Theor Appl Genet (1997) 95 : 553–567 © Springer-Verlag 1997

X. Chen · S. Temnykh · Y. Xu
Y. G. Cho · S. R. McCouch

# Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.)

Received: 13 February 1997 / Accepted: 28 February 1997

**Abstract** Ninety-four newly developed microsatellite markers were integrated into existing RFLP framework maps of four rice populations, including two doubled haploid, a recombinant inbred, and an interspecific backcross population. These simple sequence repeats (SSR) were predominantly poly(GA) motifs, targetted because of their abundance in rice. They were isolated from a previously described sheared library and a newly constructed enzyme-digested library. Differences in the average length of poly(GA) tracts were observed for clones isolated from the two libraries. The length of GA motifs averaged 21 repeat units for clones isolated from the *Tsp*-509-digested library, while motifs averaged 17 units for clones from the sheared library. There was no evidence of clustering of microsatellite markers near centromeres or telomeres. Mapping of the 94 newly developed markers as well as of 27 previously reported microsatellites provided genome-wide coverage of the 12 chromosomes, with an average distance of 1 SSLP (simple sequence repeat polymorphism) per 16–20 cM.

**Key words** Microsatellite · Simple sequence repeats (SSR) · Simple sequence length polymorphism (SSLP) · Linkage map · Molecular map · Rice (*Oryza sativa*)

Communicated by F. Salamini

X. Chen · S. Temnykh · Y. Xu · S. R. McCouch (✉)
Department of Plant Breeding and Biometry, 252 Emerson Hall, Cornell University, Ithaca, NY 14853-1901, USA

Y. G. Cho
National Institute of Agricultural Science and Technology, Suweon, Korea

## Introduction

Microsatellite markers based on simple sequence repeats (SSR) have been developed in many crop species, including rice (Wu and Tanksley 1993; Panaud et al. 1996; Akagi et al. 1996), barley (Becker and Heun 1995; Liu et al. 1996), grapevine (Thomas and Scott 1993), *Brassica* (Kresovich et al. 1995), maize (Senior and Heun 1993; http://WWW.agron.missouri.edu/), soybean (Akkaya et al. 1995), and tomato (Broun and Tanksley 1996). These markers detect simple sequence length polymorphism (SSLP) and are rapidly displacing restriction fragment length polymorphisms (RFLPs) for many kinds of genetic studies, largely because of their technical simplicity, the small amount of starting DNA required, the relatively low cost for the user, rapid turn-around time, and high power of genetic resolution. Plant SSLPs have already been demonstrated to be a powerful tool in genotype identification and variety protection (Smith and Helentjaris 1996; Rongwen et al. 1995; Olufowote et al. 1997), seed purity evaluation and germplasm conservation (Brown and Kresovich 1996; Hahn and Grifo 1996; Bretting and Widrlechner 1995; Powell et al. 1996; Olufowote et al. 1997), diversity studies (Xiao et al. 1996a; Yang et al. 1994), gene and quantitative trait locus (QTL) analysis (Blair and McCouch 1997, Koh et al. 1996, Xiao et al. 1996b), pedigree analysis and marker-assisted breeding (Ayres et al. 1997; Yang et al. 1994), and screening of large insert libraries prior to cloning (M. Blair, Cornell University, unpublished data).

Despite their apparent advantages, relatively few microsatellite markers are publicly available for plants. In contrast, high-density SSLP maps containing 5,264 and 7,377 microsatellite loci have been published for the human (Dib et al. 1996) and mouse genomes (Dietrich et al. 1996). Prior to this report, 58 microsatellite markers had been mapped in rice. Thirty-two of these were developed based on the identification of

SSR-containing sequences from GenBank (Wu and Tanksley 1993) and the DDBJ database (Akagi et al. 1996), and 26 were isolated based on the screening of genomic libraries (Panaud et al. 1996; Wu and Tanksley 1993). Although the growing pool of published DNA sequences provides a substantial reservoir for microsatellite marker development, a recent search of 11,798 rice sequences in the DDBJ database identified 369 complete SSRs (Akagi et al. 1996), while the reported estimate of SSRs in the rice genome is between 5,700 and 10,000 (McCouch et al. 1997). We therefore suggest that the construction of a high-density microsatellite map for this species will be most efficient based on targeting multiple repeat motifs and the complementary use of database searches and the screening of different types of libraries, including physically sheared and enzyme-digested libraries, genomic and cDNA libraries (McCouch et al. 1997), and microsatellite-enriched libraries (Prochzka 1996; Edwards et al. 1996).

The construction of a saturated map for rice based solely on these polymerase chain reaction (PCR)-based markers is a reasonable goal for the near future. The aim of the project reported here was (1) to develop microsatellite markers in rice using two small insert libraries (a *Tsp*509-digested library and a sheared library); (2) to compare the efficiency of marker development based on these two sources of clones; (3) to construct a microsatellite map providing genome-wide coverage of rice.

## Materials and methods

### Library construction

Genomic DNA from cv 'IR36' was used to construct both an enzyme-digested and a sheared library. The optimized cleavage size was obtained by partially digesting DNA with *Tsp*509 I (New England Biolabs, Beverly, Mass.), isolating restriction fragments ranging from 300 to 800 bp from a 2% agarose gel, and purifying them using Glas Pac/GSTM (Supply Co, San Rafael, Calif.). The digested DNA was ligated into the Lambda ZapII/*Eco*RI/CIAP cloning vector and packaged into phage coats according to the protocols provided by Stratagene (La Jolla, Calif.). The sheared library was previously described in Panaud et al. (1996). In brief, genomic DNA from cv 'IR36' was physically sheared and then treated with DNAse I. Fragments averaging 300–500 bp were ligated to synthetic *Eco*RI linkers and cloned into Lambda-Zap II/*Eco*RI vectors (Stratagene, La Jolla, Calif.).

### Isolation of clones containing poly(GA)

Twenty-three plates each containing approximately 10,000 plaques were screened for the presence of poly(GA) sequences by plaque hybridization from the enzyme-digested library. The method was that described by Panaud et al. (1996), but with the following modifications: (1) two rounds of hybridization and purification were used instead of three; (2) the stringency of the pre-sequencing screen was increased to minimize the number of clones to be mini-prepped and sequenced. In this study, lysates derived from single plaques were used as templates for PCR amplification with three combinations of primer sets to estimate insert size and confirm the presence of poly(GA) motifs: (1) T3 (5′-AATTAACCCTCACTAAAGGG-3′) plus T7 (5′-TAATACGACTCACTATAGGG-3′); (2) T3 plus internal (5′-CCCGGATCC(GA)9-3′); (3) T7 plus internal (as in 2). The PCR was performed in 25-μl reactions containing 2.0 μl of lysate, 0.2 μ*M* of each primer, 200 μ*M* deoxyribonucleotides, 50 m*M* KCl, 10 m*M* TRIS-Cl (pH 8.3), 1.5 m*M* $MgCl_2$, 0.01% gelatin and 0.5 unit of *Taq* polymerase. The PCR profile was: 94°C–1.5 min, 50°C–1.5 min, 72°C–2 min, for 25 cycles; followed by a final incubation at 72°C–5 min using a Perkin Elmer DNA Thermal Cycler 480. If no inserts, or inserts that were outside the designated 300 to 800 bp size range, were detected, lysates were discarded. Inserts in the expected size range (approximately 80%) were excised using the Exassist/SOLR system (Stratagene Co), and plasmid DNA was prepared according to the Merlin Miniprep protocol (Gopher Molecular Biology News Network). Minipreps were sequenced by the Cornell Sequencing Facility using an Applied Biosystem 373A machine (Perkin Elmer, Applied Biosystem division, Foster, Calif.).

### Primer design and evaluation

PCR primers flanking microsatellite repeat sequences were selected using the PRIMER 0.5 program (S. Lincoln, M. Daly, and E. Lander, Cambridge, Mass.) combined with manual design. The aim of the primer selection procedure was to produce well-matched primers, 17–22 nucleotides long, devoid of consecutive tracts of a single nucleotide, with a GC content around 50% (*Tm* approximately 60°C), and preferably G-or C-rich at the 3' end. Primers that met these requirements were preferentially selected to produce a PCR product in the range of 100–250 bp. Primers were synthesized by Research Genetics (Huntsville, Ala.).

### Plant material

Four mapping populations from the International Rice Research Institute (Los Baños, Philippines), the Chinese Rice Genome Project (Institute of Genetics, Chinese Academy of Sciences, Beijing, China), the Korean Rice Genome Project (National Agricultural Science and Technology Institute, Suweon, Korea), and Cornell University (Ithaca, NY) were used as the basis for placing microsatellite markers onto rice chromosomes. The first population consisted of 135 doubled haploid (DH) lines derived from a cross between 'IR64' (*indica*) and 'Azucena' (tropical *japonica*) (Guiderdoni et al. 1992; Huang et al. 1994) (hereafter referred to as the $DH_1$ population). The second population consisted of 133 DH lines derived from a cross between 'Zhai-Ye-Qing 8' (*indica*) and Jing Xi 17 (*japonica*) (hereafter referred to as the $DH_2$ population) (Zhu et al. 1993). The third population consisted of 164 $F_{11}$ recombinant inbred lines derived from a cross between 'Milyang 23' (*tongil* variety) and 'Gihobyeo' (temperate *japonica*) (Cho et al., Natl. Inst. Agr. Sci. & Techn., pers. comm.) (hereafter referred to as the RI population). The fourth population consisted of 113 individuals derived from an interspecific backcross between *O. sativa* (cv 'BS 125') and *O. longistaminata* (acc. WL02) (hereafter refer to as the SL population) (Causse et al. 1994).

### PCR amplification and evaluation of polymorphism

Newly synthesized primer pairs were tested for PCR amplification and polymorphism using DNA from the parents of all four mapping populations. PCR conditions were as described in Panaud et al. (1996). Briefly, 50-μl reactions contained 0.2 μM of each primer, 200 μ*M* deoxyribonucleotides, 50 m*M* KCl, 10 m*M* TRIS-Cl (pH 8.3), 1.5 m*M* $MgCl_2$, 0.01% gelatin, 50–100 ng of DNA and 1 unit of *Taq* polymerase. The PCR profile was: 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 2 min, and finally by 5 min at 72°C for the final extension. PCR

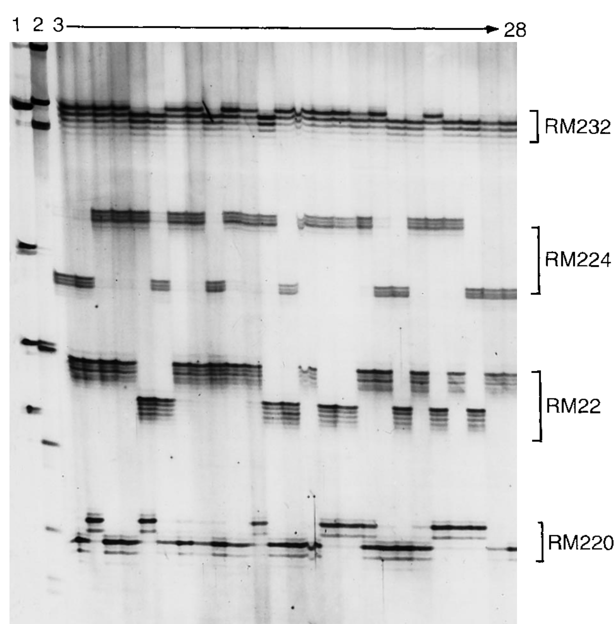



**Fig. 1** Multiplex loading of polyacrylamide gel showing 4 different microsatellite markers (*RM* designations to *right* of photo) segregating in the $DH_1$ population (*lanes 3–28*) and detected by silver staining. *Lanes 1* and *2* contain size markers, marker V and VIII (Boehringer Mannheim, Ind.), loaded with RM224. Size standard indicates molecular weight of RM224 alleles only

products were run on 4% polyacrylamide denaturing gels, and marker bands were revealed using the silver staining as described by Panaud et al (1996).

### Mapping of SSLPs

Primer pairs showing polymorphism on parental surveys were mapped onto the respective mapping populations. For the $DH_1$ population, markers were placed using a randomly selected subset of 96 individuals from the original mapping population and the RFLP dataset described in Huang et al. (1997). For the $DH_2$, RI, and SL populations, markers were placed using the complete populations. The PCR profile was the same as described above, and reactions were performed in 96-well plates in PTC100 96U thermocyclers (MJ Research, Watertown, Mass.). Individual PCR products were separated on 4% denaturing polyacrylamide gels and detected via silver staining. Multiple loading of markers was performed to increase the efficiency of the mapping process. Three or four amplification products from one or more primer pairs were loaded sequentially in a gel lane at 10- to 20-min intervals (Fig. 1). Segregation was scored, and markers were integrated into the existing RFLP framework maps for each population using the Kosambi mapping function in 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 greater than 2.0 were integrated into the framework maps, and probable locations of markers placing with a LOD less than 2.0 were also determined.

## Results

### Library screening and polymorphism survey

Primary screening of the *Tsp*509-digested library with a [$^{32}$P]-labeled poly$(GA)_{15}$ oligonucleotide probe identified 587 putative positive clones. Following a second round of plaque hybridization about 15% of these were eliminated. The remaining positive plaques were isolated and subjected to the pre-sequencing screening procedure as described above. This procedure eliminated 60% of the clones because they either failed to amplify or contained very large inserts. Finally, 139 of the original 587 (26%) clones were miniprepped and sequenced. Out of the 139 clones sequenced, 122 (88%) were found to contain poly(GA) motifs. When the $(GA)_n$ containing sequences were analyzed for redundancy, 25 of the 122 (20%) were found to be identical to one another and were eliminated from further consideration. Primer pairs flanking the microsatellite motifs could be designed for 89 of the remaining 97 (92%) of these clones, and 65 of these (73%) produced discrete PCR products of the expected size while the rest failed to amplify. Of these 65 microsatellite-containing clones 61 were polymorphic on pairs of mapping parents and could be mapped. Of these 61 markers, 90% were polymorphic on the 'IR64'/'Azucena' ($DH_1$) parents; 95% were polymorphic on the 'Zhai-Ye-Qing 8'/'Jing Xi 17' ($DH_2$) parents; 83% on the 'Milyang 23'/'Gihobyeo' (RI) parents; and 92% on the BS125/WL02 (SL) parents. The other 4 primer pairs were monomorphic on all four pairs of mapping parents.

Thirty clones were isolated and mapped from the sheared library in this study, including 23 poly(GA)-, 5 poly(TCT)-, 1 poly(ATT)-, and 1 poly(AATT)-containing clones. Two of the primer pairs bracketing poly(GA) microsatellites from the sheared library produced monomorphic banding patterns on the parents of all four populations.

### Summary of marker information

Table 1 summarizes information about the 94 microsatellite markers developed in this study, including clone name, locus designation, chromosomal location, primer sequence, size of PCR product in the reference variety ('IR36'), number of perfect repeat units, and a description of the repeat motif. These 94 SSLP loci include 86 poly(GA), 6 poly(TCT), 1 poly(ATT), and 1 poly(AATT) motifs. All marker loci are designated "RM" for *rice microsatellite*, and clones from the sheared library are in the series, RM1-RM100; clones from the *Tsp*509-digested library are designated RM200-RM299. The intensity of PCR amplification was estimated for each primer pair based on the relative signal intensity observed for the eight mapping parents on silver-stained gels. As summarized in Table 1, markers with asterisks (*) gave particularly strong amplification and are recommended for simultaneous multiplexing on silver-stained gels. Markers with ($\chi$) after the locus designation indicate faint amplification and require longer exposure times. All markers

**Table 1** Microsatellite marker information

| | Locus | Clone | Chromosome | Primer sequence[a] | Size (bp) in IR36 | Number of perfect repeats | Description of complex microsatellite motifs |
|---|---|---|---|---|---|---|---|
| 1 | RM23[b] | GA594 | 1 | CATTGGAGTGGAGGCTGG<br>GTCAGGCTTCTGCCATTCTC | 145 | 15 | |
| 2 | RM24 | GA5 | 1 | GAAGTGTGATCACTGTAACC<br>TACAGTGGACGGCGAAGTCG | 192 | 29 | |
| 3 | RM25 | GA30 | 8 | GGAAAGAATGATCTTTTCATGG<br>CTACCATCAAAACCAATGTTC | 146 | 18 | |
| 4 | RM26* | GA41 | 5 | GAGTCGACGAGCGGCAGA<br>CTGCGAGCGACGGTAACA | 112 | 15 | |
| 5 | RM27 | GA43 | 2 | TTTTCCTTCTCACCCACTTCA<br>TCTTTGACAAGAGGAAAGAGGC | 158 | 7 | $(CT)_7C(CT)_6$ |
| 6 | RM29 | GA120 | 2 | CAGGGACCCACCTGTCATAC<br>AACGTTGGTCATATCGGTGG | 250 | 7 | $(GA)_7$-18bp-$(GA)_5(AG)_4$ |
| 7 | RM30 | GA216 | 6 | GGTTAGGCATCGTCACGG<br>TCACCTCACCACACGACACG | 105 | 12 | $(AG)_9A(AGA)_{12}$ |
| 8 | RM31 | GA257 | 5 | GATCACGATCCACTGGAGCT<br>AAGTCCATTACTCTCCTCCC | 140 | 15 | |
| 9 | RM34 | GA281 | 1 | GAAATGGCAATGTGTGCG<br>GCCGGAGAACCCTAGCTC | 161 | 17 | $(CT)_{17}(TC)_2$ |
| 10 | RM35 | GA302 | 1 | TGGTTAATCGATCGGTCGCC<br>CGACGGCAGATATACACGG | 207 | 19 | |
| 11 | RM36[c] | GA306 | 3 | CAACTATGCACCATTGTCGC<br>GTACTCCACAAGACCGTACC | 192 | 23 | |
| 12 | RM38 | GA344 | 8 | ACGAGCTCTCGATCAGCCTA<br>TCGGTCTCCATGTCCCAC | 250 | 16 | |
| 13 | RM39[χ] | GA346 | 5 | GCCTCTCTCGTCTCCTTCCT<br>AATTCAAACTGCGGTGGC | 115 | 17 | $(CT)_{17}CCA(TC)_3$ |
| 14 | RM41[χ] | GA361 | 9 | AAGTCTAGTTTGCCTCCC<br>AATTTCTACGTCGTCGGGC | 265 | 15 | |
| 15 | RM42 | GA376 | 8 | ATCCTACCGCTGACCATGAG<br>TTTGGTCTACGTGGCGTACA | 166 | 6 | $(AG)_6$-11bp-$(AG)_2T(GA)_5$<br>$(AG)_5(GA)(AG)_2T(GA)A(AG)_5$ |
| 16 | RM44 | GA408 | 8 | ACGGGCAATCCGAACAACC<br>TCGGGAAAACCTACCCTACC | 99 | 16 | |
| 17 | RM47 | GA476 | 7 | ACTCCACTCCACTCCCCAC<br>GTCAGCAGGTCGGACGTC | 229 | 11 | $(AG)_7$-27bp-$(AG)_{11}$ |
| 18 | RM48 | GA479 | 2 | TGTCCCACTGCTTTCAAGC<br>CGAGAATGAGGGACAAATAACC | 204 | 17 | |
| 19 | RM49 | GA505 | 3 | TTCGGAAGTTGGTTACTGATCA<br>TTGGAGCGGATTCGGAGG | 189 | 27 | |
| 20 | RM50 | GA534 | 6 | ACTGTACCGGTCGAAGACG<br>AAATTCCACGTCAGCCTCC | 201 | 15 | $(CTAT)_4(CT)_{15}$ |
| 21 | RM51 | GA543 | 7 | TCTCGATTCAATGTCCTCGG<br>CTACGTCATCATCGTCTTCCC | 142 | 13 | |
| 22 | RM53 | GA562 | 2 | ACGTCTCGACGCATCAATGG<br>CACAAGAACTTCCTCGGTAC | 182 | 14 | |
| 23 | RM55* | GA587 | 3 | CCGTCGCCGTAGTAGAGAAG<br>TCCCGGTTATTTTAAGGCG | 226 | 17 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 24 | RM60* | AATT51 | 3 | AGTCCCATGTTCCACTTCCG<br>ATGGCTACTGCCTGTACTAC | 165 | 5 | $(AATT)_5$AATCT(AATT) |
| 25 | RM70 | ATT2 | 7 | GTGGACTTCATTTCAACTCG<br>GATGTATAAGATAGTCCC | 170 | 33 | $(ATT)_{33}$ |
| 26 | RM80 | TCT115 | 8 | TTGAAGGCGCTGAAGGAG<br>CATCAACCTCGTCTTCACCG | 142 | 25 | $(TCT)_{25}$ |
| 27 | RM81A | TCT118 | 1 | GAGTGCTTGTGCAAGATCCA<br>CTTCTTCACTCATGCAGTTC | 110 | 10 | $(TCT)_{10}$ |
| 28 | RM81B | TCT118 | 3 | GAGTGCTTGTGCAAGATCCA<br>CTTCTTCACTCATGCAGTTC | 110 | 10 | $(TCT)_{10}$ |
| 29 | RM82 | TCT122 | 7 | TGCTTCTTGTCAATTCGCC<br>CGACTCGTGGAGGTACGG | 186 | 11 | $(TCT)_{11}$ |
| 30 | RM83ᵡ | TCT123 | 12 | ACTCGATGACAAGTTGAGG<br>CACCTAGACACGATCGAG | 142 | 8 | $(TCA)_6(TCT)_8$ |
| 31 | RM84 | TCT125 | 1 | TAAGGGTCCATCCACAAGATG<br>TTGCAAATGCAGCTAGAGTAC | 113 | 10 | $(TCT)_{10}$ |
| 32 | RM201* | CT6 | 9 | CTCGTTTATTACCTACAGTACC<br>CTACCTCCTTTCTAGACCGATA | 158 | 17 | |
| 33 | RM202* | CT14 | 11 | CAGATTGGAGATGAAGTCCTCC<br>CCAGCAAGCATGTCAATGTA | 189 | 30 | |
| 34 | RM204 | CT19 | 6 | GTGACTGACTTGGTCATAGGG<br>GCTAGCCATGCTCTCGTACC | 169 | 44 | |
| 35 | RM205 | CT22 | 9 | CTGGTTCTGTATGGGAGCAG<br>CTGGCCCTTCACGTTTCAGTG | 122 | 25 | |
| 36 | RM206* | CT25 | 11 | CCCATGCGTTTAACTATTCT<br>CGTTCCATCGATCCGTATGG | 147 | 21 | |
| 37 | RM207* | CT41 | 2 | CCATTCGTGAGAAGATCTGA<br>CACCTCATCCTCGTAACGCC | 118 | 25 | |
| 38 | RM208* | CT43 | 2 | TCTGCAAGCCTTGTCTGATG<br>TAAGTCGATCATTGTGTGGACC | 173 | 17 | |
| 39 | RM209* | CT44 | 11 | ATATGAGTTGCTGTCGTGCG<br>CAACTTGCATCCTCCCCTCC | 134 | 18 | |
| 40 | RM210* | CT56 | 8 | TCACATTCGGTGGCATTG<br>CGAGGATGGTTGTTCACTTG | 140 | 23 | |
| 41 | RM211* | CT62 | 2 | CCGATCTCATCAACCAACTG<br>CTTCACGAGGATCTCAAAGG | 161 | 18 | $(TC)_3A(TC)_{18}$ |
| 42 | RM212* | CT67 | 1 | CCACTTTCAGCTACTACCAG<br>CACCCATTTGTCTCTCATTATG | 136 | 24 | |
| 43 | RM213* | CT87 | 2 | ATCTGTTTGCAGGGGACAAG<br>AGGTCTAGACGATGTCGTGA | 139 | 17 | |
| 44 | RM214 | CT91 | 7 | CTGATGATAGAAACCTCTTCTC<br>AAGAACAGCTGACTTCACAA | 112 | 14 | |
| 45 | RM215* | CT100 | 9 | CAAAATGGAGCAGCAAGAGC<br>TGAGCACCTCCTTCTCTGTAG | 148 | 16 | |
| 46 | RM216 | CT106 | 10 | GCATGGCCGATGGTAAAG<br>TGTATAAAACCACACGGCCA | 146 | 18 | |
| 47 | RM217 | CT115 | 6 | ATCGCAGCAATGCCTCGT<br>GGGTGTGAACAAAGACAC | 133 | 20 | |
| 48 | RM218 | CT125 | 3 | TGGTCAAACCAAGGTCCTTC<br>GACATACATTCTACCCCCGG | 148 | 24 | $(TC)_{24}ACT_5(GT)_{11}$ |
| 49 | RM219* | CT131 | 9 | CGTCGGATGATGTAAAGCCT<br>CATATCGGCATTCGCCTG | 202 | 17 | |

**Table 1** Continued

| | Locus | Clone | Chromosome | Primer sequence[a] | Size (bp) in IR36 | Number of perfect repeats | Description of complex microsatellite motifs |
|---|---|---|---|---|---|---|---|
| 50 | RM220* | CT158 | 1 | GGAAGGTAACTGTTTCCAAC<br>GAAATGCTTCCCACATGTCT | 127 | 17 | |
| 51 | RM221 | CT168 | 2 | ACATGTCAGCATGCCACATC<br>TGCAAGAATCTGACCCGG | 192 | 10 | $(TC)(CT)(TC)_4T_3C_3(TC)(CT)_2C_3T_2(TC)_2(CT)_{10}(TC)$ |
| 52 | RM222* | CT193 | 10 | CTTAAATGGGCCACATGCG<br>CAAAGCTTCCGGCCAAAAG | 213 | 18 | |
| 53 | RM223* | CT195 | 8 | GAGTGAGCTTGGGCTGAAAC<br>GAGGCAAGTCTTGGCACTG | 165 | 25 | |
| 54 | RM224* | CT199 | 11 | ATCGATCGATCTTCACGAGG<br>TGCTATAAAAGGCATTCGGG | 157 | 13 | $(AAG)_8(AG)_{13}G(AG)$ |
| 55 | RM225* | CT201 | 6 | TGCCCATATGGTCTGGATG<br>GAAAGTGGATCAGGAAGGC | 140 | 18 | |
| 56 | RM227 | CT211 | 3 | ACCTTTCGTCATAAAGACGAG<br>GATTGGAGAGAAAAGAAGCC | 106 | 10 | |
| 57 | RM228 | CT221 | 10 | CTGGCCATTAGTCCTTGG<br>GCTTGCGGCTCTGCTTAC | 154 | 36 | $(CA)_6(GA)_{36}$ |
| 58 | RM229* | CT224 | 11 | CACTCACACGAACGACTGAC<br>CGCAGGTTCTTGTGAAATGT | 116 | 11 | $(TC)_{11}(CT)_5C_3(CT)_5$ |
| 59 | RM230 | CT227 | 8 | GCCAGACCGTGGATGTTC<br>CACCGCAGTCACTTTTCAAG | 257 | 13 | $(AGG)_4(GA)_9A(AG)_{13}$ |
| 60 | RM231* | CT234 | 3 | CCAGATTATTTCCTGAGGTC<br>CACTTGCATAGTTCTGCATTG | 182 | 16 | |
| 61 | RM232* | CT339 | 3 | CCGGTATCCTTCGATATTGC<br>CCGACTTTTCCTCCTGACG | 158 | 24 | |
| 62 | RM233A | CT353A | 2 | CCAAATGAACCTACATGTTG<br>GCATTGCAGACAGCTATTGA | 162 | 20 | |
| 63 | RM233B | CT353B | 5 | CCAAATGAACCTACATGTTG<br>GCATTGCAGACAGCTATTGA | 162 | 20 | |
| 64 | RM234* | CT360 | 7 | ACAGTATCCAAGGCCCTGG<br>CACGTGAGACAAAGACGGAG | 156 | 25 | |
| 65 | RM235 | CT368 | 12 | AGAAGCTAGGGCTAACGAAC<br>TCACCTGGTCAGCCTCTTTC | 124 | 24 | |
| 66 | RM236 | CT16 | 2 | GCGCTGGTGGAAAATGAG<br>GGCATCCCTCTTTGATTCCTC | 191 | 18 | |
| 67 | RM237 | CT372 | 1 | CAAATCCCGACTGCTGTCC<br>TGGGAAGAGAGCACTACAGC | 130 | 18 | |
| 68 | RM238A | CT380A | 1 | GATGGAAAGCACGTGCACTA<br>ACAGGCAATCCGTAGACTCG | 147 | 15 | |
| 69 | RM238B | CT380B | 6 | GATGGAAAGCACGTGCACTA<br>ACAGGCAATCCGTAGACTCG | 147 | 15 | |
| 70 | RM239 | CT387 | 10 | TACAAAATGCTGGGTACCCC<br>ACATATGGGACCCACCTGTC | 144 | 5 | $(AG)_5TG(AG)_2T(GA)_2T(GA)_2GT(GA)_5(AG)_5$ |
| 71 | RM240 | CT388 | 2 | CCTTAATGGGTAGTGTGCAC<br>TGTAACCATTCCTTCCATCC | 132 | 21 | |
| 72 | RM241 | CT404 | 4 | GAGCCAAATAAGATCGCTGA<br>TGCAAGCAGCAGATTTAGTG | 138 | 31 | |
| 73 | RM242* | CT410 | 9 | GGCCAACGTGTGTATGTCTC<br>TATATGCCAAGACGGATGGG | 225 | 26 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 74 | RM243 | CT441 | 1 | GATCTGCAGACTGCAGTTGC<br>AGCTGCAACGATGTTGTCC | 116 | 18 | |
| 75 | RM244 | CT447 | 10 | CCGACTGTTCGTCCTTATCA<br>CTGCTCTCGGGTGAACGT | 163 | 8 | $(CT)_4(CG)_3C(CT)_6(CG)_2AC$<br>$(CT)_2$-6bp-$(TC)_4$-8bp-$(CT)_8$ |
| 76 | RM245 | CT453 | 9 | ATGCCGCCAGTGAATAGC<br>CTGAGAATCCAATTATCTGGGG | 150 | 14 | |
| 77 | RM246 | CT461 | 1 | GAGCTCCATCAGCCATTCAG<br>CTGAGTGCTGCTGCGACT | 116 | 20 | |
| 78 | RM247* | CT462 | 12 | TAGTGCCGATCGATGTAACG<br>CATATGGTTTTGACAAAGCG | 131 | 16 | |
| 79 | RM248 | CT469 | 7 | TCCTTGTGAAATCTGGTCCC<br>GTAGCCTAGCATGGTGCATG | 102 | 25 | |
| 80 | RM249* | CT481 | 5 | GGCGTAAAGGTTTTGCATGT<br>ATGATGCCATGAAGGTCAGC | 121 | 14 | $(AG)_5A_2(AG)_{14}$ |
| 81 | RM250 | CT482 | 2 | GGTTCAAACCAAGCTGATCA<br>GATGAAGGCCTTCCACGCAG | 153 | 17 | |
| 82 | RM251 | CT489 | 3 | GAATGGCAATGGCGCTAG<br>ATGCGGTTCAAGATTCGATC | 147 | 29 | |
| 83 | RM252* | CT206 | 4 | TTCGCTGACGTGATAGGTTG<br>ATGACTTGATCCCGAGAACG | 216 | 19 | |
| 84 | RM253* | CT506 | 6 | TCCTTCAAGAGTGCAAAACC<br>GCATTGTCATGTCGAAGCC | 141 | 25 | |
| 85 | RM254 | CT514 | 11 | AGCCCCGAATAAATCCACCT<br>CTGGAGGAGCATTTGGTAGC | 165 | 11 | $(TC)_6ATT(CT)_{11}$ |
| 86 | RM255 | CT519 | 4 | TGTTGCGTGTGGAGATGTG<br>CGAAACCGCTCAGTTCAAC | 144 | 16 | $(AGG)_5(AG)_2ATC(AG)(GA)_{16}$ |
| 87 | RM256 | CT520 | 8 | GACAGGGAGTGATTGAAGGC<br>GTTGATTTCGCCAAGGGC | 127 | 21 | |
| 88 | RM257 | CT522 | 9 | CAGTTCCGAGCAAGAGTACTC<br>GGATCGGACGTGGCATATG | 147 | 24 | |
| 89 | RM258 | CT531 | 10 | TGCTGTATGTAGCTCGCACC<br>TGGCCTTTAAAGCTGTCGC | 148 | 21 | $(GA)_{21}(GGA)_3$ |
| 90 | RM259* | CT550 | 1 | TGGAGTTTGAGAGGAGGG<br>CTTGTTGCATGGTGCCATGT | 162 | 17 | |
| 91 | RM260 | CT553 | 11 | ACTCCACTATGACCCAGAG<br>GAACAATCCCTTCTACGATCG | 111 | 34 | |
| 92 | RM261 | CT563 | 4 | CTACTTCTCCCCTTGTGTCG<br>TGTACCATCGCCAAATCTCC | 125 | 8 | $C_9(CT)_8$ |
| 93 | RM262 | CT565 | 2 | CATTCCGTCTCGGCTCAACT<br>CAGAGCAAGGTGGCTTGC | 154 | 16 | |
| 94 | RM263* | CT580 | 2 | CCCAGGCTAGCTCATGAACC<br>GCTACGTTTGAGCTACCACG | 199 | 34 | |

[a] Forward primer listed on the first line and Reverse primer listed underneath
[b] An asterisk (*) after the locus designation indicates a particularly strong PCR amplification of microsatellite markers
[c] A $\chi$ after the locus designation indicates a particularly faint PCR amplification of microsatellite markers

were genotyped unequivocally on the mapping populations used in this study.

Genetic mapping of microsatellites

The 94 microsatellite markers developed in this study, as well as the 27 previously developed by Wu and Tanksley (1993) and Panaud et al. (1996), were mapped onto the same populations so that the locations of all 121 SSLPs could be integrated into a common map (Fig. 2). Of the SSLPs 90% have been mapped onto at least two populations, with identical positions being demonstrated in every case. Ninety-seven primer pairs showed polymorphism and were mapped onto the $DH_1$ ('IR64'/'Azucena') population; 92 were mapped onto the $DH_2$ (ZYQ/JX) population, including 12 primer pairs that were not polymorphic in the $DH_1$ parents; and 90 were mapped onto the RI ('Milyang 23'/ 'Gihobyeo') population, including 2 new polymorphisms. Though the SL map contained the greatest number of RFLP markers (Causse et al. 1994) and the frequency of polymorphism between the two parents, BS125/*O. longistaminata*, was high (92%), only 65% heterozygosity was detected in the $F_1$ interspecific hybrid. This can be explained by the fact that the *O. longistaminata* parent is a highly heterozygous, obligate outcrossing species that sometimes shared one of its alleles with BS125 such that the hybrid $F_1$ only inherited a polymorphic allele 65% of the time. Because a higher level of polymorphism was detected in the *indica*/*japonica* crosses, as well as a higher rate of recombination, microsatellite markers were preferentially mapped onto the $DH_1$, $DH_2$, or RI populations. The interspecific SL population was used primarily to confirm the locations of markers that detected polymorphism in only one of the other populations or that showed distorted segregation in one or more populations.

To provide a convenient point of reference, we present the RFLP framework map based on the SL population (used extensively for gene and QTL localization) alongside the $DH_1$ map in Fig. 2, and common RFLP markers are joined by a dashed line to facilitate identification of microsatellite locations relative to the high-density map published by Causse et al. (1994). Markers that were mapped only onto the $DH_2$, RI, or SL populations are underlined and shown to the right of markers mapped directly onto the $DH_1$ population (Fig. 2).

Genome coverage and distribution

The 121 microsatellite loci were distributed throughout the 12 rice chromosomes and showed no significant clustering near centromeres or distal regions of chromosomes, at least not at the level of resolution examined here. To test whether the microsatellite markers were in accordance with the physical length of prometaphase chromosomes estimated by image analysis (Fukui and Iijima 1991), we used chi-square analysis to compare the observed to the expected number of SSLPs per chromosome (Table 2). On a genome-wide basis, there was excellent agreement between number of markers and cytogenetic length of chromosomes (no significant deviation from expected: $\chi^2 = 9.94$, $P > 0.5$).

On the map reported by Causse et al. (1994) there were fewer markers on chromosome 10 than expected. In this study, 4 microsatellites extended the short arm of chromosome 10 by approximately 20 cM, filling in a region that had been poorly represented on the RFLP map published by Causse et al. (1994). These SSLP markers mapped to a region of chromosome already defined by RFLP markers on the map published by Kurata et al. (1994), as witnessed by the placement of the Japanese marker G 1084.

Maps derived from the $DH_1$ and $DH_2$ populations used in this study contained similar numbers of RFLP framework markers (approximately 150) and SSLPs (80–90) and similar total map distances (approximately 1900 cM), and both averaged 1 microsatellite every 16–20 cM. However, the map distances of these DH populations were approximately 12–13% larger than was observed for the interspecific backcross population reported by Causse et al. (1994) (1,491 cM defined by

**Table 2** Distribution of microsatellite markers based on cytogenetic length of chromosomes

| Chromosome[a] | Relative cytogenic length (%) in pro-metaphase | Number of microsatellite loci per chromosome | | Z-score[d] |
|---|---|---|---|---|
| | | Expected[b,c] | Observed[c] | |
| 1 | 13.6 | 16.5 | 17 | 0.12 |
| 2 | 10.9 | 13.2 | 17 | 1.05 |
| 3 | 11.7 | 14.2 | 15 | 0.21 |
| 4 | 9.1 | 11.0 | 4 | – 2.11 |
| 5 | 6.6 | 8.0 | 8 | 0.00 |
| 6 | 8.3 | 10.0 | 8 | – 0.63 |
| 7 | 6.1 | 7.4 | 11 | 1.32 |
| 8 | 7.6 | 9.2 | 9 | – 0.07 |
| 9 | 5.8 | 7.0 | 8 | 0.38 |
| 10 | 5.8 | 7.0 | 6 | – 0.38 |
| 11 | 6.6 | 8.0 | 11 | 1.06 |
| 12 | 7.9 | 9.6 | 7 | – 0.84 |
| Total | 100 | 121.0 | 121 | |

[a] Chromosome numbering systems used by Fukui and Iijima (1991) and McCouch and Tanksley (1991) were harmonized according to Khush and Kinoshita (1991)

[b] Based on percentage of physical length

[c] Chi-square test [(observed-expected)/expected; degree of freedom = 11] indicated that there is no significant deviation from random distribution of microsatellites/chromosome based on their cytogenetic size ($\chi^2 = 9.94$, $P < 0.5$)

[d] Z-score = (observed-expected)/expected)$^{1/2}$, Z-score $> 3$ or $< -3$ indicates significant deviation from expected number of markers/chromosome

680 DNA markers) or the intraspecific $F_2$ population reported by Kurata et al. (1994) (1,575 cM defined by 1,383 DNA markers). This may be due to a higher recombination frequency of male (anther culture)-derived populations than female-derived populations, as has been reported for barley (Devaux et al. 1995). The higher rate of recombination provided an advantage for mapping SSLP markers in rice.

## Multiple loci

Mapping of the amplification products of 91 primer pairs developed during this study revealed 94 loci. Three of the markers involved multiple loci, 2 from the enzyme-digested library and 1 from the sheared library. Markers that detected more than one locus were given a suffix of A or B following the RM designation (Table 1 and Fig. 2). Multiple-copy markers identified in this study included RM238A on chromosome 1 and RM238B on chromosome 6, RM233A on chromosome 2 and RM233B on chromosome 5, and RM81A on chromosome 1 and RM81B on chromosome 3. Two additional multiple-copy markers, RM4 and RM20, had been mapped to a conserved region of chromosomes 11 and 12 in a previous study (Panaud et al. 1996).

## Redundancy

The basic local alignment search tool (BLAST) (Altschul et al. 1990) was used to identify redundant poly (GA)-containing clones among the 292 DNA sequences obtained from the small-insert libraries used

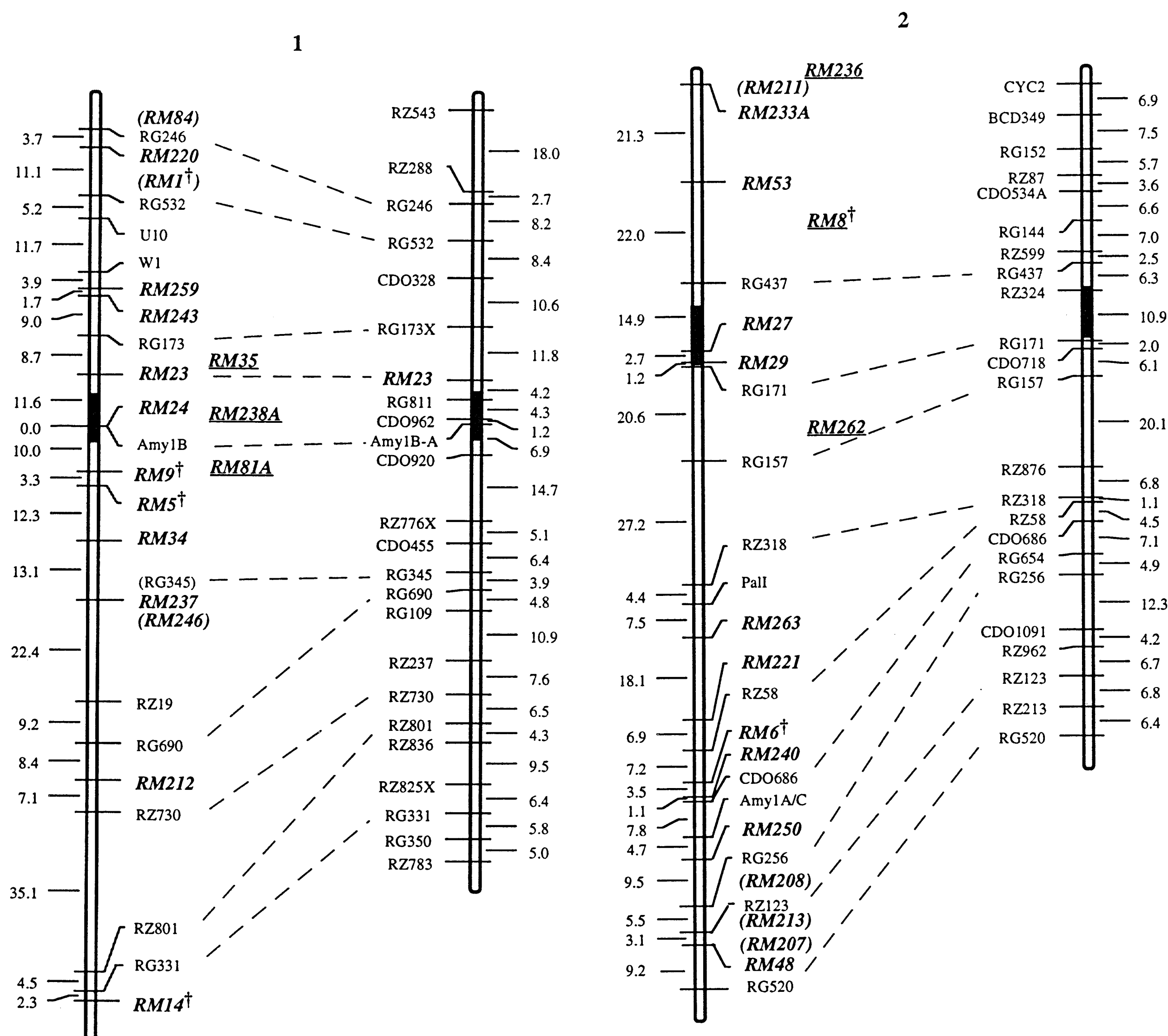



**Fig. 2** See page 565 for legend

in this study. Clones from the *Tsp*509-digested library contained a higher level of redundancy (20%) than did clones from the sheared library (7%). Out of 122 clones from the *Tsp*509-digested library, 12 sequences were each isolated twice, 4 sequences were isolated three times, and 1 sequence was isolated four times, leaving 97 unique poly (GA)n clones for primer design and motif analysis. Among overlapping or redundant clones isolated from the *Tsp*509-digested library, about 50% differed in insert size, an observation that is consistent with the fact that partial digestion was used to generate the library. Of the 170 DNA sequences from the sheared library (Panaud et al. 1996; and this study), 10 sequences were each isolated twice, and 1 sequence was isolated three times, leaving 159 unique poly (GA)-containing clones. Redundancy between libraries was minimal, with only 3 of the 292 clones (1%) having common sequences. These results indicate that the two libraries provided largely independent sources of microsatellite markers.

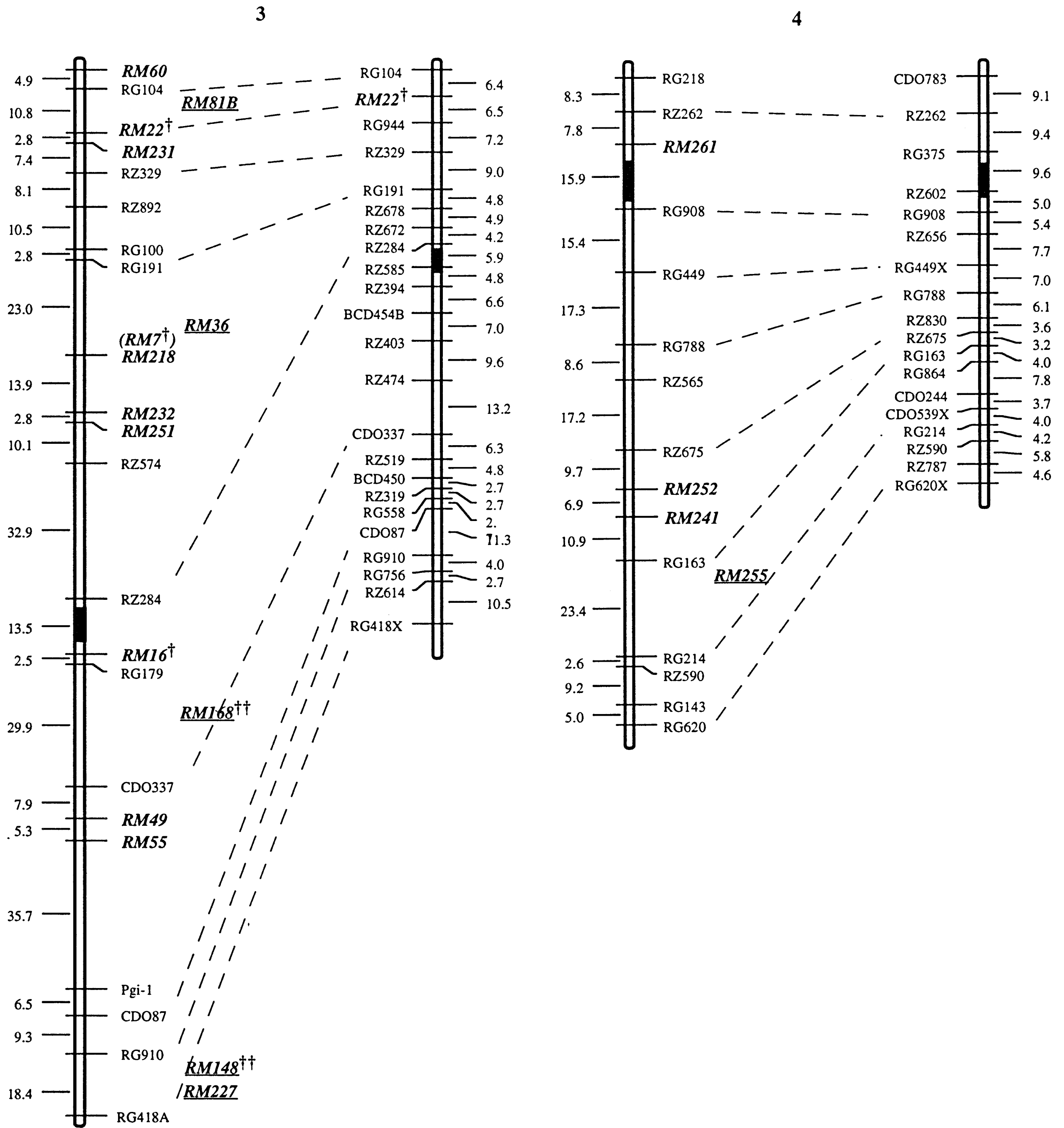



**Fig. 2** See page 565 for legend

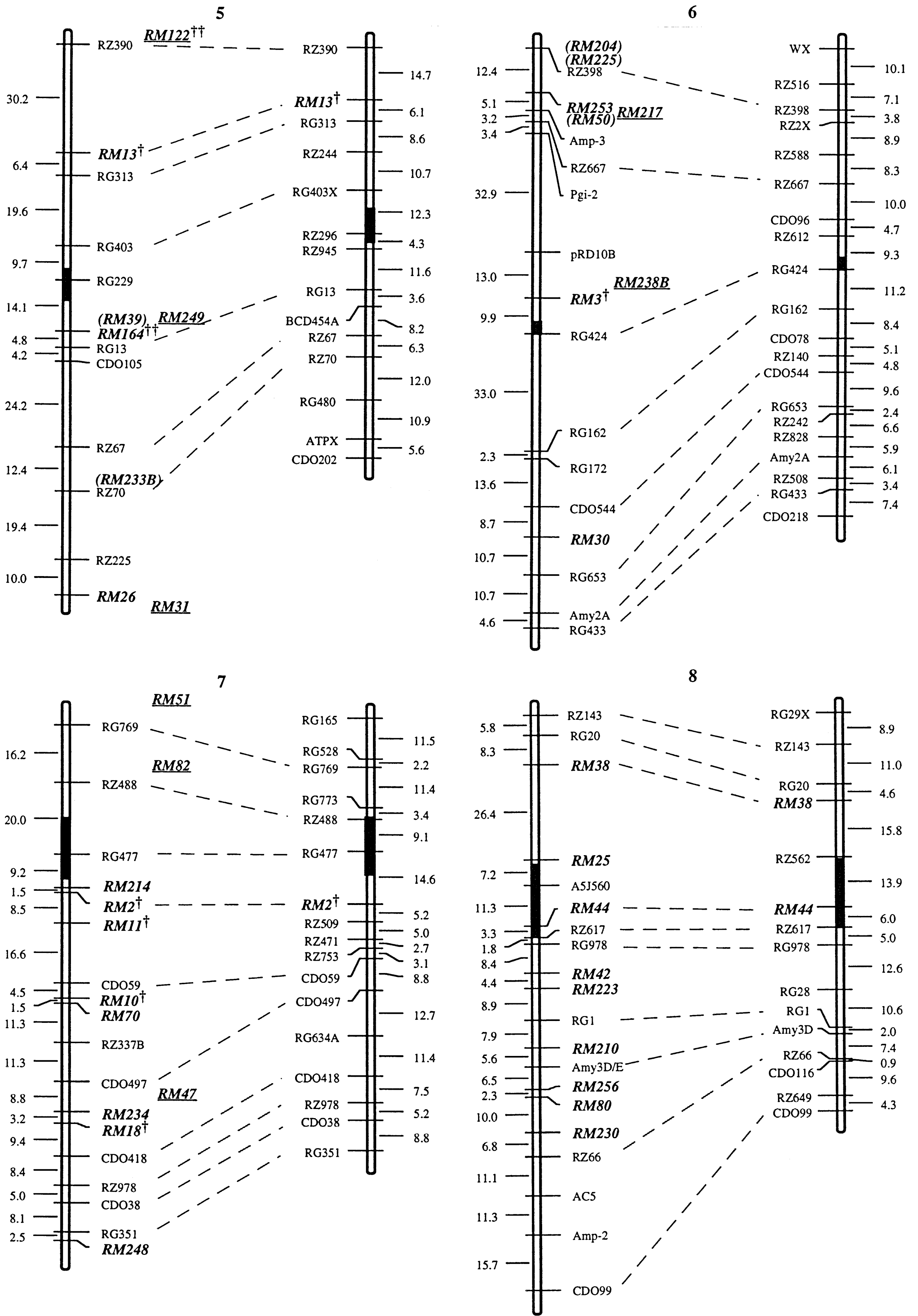
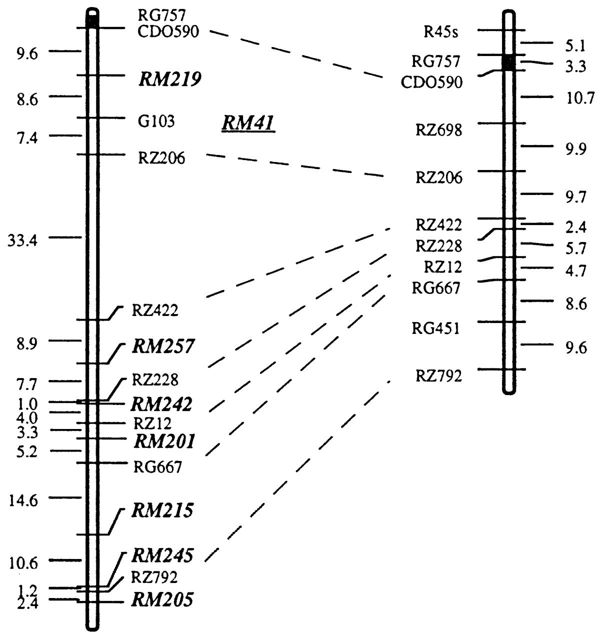
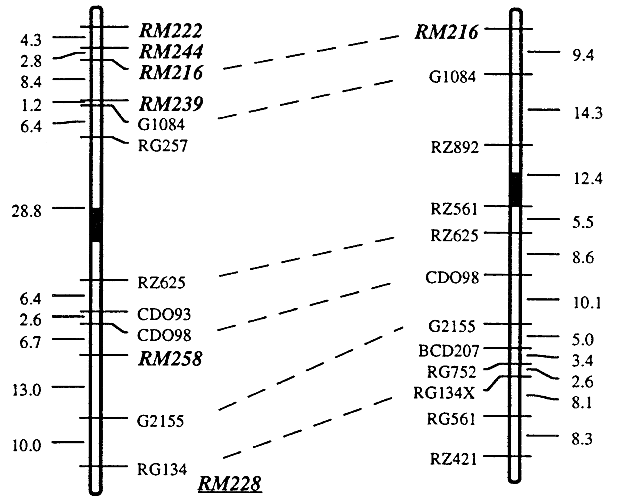
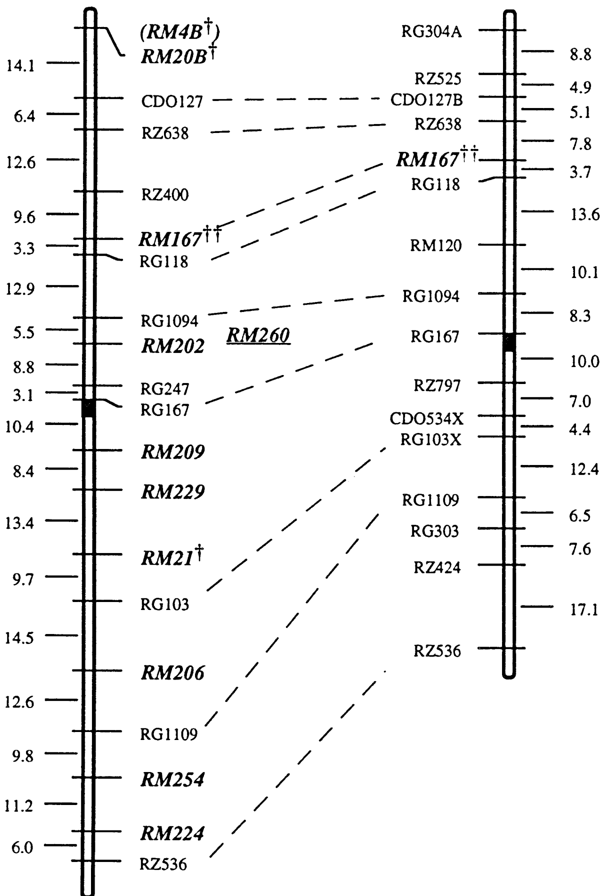



**Fig. 2** See page 565 for legend


9
RG757
CDO590
RM219
G103
RM41
RZ206
RZ422
RM257
RZ228
RM242
RZ12
RM201
RG667
RM215
RM245
RZ792
RM205
9.6
8.6
7.4
33.4
8.9
7.7
1.0
4.0
3.3
5.2
14.6
10.6
1.2
2.4
R45s
RG757
CDO590
RZ698
RZ206
RZ422
RZ228
RZ12
RG667
RG451
RZ792
5.1
3.3
10.7
9.9
9.7
2.4
5.7
4.7
8.6
9.6


10
RM222
RM244
RM216
RM239
G1084
RG257
RZ625
CDO93
CDO98
RM258
G2155
RG134
RM228
4.3
2.8
8.4
1.2
6.4
28.8
6.4
2.6
6.7
13.0
10.0
RM216
G1084
RZ892
RZ561
RZ625
CDO98
G2155
BCD207
RG752
RG134X
RG561
RZ421
9.4
14.3
12.4
5.5
8.6
10.1
5.0
3.4
2.6
8.1
8.3


11
(RM4B†)
RM20B†
CDO127
RZ638
RZ400
RM167††
RG118
RG1094
RM202
RM260
RG247
RG167
RM209
RM229
RM21†
RG103
RM206
RG1109
RM254
RM224
RZ536
14.1
6.4
12.6
9.6
3.3
12.9
5.5
8.8
3.1
10.4
8.4
13.4
9.7
14.5
12.6
9.8
11.2
6.0
RG304A
RZ525
CDO127B
RZ638
RM167††
RG118
RM120
RG1094
RG167
RZ797
CDO534X
RG103X
RG1109
RG303
RZ424
RZ536
8.8
4.9
5.1
7.8
3.7
13.6
10.1
8.3
10.0
7.0
4.4
12.4
6.5
7.6
17.1

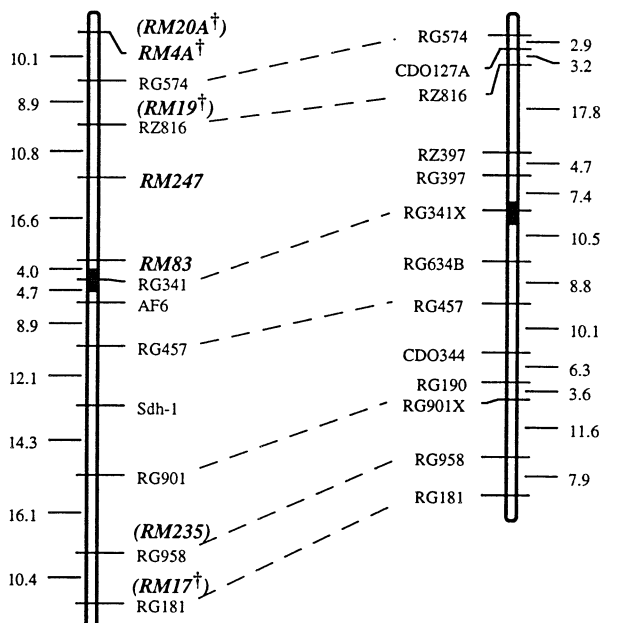

12
(RM20A†)
RM4A†
RG574
(RM19†)
RZ816
RM247
RM83
RG341
AF6
RG457
Sdh-1
RG901
(RM235)
RG958
(RM17†)
RG181
10.1
8.9
10.8
16.6
4.0
4.7
8.9
12.1
14.3
16.1
10.4
RG574
CDO127A
RZ816
RZ397
RG397
RG341X
RG634B
RG457
CDO344
RG190
RG901X
RG958
RG181
2.9
3.2
17.8
4.7
7.4
10.5
8.8
10.1
6.3
3.6
11.6
7.9

**Table 3** Summary of microsatellite motifs selected from the sheared and the *Tsp*509-digested libraries

| Library | Number of clones sequenced | Average number of $(GA)_n$ repeats | Distribution of $(GA)_n$ repeats | | |
|---|---|---|---|---|---|
| | | | Perfect | Imperfect | Compound |
| Enzyme-digested | 97 | $20.53 \pm 8.99$ | 76.29% | 11.34% | 12.37% |
| Sheared | 159 | $16.90 \pm 6.38$ | 77.36% | 17.61% | 5.02% |

### Length and complexity of rice microsatellite motifs

As summarized in Table 3, rice microsatellites were of varying complexity and length. The motifs were comprised of perfect repeat sequences without interruptions, imperfect repeats with two or more runs of repeats interspersed with single nucleotides, and compound repeats with complex combinations of microsatellites (Weber 1990). In this study, 256 poly(GA)-containing sequences (159 from the sheared library, 97 from the *Tsp*509-digested library) were analyzed based on the length of microsatellite repeats and the complexity of those motifs. The percentage of clones containing perfect $(GA)_n$ repeats was similar in the two libraries, but the clones from the *Tsp*509-digested library contained longer perfect poly(GA) tracts, with a mean of 21 repeat units, compared to a mean of 17 for the sheared library ($t$-test $= 3.47$, $P < 0.001$) (Table 3).The longest perfect repeat motif observed from the *Tsp*509-digested library was poly$(GA)_{64}$ and that from the sheared library was poly$(GA)_{39}$. The libraries also differed significantly in the proportions of imperfect and compound repeats, with more imperfect repeats found in the sheared library but more than twice as many compound repeats found in the *Tsp*509-digested library (Table 3).

## Discussion

The rate of successful amplification from primers designed from *Tsp*509 clone sequences (73%) was higher than that previously reported by Panaud et al. (1996) from sheared clones (30%). We suggested that the low rate of primer amplification from sheared library clones might be due to the presence of chimeric clones in the sheared library or to structural differences in regions of the genome that were cloned in the two genomic libraries. The *Tsp*509-digested library was developed to test the possibility that we could increase the efficiency of microsatellite development. Other changes in our procedures included subjecting this library to a higher-stringency pre-sequencing screening procedure, which reduced the number of clones to be mini-prepped and sequenced, and optimization of our technique based on experience.

There was a higher level of redundancy among clones from the *Tsp*509-digested library compared to those from the physically sheared one. As hypothesized by Rafalski et al. (1996), this may be partly due to the fact that a size-selected library derived from digestion with a single enzyme represents only a subset of the DNA fragments potentially available in a genome. Libraries generated from digestion with multiple enzymes might provide a less biased array of clones, which would potentially lower the level of redundancy in an enzyme-digested library. The problem of redundancy can be expected to increase as more clones are screened from any given library. While sheared libraries might appear to provide the most random array of genomic fragments, in our sheared library there appears to be a tendency toward physical breakage within extended repeat motifs, which would explain the smaller length of poly(GA) motifs in the clones that we isolated. This would be a disadvantage given the positive relationship between the length of the microsatellite motif and allelic diversity in rice (Yang et al. 1994), human (Weber 1990), mouse (Love et al. 1990), *Arabidopsis* (Bell and Ecker 1994), and grapevine (Thomas and Scott 1993). Future work involving more extensive germplasm surveys will reveal the degree of polymorphism (or informativeness) of these markers and the relationship between number of alleles and length and complexity of repeat sequences (Ishii et al., in prep.).

The mapping of 121 microsatellite loci in rice supported the hypothesis that they were randomly distributed throughout the genome and therefore that saturated genome coverage would be possible. The 121 markers covered all 12 chromosomes, and the number of markers per chromosome was approximately proportional to the physical length at pro-metaphase,

◀

**Fig. 2** Rice map showing the location of 121 microsatellite marker loci (shown in *bold* and *italics*). To the *left* is the framework map based on the 'IR64'/'Azucena' doubled haploid population ($DH_1$) (Huang et al. 1994). Microsatellite markers with a LOD $>2$ are integrated into the RFLP framework, and markers with a LOD $<$ 2 are indicated in *parenthesis* in their most probable locations along the chromosomes. Markers mapped onto the $DH_2$, RI or SL populations (see text) are *underlined* and placed to the *side* of the $DH_1$ framework map. Microsatellites marked with † were previously reported by Panaud et al. (1996), and those marked with †† were reported by Wu and Tanksley (1993). To the *right* is the SL framework map based on the interspecific backcross population reported by Causse et al. (1994). Common RFLP markers are joined by a *dashed line* between the two maps to facilitate interpretation of microsatellite locations. *Stippled regions* along the chromosomes represent locations of centromeres as estimated by Singh et al. (1996)

as estimated by Fukui and Iijima (1991). Although our data showed the predicted distribution of microsatellite markers across chromosomes, there was a nearly threefold underrepresentation of markers on chromosome 4 and an overrepresentation of markers on chromosomes 2, 7 and 11 though these deviations were not statistically significant (Table 2). The deviation may be due to the relatively small number of microsatellite markers investigated to date, but differences in SSR frequency per chromosome cannot be ruled out at this time.

Simple sequence repeat polymorphisms have been found linked to several genes of agricultural importance in rice. A microsatellite located in the 5′ end of the waxy gene (at the distal end of the short arm of chromosome 6) (Bligh et al. 1995) revealed high polymorphism and was found to be a valuable predictor of starch quality in commercial rice breeding (Ayres et al. 1997). On chromosome 7, two microsatellite markers developed by Panaud et al. (1996) were recently found to be closely linked to the giant embryo gene, $ge^s$ (Koh et al. 1996), which controls enlarged embryo size and high protein, oil, and vitamin content. Two of the microsatellite markers developed in this study showed linkage to the wide compatibility trait, $S\text{-}5^n$, in *indica/japonica* hybrids of rice (Yanagihara and Temnykh, unpublished data), suggesting that they can be useful for tracing the inheritance of "wide compatibility" in a breeding program. The work done by Blair and McCouch (1997) showed that 3 microsatellite loci were tightly linked to the resistance gene, *xa-5*, and were efficiently used with RFLP markers to establish a high-density map around this gene. Xiao et al. (1996b) analyzed QTLs for rice yield and demonstrated that 2 microsatellite markers were located in the vicinity of a QTL for yield on chromosome 1.

This work indicates that construction of a saturated microsatellite map should be feasible for rice. The current level of genome coverage is approximately 1 microsatellite every 16–20 cM depending on the cross used to map, which is sufficient for diversity analysis and genotype identification (Yang et al. 1994; Xiao et al. 1996a; Olufowote et al. 1997), gene and QTL analysis (Koh et al. 1996; Xiao et al. 1996b), and marker-assisted selection (Blair and McCouch 1997). A concerted effort to generate and map more of these highly informative, PCR-based, co-dominant markers will improve the resolution of the rice map and provide the basis for fine-scale genome analysis, positional cloning and targeted genetic improvement.

**Acknowledgements** We gratefully acknowledge the contributions of mapping populations and RFLP data from N. Huang at the IRRI, Philippines ($DH_1$); L. Zhu at the Chinese Academy of Sciences, China ($DH_2$); and M. Y. Eun of the Korean Rice Genome project, Suweon, Korea (RI). We are grateful for financial support from the USDA (NRI Competitive Grant 93-37300-8703) and the Rockefeller Foundation (RF 95001, no. 315). Our thanks are extended to Steve Daughhetee (Biochemistry Department, Cornell University) for providing the sequence data for CT 100, Dr. G. Churchill for statistical advice, Carole Morehouse for help in formatting this manuscript and to E. Radwanski, T. Fulton, and S. Grandillo for helpful suggestions on the manuscript.

## References


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

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, Lipman DJ (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

Becker J, Heun M (1995) Barley microsatellites: allele variation and mapping. Plant Mol Biol 27: 835–845

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

Blair M, McCouch SR (1997) Microsatellite and sequence-tagged site markers diagnostic for the bacterial blight resistance gene, *xa-5*. Theor Appl Genet 95: 174–184

Bligh HFJ, Till RI, Jones CA (1995) A microsatellite sequence closely linked to the *Waxy* gene of *Oryza sativa*. Euphytica 86: 83–85

Bretting PK, Widrlechner MP (1995) Genetic markers and plant genetic resource management. Plant Breed Rev 13: 11–86

Broun P, Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol Gen Genet 250: 39–49

Brown, SM, Kresovich S (1996) Molecular characterization for plant genetic resources conservation. In: Paterson AH (ed) Genome mapping in plants. RG Landes, New York, pp 85–93

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

Devaux P, Kilian A, Kleinhofs A (1995) Comparative mapping of the barley genome with male and female recombination-derived, doubled haploid populations. Mol Gen Genet 249: 600–608

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morisette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380: 152–154

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

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. BioTechniques 20: 758–760

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

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

Hahn WJ, Grifo FT (1996) Molecular markers in plant conservation genetics. In: Sobral BWS (ed) The impact of plant molecular genetics, Birkhauser, Boston, pp 113–136

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

Huang N, Parco A, Mew T, Magpantay G, McCouch SR (1997) RFLP mapping of isozymes, RAPD and QTLs for grain shape, brown planthopper resistance in a double haploid rice population. Mol Breed 3 : 105–113

Khush GS, Kinoshita T (1991) Rice karyotype, marker genes, and linkage groups. In: Khush GS, Toenniessen GH (eds) Rice biotechnology. CAB International and IRRI, Oxon, UK, pp 83–108

Koh HJ, Heu MH, McCouch SR (1996) Molecular mapping of the *ge* gene controlling the super-giant embryo character in rice (*Oryza sativa* L.). Theor Appl Genet 93 : 257–261

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

Kurata N, Nagamura Y, Yamamoto K, Harushima Y, Sue N, Wu J, Antonio BA, Shomura A, Shimizu T, Lin SY, Inoue T, Fukuda A, Shimano T, Kuboki Y, Toyama T, Miyamoto Y, Kirihara T, Hayasaka K, Miyao A, Monna L. Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y (1994) A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nat Genet 8 : 365–372

Lander ES, Green P, Abrahamson J, Barlow A, 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 ZW, Biyashev RM, Saghai Maroof MA (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93 : 869–876

Love JM, Knight AM, Macleer MA, Todd JA (1990) Towards construction of a high resolution map of the mouse genome using PCR-analyzed microsatellites. Nucleic Acids Res 18 : 4123–4130

McCouch SR, Tanksley SD (1991) Development and use of restriction fragment length polymorphism in rice breeding and genetics. In: Khush GS, Toenniessen GH (eds) Rice Biotechnology. CAB International and IRRI, Oxon, UK, pp 109–133

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 (in press)

Olufowote JO, Xu Y, Chen X, Park WD, Beachell HM, Dilday RH, 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 (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

Prochzka M (1996) Microsatellite hybrid capture technique for simultaneous isolation of various STR markers. Genome Res 6 : 646–649

Rafalski JA, Vogel JM, Morgante M, Powell W, Andre C, Tingey SV (1996) Generating and using DNA markers in plants. In: Birren B and Lai E (eds) Nonmammalian genomic analysis: a practical guide, Academic Press, New York, pp 75–134

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

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36 : 884–889

Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS (1996) Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). Proc Natl Acad Sci USA 93 : 6163–6168

Smith S, Helentjaris T (1996) DNA fingerprinting and plant variety protection. In: Paterson AH (ed) Genome mapping in plants. RG Landes, New York, pp 95–110

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor Appl Genet 86 : 986–990

Weber JL (1990) Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. Genomics 7 : 517–524

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 24 : 225–235

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

Xiao J, Grandillo, Ahn SN, McCouch SR, Tanksley SD, Li J, Yuan L (1996b) Genes from wild rice improve rice yield. Nature 384 : 223–224

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

Zhu LH, Chen Y, Xu YB, Xu JC, Cai HW, Ling ZZ (1993) Construction of a molecular map of rice and gene mapping using a double haploid population of a cross between Indica and Japonica varieties. Rice Genet Newsl 10 : 132–134