microsatellites were developed, and Nipponbare was the source of sequence information for the other four markers, RM105, RM110, RM118 and RM135 (Akagi et al. 1996; Temnykh et al. 2000). Total genomic DNA was extracted as described in Cho et al. (2000) and Panaud et al. (1996) from single plants for germplasm set 1 and from bulks of 20 plants each for germplasm set 2, respectively.

Multiplexing of fluorescent-labeled microsatellites

Four panels (A, B, C and D) of multiplexed fluorescent-labeled microsatellites were created for this study (Table 1). The panels included 27 microsatellites that were selected from over 500 previously mapped markers on the basis of their published molecular weight, reliability of amplification signal, polymorphism information content, proximity to known disease resistance genes and distribution throughout the rice genome (Chen et al. 1997; Cho et al. 2000; Temnykh et al. 2000, 2001). The forward primers of the selected markers were labeled with either hexachloro-6-carboxyfluorescein (HEX), tetrachloro-6-carboxyfluorescein (TET) or 6-carboxyflourescein (6-FAM) dye phosporamidites and synthesized on an Applied Biosystem 392 (Applied Biosystem, Foster City, Calif) by the Cornell Bioresource Facility. The allele size-ranges of the microsatellites, i.e., the size difference between the largest and smallest alleles, were inferred from a previous germplasm survey (Cho et al. 2000; Temnykh et al. 2000; http://www.gramene.org//).

Individual PCR amplifications of each microsatellite were carried out in a total volume of 25 µl containing 1 µl of genomic DNA (50 ng) and 0.5 µl of each primer (diluted to a 10-mM concentration), 2.5 µl of 10 times buffer (0.1 M Tris pH 8.3; 0.5 M KCl; 7.5 mM $MgCl_2$; 0.1% gelatin), 100 mM each of dNTPs and 1.0 unit of *Taq* polymerase. The PCR profile used was a hot start of 94 °C for 5 min; followed by 35 cycles of denaturing at 94 °C for 1 min; annealing at 55 °C for 1 min; extension at 72 °C for 2 min; and product extension for 4 min at 72 °C. The PCR products for each set of microsatellites were mixed together in a ratio of 1:2:4 FAM:TET:HEX, that was varied depending on the amplification intensity for individual markers as determined on agarose gels. PCR products labeled with HEX were added in higher amounts and those labeled in FAM were added in lower amounts because of the different signal intensities of these fluoropohores. One microliter of the mixed microsatellite samples was combined with 3 µl of a loading buffer (98% formamide, 10 mM EDTA, blue dextran) and 0.5 µl of an internal-lane size standard, TAMRA-labeled Genescan-350, in preparation for co-electrophoresis. The samples were denatured at 90 °C for 2 min and run on 64-well, 36-cm plates with a 24-cm well-to-read distance, 5% denaturing Longranger polyacrylamide gels (8.0 M urea) in 1 × TBE buffer with the recommended run module (constant 30 W) and with filter set B. Gels were run for 6 h on a 373A automated DNA sequencer (Applied Biosystems).

Data analysis

Molecular weights for microsatellite products, in base-pairs, were estimated with Genescan software (Applied Biosystems) by the local Southern method. The individual fragments were assigned as alleles of the appropriate microsatellite loci with Genotyper software (Applied Biosystems). Allele binning was conducted with the following steps: (1) the fragments were sorted in descending order by size; (2) fragments that were separated by less than 0.4 bp were "binned" together, keeping the standard deviation of each bin below 0.2 bp; and (3) the mean was determined and rounded off to the nearest whole base-pair integer to give a molecular weight for the allele (Ghosh et al. 1997; Idury and Cardon 1997). Data obtained on silver-stained gels was used as a reference to avoid band-scoring errors and eliminate artifacts (Dovc et al. 1994). The polymorphism information content (PIC) for each microsatellite marker in panels C and D was calculated according to Anderson et al. (1993), based on the allele patterns of all the varieties analyzed. Genetic distances and dissimilarity matrices were calculated using the software program Microsat (Minch et al. 1997). Distances were based on the proportion of shared alleles assuming an "infinite alleles" model. The genotypes were clustered following the unweighted pair-group method using an arithmetic average (UPGMA) with the "neighbor" subprogram of the PHYLIP package (Felsenstein 1993).

Results

Allele scoring with fluorescent-labeled microsatellites

Four multiplex panels were optimized for the evaluation of the two germplasm sets. In the initial part of the study, multiplex panels A, consisting of 12 markers, and panel

Table 1 Fluorescently-labeled microsatellite markers used in four multiplex panels (A, B, C & D)

Panel[a]	Label	Marker	Genbank or clone no.	M.Wt. class	Chrom.	Reference
A	FAM	RM253	CT452	Low	6	Chen et al. 1997
A	FAM	RM135	D41754	Low	3	Temnykh et al. 2000
A	FAM	RM280	CT780	Med	4	Temnykh et al. 2000
A	FAM	RM8	GA317	High	2	Panaud et al. 1996
A	TET	RM170	D15716	Low	6	Temnykh et al. 2000
A	TET	RM284	CT794	Med	8	Temnykh et al. 2000
A	TET	RM118	D24674	Med	7	Temnykh et al. 2000
A	TET	RM55	GA587	High	3	Chen et al. 1997
A	HEX	RM286	CT806	Low	11	Temnykh et al. 2000
A	HEX	RM110	D22203	Med	2	Temnykh et al. 2000
A	HEX	RM231	CT234	Med	3	Chen et al. 1997
A	HEX	RM38	GA344	High	8	Chen et al. 1997
B	FAM	RM1	GA12	Low	1	Panaud et al. 1996
B	FAM	RM105	D15582	Low	9	Temnykh et al. 2000
B	FAM	RM18	GA97	Med	7	Panaud et al. 1996
B	FAM	RM122	RM122	High	5	Wu and Tanksley, 1993
B	TET	RM5	GA273	Low	1	Panaud et al. 1996
B	TET	RM17a	GA56	Med	12	Panaud et al. 1996
B	TET	RM222	CT193	High	10	Chen et al. 1997
B	HEX	RM13	GA478	Low	5	Panaud et al. 1996
B	HEX	RM337	CTT64	Med	8	Temnykh et al. 2000
C	FAM	RM1	GA12	Low	1	Panaud et al. 1996
C	FAM	RM70	ATT2	Med	7	Chen et al. 1997
C	FAM	RM122	n.a.	High	5	Wu and Tanksley, 1993
C	TET	RM5	GA273	Low	1	Panaud et al. 1996
C	TET	RM55	GA587	High	3	Chen et al. 1997
C	TET	RM17b	GA44	High	12	Panaud et al. 1996
C	HEX	RM248	CT469	Low	7	Chen et al. 1997
C	HEX	RM231	CT234	Med	3	Chen et al. 1997
C	HEX	RM38	GA344	High	8	Chen et al. 1997
D	FAM	RM253	CT452	Low	6	Chen et al. 1997
D	FAM	RM224	CT199	Med	11	Chen et al. 1997
D	FAM	RM252	CT206	High	4	Chen et al. 1997
D	TET	RM229	CT224	Low	11	Chen et al. 1997
D	TET	RM17a	GA56	Med	12	Panaud et al. 1996
D	TET	RM222	CT193	High	10	Chen et al. 1997
D	HEX	RM44	GA408	Low	8	Chen et al. 1997
D	HEX	RM13	GA478	Med	5	Panaud et al. 1996

[a] Panels A and B tested on 14 mapping parents, Panels C and D tested on 59 varieties

B, consisting of nine markers (Table 1), were designed for genotype germplasm set 1. The varieties in germplasm set 1 are considered to be standard genotypes for evaluating allelic diversity in rice and have been analyzed previously with all the available microsatellites by Cho et al. (2000) and Temnykh et al. (2000). Subsequently, we constructed panels C, with ten markers and panel D, with eight markers, to evaluate the larger number of varieties in germplasm set 2. The varieties in germplasm set 2 had been analyzed previously with inter-simple sequence repeat and amplified fragment length fingerprinting (Blair et al. 1999, accession numbers included) as well as with 150 RFLP probes (G. Second, personal communication; while subsets of the same cultivars were also analyzed by Glaszmann (1987) for isozymes and by Wang and Tanksley (1989) for other RFLP markers. This study is the first to use microsatellite markers to analyze the genetic diversity among the varieties in germplasm set 2. Ten markers (RM1, RM5, RM13, RM17a, RM38, RM55, RM122, RM222, RM231 and RM253) were used in more than one multiplex panel and were evaluated on both sets of germplasm (Table 1).

The four multiplex panels produced a total of 1,294 alleles from all the genotypes analyzed (159, 114, 563 and 458 alleles for panels A, B, C and D, respectively). More alleles were observed for panels C and D, because these panels were tested against a wider array of genotypes (59) than were panels A and B (13). A total of 2.5 PAGE gels were needed to evaluate all the genotype × marker combinations, resulting in an average data-collection efficiency of 518 alleles detected per gel.

Figure 1 shows the peak height and fluorescence intensity of the TAMRA-350 size standard and the microsatellites from panel C amplified for the reference variety IR36. The sizes of the alleles observed with automated fluorescent detection for IR36 were comparable to the sizes predicted from sequencing the clone used to isolate the microsatellites (Panaud et al. 1995, 1996; Chen et al. 1997; Temnykh et al. 2000) (data not shown). The size standard was effective for estimating molecular weights for all the PCR products amplified across all the multi-

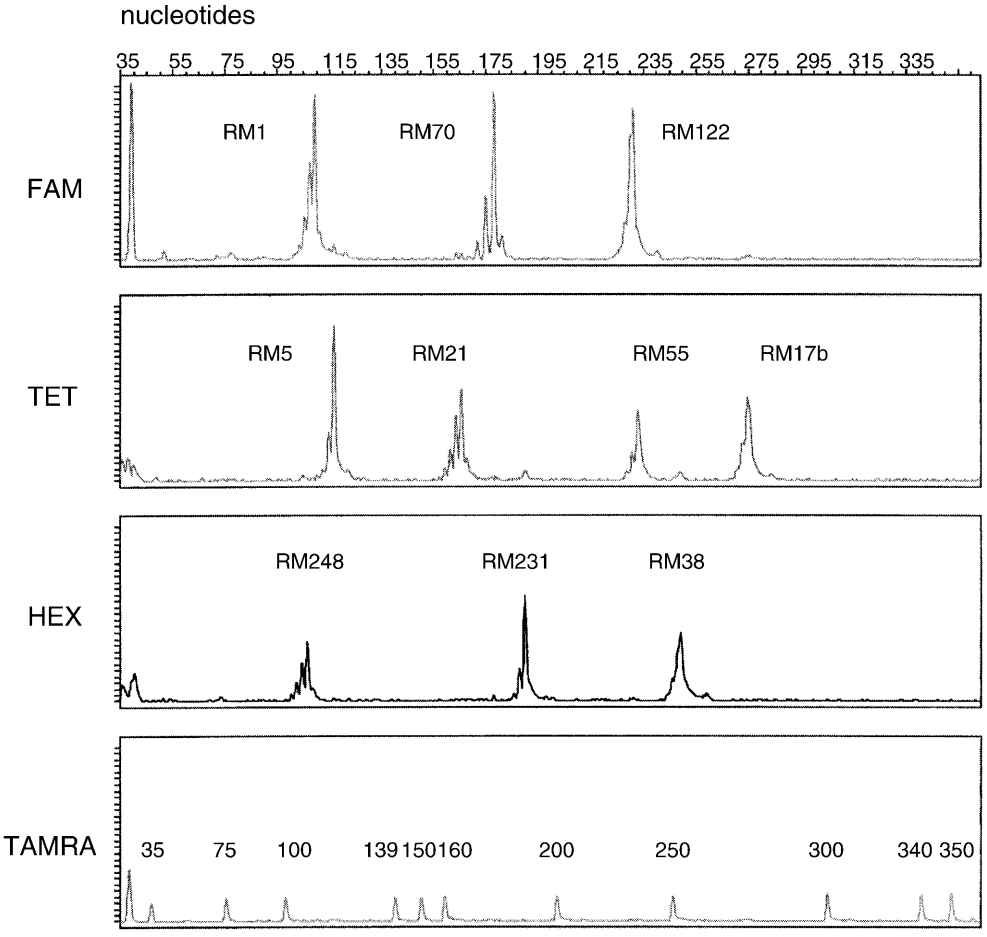

Fig. 1 Electropherograms showing the analysis of multiplexed FAM, TET and HEX labeled microsatellite PCR products from panel C for the standard variety IR36. Peaks represent fluorescence intensities (vertical axis) measured by an ABI373A DNA sequencer with size estimates in nucleotides (horizontal axis) determined by Genescan 672 software based on TAMRA-350 labeled size standards, shown below. Microsatellite alleles are identified above the peaks

plex panels, which ranged from 73 to 277-bp long. Each individual microsatellite had a typical fragment pattern with a characteristic number of stutter-peaks, signal peak width, slope and spread that could easily be recognized. Long stutters on silver-stained gels showed fewer stutter peaks on the ABI due in part to the fact that only one strand is labeled with fluorescence while both strands of the DNA molecule are stained with silver. Because only the most intense fragment was evaluated for molecular weight, the stuttering did not interfere with allele calling.

A variety was assigned a null allele for a microsatellite locus whenever an amplification product could not be detected for a particular genotype × marker combination in repeated experiments. In germplasm set 2, six out of the 17 loci evaluated (RM5, RM55, RM70, RM224, RM231 and RM253) showed null alleles for only a single variety, one locus (RM44) showed null alleles for two varieties, and another locus (RM38) showed null alleles for 11 varieties. Meanwhile, for germplasm set 1, null alleles occurred at four microsatellites (RM38, RM118, RM253 and RM280). The loci, RM38 and RM253, had null alleles among both sets of germplasm.

Multiple alleles per sample were observed at four loci (RM222, RM224, RM229 and RM252). Rice is a naturally inbred crop, so multiple alleles in a variety are generally an indication of mixed pure lines or seed mixtures (heterogeneity), rather than genetic heterozygosity, and the phenomenon is especially common in landrace varieties (Olufowote et al. 1997). Heterogeneity was observed for a total of four varieties in germplasm set 2 (SLO17, Carreon, JC157 and Azucena). All four showed multiple alleles at RM252 which is consistent with unpublished data (M. Semon, Cornell University, personal communication) suggesting that this SSR has an unusually high rate of mutation in rice. Because all the varieties were genotyped based on a bulk DNA sample, the multiple alleles detected in these data are likely to accurately reflect the degree of heterogeneity encountered in the IRGC seed stocks of these varieties. Multiple alleles were not observed for the varieties in germplasm set 1, which were genotyped based on DNA samples extracted from single plants.

Allele size range and performance of the multiplex panels

The size range between smallest and largest allele observed for a given microsatellite varied from 22 to 95 bp. A narrower average size-range was found for the 21 markers in panels A and B than for the 17 markers in panels C and D. This difference was probably due to the

Table 2 Number of alleles, polymorphism information content (PIC), number of varieties with null or heterozygous loci, allele size-range and mean alleles sizes found among 59 rice cultivars (29 isozyme group-I *Indica*, 17 isozyme group-VI *Japonica*) for 17 fluorescently labeled microsatellites from multiplex panels C and D (see text for explanation)

Marker	Allele no.			PIC			Varieties with		Allele size (bp)			Mean size (bp)	
	All	*Ind.*	*Jap.*	All	*Ind.*	*Jap.*	Null	Multiple	Min.	Max	Range	*Ind.*	*Jap.*
RM1	16	11	6	0.880	0.819	0.734	0	0	76	119	43	100.1	87.5
RM5	10	5	7	0.829	0.675	0.766	1	0	109	133	24	115.1	119.4
RM13	13	7	5	0.841	0.713	0.602	0	0	129	164	35	143.4	134.4
RM17a	11	7	3	0.817	0.785	0.491	0	0	163	190	27	177.5	163.5
RM17b	11	7	3	0.744	0.771	0.215	0	0	248	276	28	263.6	251.1
RM38	14	5	5	0.850	0.573	0.720	11	0	238	263	25	250.8	260.3
RM44	12	7	6	0.822	0.685	0.711	2	0	80	118	38	106.4	102.1
RM55	6	5	2	0.757	0.704	0.430	1	0	221	243	22	237.5	241.6
RM70	31	18	11	0.957	0.926	0.886	1	0	121	216	95	175.8	160.9
RM122	11	7	6	0.779	0.685	0.754	0	0	229	252	23	235.5	235.2
RM222	16	12	6	0.879	0.866	0.734	0	1	209	257	48	226.0	216.0
RM224	18	11	5	0.906	0.854	0.692	1	2	123	180	57	149.2	137.6
RM229	16	9	7	0.904	0.818	0.810	0	1	112	135	23	120.3	131.1
RM231	12	9	5	0.848	0.747	0.727	1	0	170	196	26	185.4	189.1
RM248	16	11	5	0.876	0.799	0.740	0	0	73	109	36	99.8	85.8
RM252	30	8	4	0.943	0.860	0.720	0	4	193	277	84	215.7	198.2
RM253	19	11	10	0.924	0.859	0.865	1	0	89	119	30	112.4	109.2
Average	15.3	8.7	5.7	0.853	0.768	0.682	n.a.	n.a.	n.a.	n.a.	n.a.	172.8	167.8

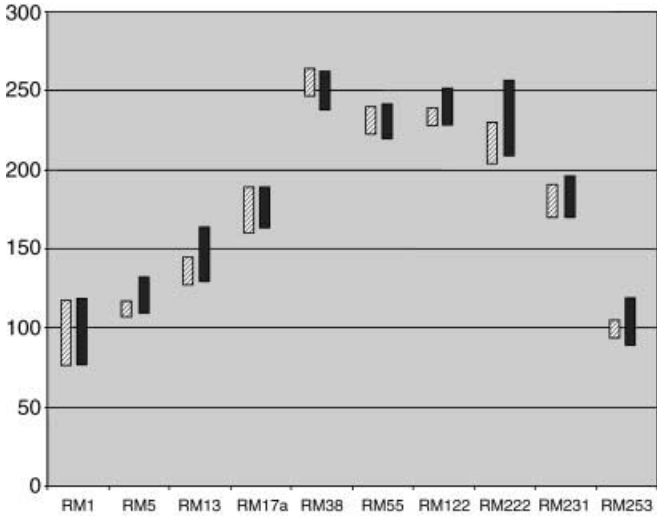

Fig. 2 Allele size-ranges in base pairs (vertical axis) for ten microsatellites (shown at the horizontal axis) evaluated on two sets of germplasm, consisting of 13 mapping parents (*light bars*) and 59 cultivars (*dark bars*) of rice

fact that fewer genotypes were evaluated for Panels A and B but it could also be due to the particular set of loci tested against either germplasm set. To differentiate between these possibilities we compared the allele size-ranges for ten microsatellite markers tested on both germplasm sets (shown diagramatically in Fig. 2). For these markers, we again found that a significantly wider allele size-range was observed in germplasm set 2 (30 bp ± 8.9) than in germplasm set 1 (20 bp ± 9.0) ($P = 0.002$, paired *t*-test). The allele range was from 0 to 22 bp wider in the larger germplasm set compared to the smaller standard germplasm set. At five out of the ten microsatellites common to both datasets, the total allele size-range was increased by more than 10 bp. Because we had used germplasm set 1 to predict the range of molecular weights for individual microsatellites, we were concerned that the wider allele size-range found in germplasm set 2 might cause problems of overlapping alleles from different microsatellites within a single dye color. However, the markers were not observed to overlap with each other, and allele-calling was straightforward and unequivocal in all the panels. This occurred because we had allowed for a 20-bp "buffer" between markers in the design of the multiplex panel, and alleles falling outside of the predicted size-range did not exceed the buffer zone. Furthermore, we had allowed for a 20-bp "buffer" between markers in the design of the multiplex panel. Therefore it seems that the multiplex panels can be confidently designed based on the results of testing microsatellites on germplasm set 1 or the equivalent standard panel described in Cho et al. (2000).

When alleles at a locus were ordered from smallest to largest, 54% of adjacent alleles differed by the number of base-pairs in the simple sequence repeat (data not shown). For the microsatellites based on di-nucleotide GA repeats, which included all but one of the markers used in these panels, the most-common alleles were separated by two base-pairs. Meanwhile, for the tri-nucleotide-containing locus, RM70, the size-ordered alleles were usually separated by three base-pairs. However, an average of 46% of the microsatellite alleles were separated in size from the next largest allele by a single base-pair, a phenomenon that was associated with the least-common alleles. While some of these alleles may represent scoring round-off errors, the frequency of these events suggests a biological or possibly a technical explanation. A similar situation was de-

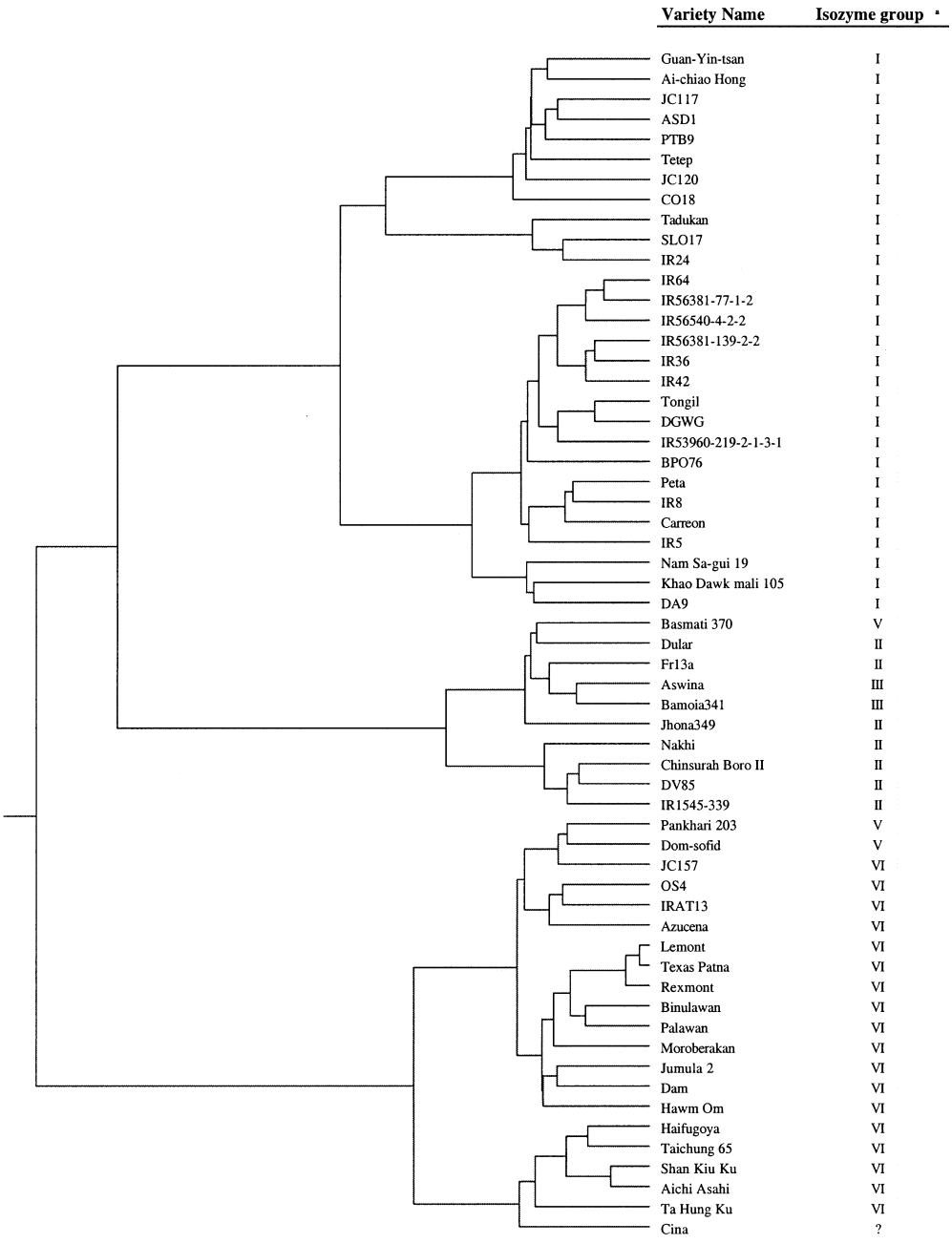

Fig. 3 Dendrogram showing the associations among 59 rice cultivars derived from a UPGMA cluster analysis using the Nei and Li genetic-distance coefficient based on alleles present at 17 microsatellite loci

scribed by Chen et al. (submitted) where a relatively high rate of insertion/deletion events in regions flanking SSR loci contributed to allelic differences that would not be expected based on the SSR motif itself. It is also consistent with the results presented by Coburn et al. (2002) showing that the standard deviation of allele molecular weight estimates on an ABI 373 can be significantly reduced by using T4 DNA polymerase to minimize the effect of plus-A modification during PCR.

Diversity analysis

The level of polymorphism among the 59 varieties of germplasm set 2 was evaluated by calculating allele number and PIC values for each of the 16 microsatellites evaluated (Table 2). The average number of alleles was 15.3 and ranged from six alleles for RM55 to 31 alleles for RM70, respectively. The PIC values, a reflection of allele diversity and frequency among the varieties, were very high for all the microsatellites, averaging 0.853, and ranging from a low of 0.744 for RM55 to a high of 0.957 for RM70. The PIC value was significantly correlated with the number of alleles ($r = 0.848$, $P < 0.01$) and the allele size range ($r = 0.682$, $P < 0.05$) for the microsatellites evaluated in this study. The allele size range and the number of alleles were themselves also highly correlated ($r = 0.906$, $P < 0.001$).

The highest frequency allele was present in up to 24 out of 59 (40.6%) of the varieties, as in the case of

RM122, but on average fewer than 15 (25%) genotypes shared a common allele. At all the microsatellite loci, there was at least one variety with an allele that was not shared by any other varieties in this study (i.e., a "rare" allele). The number of rare alleles per locus was significantly correlated with the allele size range (r = 0.75; $P < 0.01$). RM70, based on a poly (ATT) tri-nucleotide repeat, had the largest allele size range, the highest frequency of rare alleles, the highest number of alleles, and highest PIC value of any marker in this study. Four loci with moderate allele size-ranges harbored eight rare alleles each (RM1, RM222, RM 224 and RM248), while RM252, a poly-(GA) dinucleotide repeat, had the second to largest allele size range and was a close second to RM70 in terms of total allele number and PIC value, but had only five rare alleles.

A cluster analysis was used to group the varieties and to construct a dendrogram (Fig. 3). The classification of the germplasm was in good agreement with previous studies of the same varieties (Glaszmann 1987; Wang and Tanksley 1989; Blair et al. 1999). Two main branches could be discerned, corresponding to the *Indica* (isozyme group I) and *Japonica* (isozyme group VI) subspecies of cultivated rice. The IRRI varieties and cultivars that have contributed to their pedigrees, such as Carreon, DGWG, IR5, IR8, IR24, IR36, IR42, IR64, Peta, SLO17 and others, formed one cluster within the *Indica* subgroup; while landraces and breeding lines that are not extensively represented in the pedigrees of improved varieties, such as Ai Chiao Hong, Asd1, CO18, Guanyintsan, JC117, JC120 and Ptb9, formed another. Within the *Indica* branch, the most distinct cluster was a group of varieties, including Aswina, Bamoia 341, Basmati 370, Chinsurah Boro II, Dular, DV85, Fr13a, IR1545-339, Jhona 349 and Nakhi, which have previously been classified as members of isozyme groups II, III and V (Glaszmann 1987).

The microsatellite locus RM17 was evaluated twice by amplification with the same fluorescently labeled forward primer and either of two different reverse primers, one producing a large PCR product ranging from 248 to 276-bp long, the other producing a smaller PCR product, ranging from 163 to 190 bp. The pattern of diversity detected at the locus was the same for both markers except for a subset of 12 varieties, from the isozyme group II, III and V varieties mentioned above, which all had a 7–10-bp insertion in the larger PCR product that was not present in the smaller one. Otherwise, for the remainder of the varieties, the sizes of alleles estimated by the two markers were highly correlated (r = 0.995, $P < 0.001$).

Several significant differences were detected between the microsatellite alleles of the varieties clustering with the *Indica* and *Japonica* isozyme groups (Table 2). The *Indica* subset of varieties was more diverse, producing a significantly greater average number of alleles (8.7 compared to 5.7; P = 0.0005, paired t-tests) and a higher PIC value (0.772 compared to 0.691; P = 0.03) than the *Japonica* varieties. The larger number of *Indica* (29) than *Japonica* (17) varieties included in the germplasm-survey may partially explain these differences. However, significant differences were found between the mean molecular weight for the amplified fragments produced by the two subgroups (P = 0.012, paired t-tests); *Indica* alleles were 5–8-bp larger than *Japonica* alleles. The inclusion of genetically diverse *Indica* genotypes extended the allele size-range in the larger set of 59 varieties compared to the standard set of 13 parents. Although no single *Indica* variety consistently had the longest allele for all microsatellites, several of the Indian landraces, such as Co18, JC117 and Ptb9, had among the longest alleles and would be of interest to include in an expanded standard germplasm set when testing allele size-ranges prior to the multiplexing of microsatellites.

Discussion

Because rice is a self-pollinated diploid crop, microsatellite markers usually reveal single-copy, homozygous loci and allelic heterogeneity is rare in pure-line varieties. These facts simplified the work of multiplexing microsatellites for the analysis of genetic diversity of rice cultivars. The diverse range of allele sizes of available rice microsatellite markers also facilitated the design of the four panels of 8 to 12 fluorescent-labeled microsatellites. Multiplexing panels with even larger numbers of microsatellites, up to 24 per lane, have been achieved for the human genome (Reed et al. 1994). To increase the ability to multiplex rice microsatellites, it would be advisable to design primer pairs that produce larger PCR products (250–350 bp) since the majority of markers that are currently available average 100–250 bp in size (Panaud et al. 1996; Chen et al. 1997; Temnykh et al. 2000, 2001). The opportunity to re-design primers to produce alleles in a higher or lower molecular-weight category at the same locus offers flexibility in the design of multiplex panels. A larger size standard, such as the TAMRA-500, would be needed when adding a larger tier of markers. To increase the efficiency of genotyping, re-designed primers should be made with compatible melting temperatures and, where possible, similar amplification intensities to allow for co-amplification of microsatellites in a single PCR reaction.

The most-important factor for the design of multiplex panels was to ensure that the microsatellites that were labeled in the same color were non-overlapping in the ranges of their alleles and, as long as the ranges in allele sizes for the microsatellites are compatible, the identity of amplification products can be readily determined. The range of allele sizes found for each microsatellite marker could be related to the identity of the simple sequence repeat motif itself (Ghosh et al. 1997). As reported for mammals, the range of allele sizes tends to be wider for loci with tri- and tetra-nucleotide simple repeats than for loci with di-nucleotide repeats (Hall et al. 1996). Therefore, an advantage of our using di-nucleotide microsatellites in the rice multiplex panels was that intra-allelic base-pair differences were generally small, and a dense

set of markers could be created. Meanwhile the advantage of using tri- and tetra-nucleotide microsatellites in a multiplex panel would be that individual alleles are more-easily discriminated because they tend to be separated by a larger repeat value (Ghosh et al. 1997). However, the wide range of allele sizes for microsatellites with tri-nucleotide motifs would reduce the number of markers that can be used per multiplex panel (Hall et al. 1996). In this study of rice microsatellites, the single tri-nucleotide targeting marker had a wider allele range than any of the di-nucleotide-based markers and was, therefore, combined with other HEX labeled markers that were much smaller in size range so that they would not overlap.

The germplasm set studied in the present research was a good test of the multiplex microsatellite panels. The polymorphism information content (PIC) values were high even compared to previous estimates for rice microsatellites (Wu and Tanksley 1993; Panaud et al. 1996; Olufowote et al. 1997; Cho et al. 2000), demonstrating the diverse nature of the selected genotypes. Because of the wide cross-section of germplasm that these varieties represent, it is likely that the allele size-ranges reported here are close to the maximum for cultivated rice. On the other hand, wild relatives of Asian rice (e.g., *O. rufipogon* and *O. nivara*) as well as cultivated and wild forms of African rice (*O. glaberrima, O. barthii* and *O. longistaminata*) often have markedly smaller or larger alleles than the rice evaluated in this study, and this increase in the size range would limit the number of microsatellites that could be combined in a multiplex panel if wild accessions were included in the analysis (Harrington 2000). Wild rice species are also more-likely to present null alleles due to divergence of the primer annealing sequences (Panaud et al. 1996; Ishii and McCouch 2000; Temnykh et al. 2000; Chen et al. submitted). Null alleles were relatively rare for the cultivated varieties analyzed in this study.

The double-evaluation of locus RM17 with two markers that amplified different size products uncovered a larger than expected allele for the longer of the two PCR products and led to an interesting hypothesis: namely, that an insertion-deletion or a second microsatellite motif is probably located within some alleles of this locus. The smaller product is known to contain a perfect GA repeat that is 26-units long in IR36 (Panaud et al. 1996). The revelation that an additional allele variant existed for the locus was interesting from a breeding perspective, since new alleles at any locus are useful for the selection of traits that map nearby. On a biological level, the inclusion of larger portions of flanking sequences in the PCR product increases the chances of detecting sequence differences outside the microsatellite itself and can result in the discovery of new levels of diversity. In this case, the new allele observed for the RM17 locus was diagnostic for a specific phylogenetic divergence, occurring in 12 varieties which were all traditional landraces and improved lines from the Indian subcontinent that are members of isozyme groups II, III and V (Glaszmann 1987). This observation supports the conclusions of Chen et al. (submitted) that some of the variation observed as size differences in PCR-analyzed products is the result of mutation events other than SSR expansion/contraction in regions immediately flanking the repeat motif. It is also consistent with our observation that 46% of the alleles in this study differ from each other by a single base-pair.

The alleles of the *Japonica* varieties were significantly smaller on average than the alleles of the *Indica* varieties in this study. This difference may have resulted from an ascertainment bias that occurred during the development of the microsatellites based on screening of a small insert. In the *Indica* library clones with long repeat motifs are preferentially selected. A similar ascertainment bias has been observed for microsatellites from two species of *Drosophila* (Hutter et al. 1998). One implication of these results is that library based screening approaches to microsatellite development can be used to identify long simple-sequence repeats specific to each subspecies of rice that will tend to be more polymorphic than shorter arrays (Cho et al. 2000). There was no bias in SSR length when microsatellites were derived from randomly sequenced cDNAs and BAC ends of the *Japonica* variety Nipponbare (Akagi et al. 1996; Cho et al. 2000; Temnykh et al. 2000, 2001). It would be interesting to see whether library based screening of a *Japonica* variety could be used to develop more microsatellite markers, with enhanced rates of polymorphism, specifically for this economically important rice subspecies. Finally our, results show that di-nucleotide microsatellites are informative for clustering genotypes within *O. sativa* and its subspecies and therefore are likely to have a controlled mutation rate.

In this study, automated fluorescence detection was used to improve the multiplex ratio of microsatellite analysis to a level comparable to that of multi-locus fingerprinting techniques. While design of multiplex panels is a highly empirical exercise, information such as that presented here lays the foundation for the development of genome-wide panels of microsatellites in rice. An advantage of multiplexing is that new panels can be designed to target specific regions of the genome, or genes of interest, using markers of known map position. These microsatellites panels can be expected to greatly increase the efficiency of genetic-diversity assessment, variety fingerprinting and identification, the genetic and physical mapping of genes and quantitative trait loci and marker-assisted selection during breeding.

Acknowledgements Our thanks to C. Bayles for technical assistance, G. Second and Y.G. Cho for providing DNA samples, S. Kresovich, S. Temnykh and J. Coburn for critical reviews and to C. Morehouse and L. Swales for help with formatting. This work was supported by a Plant Science Center (USDA-DOE-NSF-Cornell Biotechnology Program) fellowship to M.W.B. and by a grant from the Rockefeller Foundation (RF99001#726) to S. McCouch.

References

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 93:1071–1077

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

Blair MW, Panaud O, McCouch SR (1999) Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.) Theor Appl Genet 98:780–792

Bredemeijer GMM, Arens P, Wouters D, Visser D, Vosman B (1998) The use of semi-automated fluorescent microsatellite analysis for tomato cultivars. Theor Appl Genet 97:584–590

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 95:553–567

Chen X, Cho YG, McCouch SR (2001) Sequence divergence of rice microsatellites in *Oryza* and other plant species. Mol Genet Gen (submitted)

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

Coburn J, Temnykh S, McCouch SR (2002) Chromosome-specific panels of microsatellite markers for semi-automated genotyping of rice (*Oryza sativa* L.). Crop Sci (in press)

Diwan N, Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor Appl Genet 95:723–233

Dovc P, Brikenmaier S, Forester M (1994) Nonradioactive detection of microsatellite polymorphisms. Biotechniques 8:183–186

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) v. http://www.evolution.genetics.washington.edu/pub/phylip

Ghosh S, Karanjawala ZE, Hauser ER, Ally D, Knapp JI, Rayman JB, Musick A, Tannenbaum J, Te C, Shapiro S, Eldridge W, Musick T, Martin C, Smith JR, Carpten JD, Brownstein MJ, Powell JI, Whiten R, Chines P, Nylund SJ, Magnuson VL, Boehnke M, Collins FS (1997) Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. Genome Res 7:165–178

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 42:233–240

Hall JM, LeDuc CA, Watson AR, Roter AH (1996) An approach to high-throughput genotyping. Genome Res 6:781–790

Harrington SE (2000) A survey of genetic diversity of eight AA genome species of *Oryza* using microsatellite markers. MS thesis, Cornell University

Hutter CM, Schug MD, Aquadro CF (1998) Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. Mol Biol Evol 15:1620–1636

Idury RM, Cardon LR (1997) A simple method for automated allele binning in microsatellite markers. Genome Res 7:1104–1109

Ishii T, McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. Theor Appl Genet 100:1257–1266

Ishii T, Xu Y, McCouch SR (2001) Nuclear and cholorplast microsatellite variation in A-genome species of rice. Genome 44:648–666

Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor Appl Genet 91:206–211

Mansfield DC, Brown AF, Green DK, Carothers AD, Morris SW, Evans H-J (1994) Automation of genetic linkage analysis using fluorescent microsatellite markers. Genomics 24:225–233

Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL (1997) Microsat (v.1.5.): a computer program for calculating various statistics on microsatellite allele data. http://www.lotka.stanford.edu/microsat.html

Mitchell SE, Kresovich S, Jester CA, Hernandez CJ, Szewc-McFadden AK (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. Crop Sci 37:617–624

Moncada MP, Martínez CP, Tohme J, Guimaraes E, Chatel M, Borrero J, Gauch H, Jr. and McCouch SR (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* ×*Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor Appl Genet 102:41–52

Olufowote JO, Xu Y, Chen X, Goto M, McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40:370–378

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 microsatellites and characterization of simple sequence repeat polymorphism in rice (*Oryza sativa* L.). Mol Gen Genet 252:597–607

Ponce MR, Robles P, Micol JL (1999) High-throughput genetic mapping in *Arabidopsis thaliana*. Mol Gen Genet 261:408–415

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Reed PW, Davies JL, Copeman JB, Bennett 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

Schwengel DA, Jedlicka AE, Nanthakumar EJ, Weber JL, Levitt RC (1994) Comparison of fluorescence-based semi-automated genotyping of multiple microsatellite loci with autoradiographic techniques. Genomics 22:46–54

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. Theor Appl Genet 95:163–173

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Wang ZY, Tanksley SD (1989) Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32:1113–1118

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

Xiao J, Li J, Yuan L, McCouch SR, Tanksley SD (1996) Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor Appl Genet 92:637–643

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

Ziegle JS, Su Y, Corcoran KP, Nie L, Maynard PE, Hoff LB, McBride LJ, Kronick MN, Diehl, SR (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14:1026–1031