Theor Appl Genet (1999) 98: 780–792

M. W. Blair · O. Panaud · S. R. McCouch

Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.)

Received: 25 May 1998 / Accepted: 17 September 1998

Abstract Inter-simple sequence repeat (ISSR) amplification was used to analyze microsatellite motif frequency in the rice genome and to evaluate genetic diversity among rice cultivars. A total of 32 primers, containing different simple sequence repeat (SSR) motifs, were tested for amplification on a panel of 59 varieties, representative of the diversity of cultivated rice (*Oryza sativa* L.). The ISSR analysis provided insights into the organization, frequency and levels of polymorphism of different simple sequence repeats in rice. The more common dinucleotide motifs were more amenable to ISSR analysis than the more infrequent tri-, tetra- and penta-nucleotide motifs. The ISSR results suggested that within the dinucleotide class, the poly(GA) motif was more common than the poly(GT) motif and that the frequency and clustering of specific tri- and tetra-nucleotide simple sequence repeats was variable and motif-specific. Furthermore, trinucleotide ISSR markers were found to be less polymorphic than either dinucleotide or certain tetranucleotide ISSR markers, suggesting which motifs would be better targets for microsatellite marker development. The ISSR amplification pattern was used to group the rice genotypes by cluster analysis. These results were compared to surveys of the same varieties for amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP) and isozyme markers. The ISSR fingerprint could be used to differentiate the genotypes belonging to either *Japonica* or *Indica* subspecies of cultivated rice and to dissect finer levels of diversity within each subspecies. A higher percentage of polymorphic bands was produced with the ISSR technique than the AFLP method, based on a similar PCR reaction. Therefore, ISSR amplification proved to be a valuable method for determining genetic variability among rice varieties and for rapidly identifying cultivars. This efficient genetic fingerprinting technique would be useful for characterizing the large numbers of rice accessions held in national and international germplasm centers.

Key words Rice (*Oryza sativa*) · Genome analysis · Inter-simple sequence repeats (ISSR) · DNA fingerprinting · Microsatellite motif frequency

Communicated by F. Salamini

M. W. Blair · O. Panaud[1] · S. R. McCouch (✉)
Department of Plant Breeding, 252 Emerson Hall,
Cornell University, Ithaca, NY 14853-1901, USA
Fax: 607-255-6683
E-mail: srm@cornell.edu

Present address:
[1]GEPC, Laboratoire d'évolution et systématique végétale,
Bat. 362, Université Paris XI, 91405 Orsay, Cedex, France

Introduction

Inter-simple sequence repeat (ISSR) analysis involves the polymerase chain reaction (PCR) amplification of regions between adjacent, inversely oriented microsatellites using a single simple sequence repeat (SSR)-containing primer (Zietkiewicz et al. 1994). This technique can be undertaken for any species that contains a sufficient number and distribution of SSR motifs and has the advantage that genomic sequence data is not required (Gupta et al. 1994; Goodwin et al. 1997). The primer used in ISSR analysis can be based on any of the SSR motifs (di-, tri-, tetra- or penta-nucleotides) found at microsatellite loci, giving a wide array of possible amplification products, and can be anchored to genomic sequences flanking either side of the targeted simple sequence repeats (Zietkiewicz et al. 1994).

For ISSR analysis to be successful, pairs of simple sequence repeats must occur within a short distance (in base-pairs) that is amplifiable by a PCR reaction which produces a band that is resolvable on standard

polyacrylamide or agarose gels (Zietkiewicz et al. 1994). The potential supply of ISSR markers depends on the variety and frequency of microsatellites, which changes with the species and the SSR motifs that are targeted (Morgante and Olivieri 1993; Depeiges et al. 1995). If microsatellite distribution is completely random, the length of the intervening regions between simple sequence repeats of the same motif are a function of their frequency. As a result, we hypothesize that the number of bands produced by an ISSR primer with a given microsatellite repeat should reflect the relative frequency of that motif in a given genome and would provide an estimate of the motif's abundance different from library hybridization. Because of our laboratory's interest in microsatellite mapping in rice, we also wanted to use the ISSR technique to investigate the organization, frequency and characteristic level of polymorphism of different SSR motifs to help select motifs to target for future development of new microsatellite markers.

Another important application of the ISSR technique in rice is for the evaluation of genetic diversity. Because the ISSR technique amplifies large numbers of DNA fragments per reaction, representing multiple loci from across the genome, it is an ideal method for fingerprinting rice varieties and a useful alternative to single-locus or hybridization-based methods (Zietkiewicz et al. 1994; Goodwin et al. 1997). The ISSR method has proven especially useful in the Gramineae family for the analysis of nearly isogenic lines (Akagi et al. 1996b) and varieties (Parsons et al. 1997) of rice, inbred lines of corn (Kantety et al. 1995), populations of finger millet (Salimath et al. 1995), and accessions of sorghum (Yang et al. 1996) and pearl millet (O. Panaud, personal communication). An alternative fingerprinting method consists in amplified fragment length polymorphism (AFLP) analysis (Vos et al. 1995); however, previous studies have shown that this technique only detects moderate to low levels of polymorphism in rice (Mackill et al. 1996; Maheswaran et al. 1997; Cho et al. 1997). Efficient molecular fingerprinting techniques would be valuable to both rice breeders and germplasm curators who are faced with the enormous task of conserving, evaluating and using over a quarter million accessions of rice that have been preserved in international and national seed banks in over 42 countries around the world (Jackson 1997).

The specific objectives of the research presented here were to determine the relative abundance and level of polymorphism of different ISSR loci in the rice genome, to compare 5′- and 3′-anchored ISSR primers of different simple sequence repeat motifs, to use the ISSR technique to fingerprint a wide sample of rice cultivars and to evaluate the ISSR and AFLP techniques for this purpose.

Materials and methods

Plant material

Fifty-nine genotypes previously selected to represent the diversity of cultivated rice and analyzed with 150 restriction fragment length polymorphism (RFLP) probes (G. Second, personal communication) were used in this study (information on accession number, cultivar origin and marker genotypes is available in the RiceGenes database, http://probe.nalusda.8300). A subset of the same cultivars was also analyzed by Glaszmann (1987) for isozyme markers and by Wang and Tanksley (1989) for additional RFLPs. Seed for these varieties was kindly provided by G. Second from his original study performed at the International Rice Research Institute (IRRI), and DNA was extracted as a pooled sample from freshly harvested leaves of 20 plants according to Causse et al. (1994).

ISSR and AFLP amplifications

A total of 32 primers were tested for ISSR amplification in the rice genome (Table 1). Six of the primers were anchored at the 3′ end, 26 primers were anchored at the 5′ end. For anchoring to the 3′ border 1 selective nucleotide differing from the repeat motif was used. The six 3′-anchored primers (Integrated DNA Technologies) were all 19-mers containing a single dinucleotide motif, either GA or GT repeated nine times, with 1 selective nucleotide (A, C or T) following the repeat motif. For anchoring to the 5′ border of the simple sequence repeat, a tail consisting of the bases GATC was incorporated into the design of the primers. The 5′-anchored primers (Genosys) ranged from 22 to 36 nucleotides in length, and were designed for four dinucleotide motifs, eight trinucleotide repeats, 11 tetranucleotide repeats and three pentanucleotide repeats.

A single primer was used in each PCR reaction, which was carried out in a total volume of 25 µl containing 1 µl of genomic DNA (100 ng) and 2 µl of primer (diluted to 10 µ*M* concentration), 2.5 µl of 10 × buffer (0.1 *M* TRIS pH 8.3; 0.5 *M* KCl; 7.5 m*M* $MgCl_2$; 0.1% gelatin), 1.0 µl of 100 µ*M* dNTPs and 1.0 unit of *Taq* polymerase. PCR amplifications were performed in 96-well plates on a PTC100 (MJ Research) Thermal Cycler under the following conditions: 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 ISSR amplification products were mixed with a loading buffer (98% formamide, 10 m*M* EDTA, 0.005% each of xylene cyanol and bromophenol blue as tracking dyes), denatured at 90°C for 2 min, run for 2.5 h on 4% denaturing polyacrylamide gels at a constant 70 W and detected by silver staining (Promega, Madison, Wis.). The molecular-weight size ladders, Markers V and VIII (Boehringer Mannheim, Indianapolis, Ind.) were used for band sizing.

A subset of 54 of the 59 varieties were genotyped with the AFLP technique. The digestion and ligation of genomic DNA with *Pst*I/*Taq*I adaptors and PCR amplification to prepare the template were as described by Cho et al. (1996). Three PCR reactions were performed using one of three *Taq* primers, TQ1 (TGAGTCCTGACCGA-ACC), TQ2 (TGAGTCCTGACCGA-ACA) or TQ3 (TGAGTCCTGACCGA-CAA), all with 3 selective nucleotides (indicated after the hyphen), and one of two *Pst* primers, PS1 (GACTGCGTACATGCAG-A), with only 1 selective nucleotide or PS2 (GACTGCGTACATGCAG-AC), with 2 selective nucleotides (Vos et al. 1995). Three primer combinations, sampling the five possible primers, were used: PS1/TQ1, PS2/TQ2 and PS2/TQ3, with an annealing temperature of 60°C. PAGE separation and silver staining were performed as described above.

Table 1 Inter-simple sequence repeat (ISSR) primers: their sequence, GC content and melting temperature (T_m), the number and size range of bands that they produce and the level of polymorphism and average genetic similarity between rice varieties that they detect (*n.a.* not available)

Primer type	Sequence[a]	GC content (%)	T_m[b] (°C)	Total no. of bands	Total no. of monomorphisms	Total polymorphism (%)	MW size range[c]	Average similarity index[d]
3′-anchored primers								
di-nt	(**GA**)9-A	47.4	53.1	63	3	95.2	200–1200	32.5
	(**GA**)9-C	52.6	55.3	54	1	98.1	100–940	25.6
	(**GA**)9-T	47.4	53.1	55	3	94.5	165–1500	39.1
	(**GT**)9-A	47.4	53.1	14	0	100.0	500–1600	22.7
	(**GT**)9-C	52.6	55.3	19	0	100.0	270–1500	26.5
	(**GT**)9-T	47.4	53.1	5	0	100.0	600–1100	n.a.
5′-anchored primers								
di-nt	CCCGGATCC(**GA**)9	59.3	65.8	66	6	90.9	150–1000	39.1
	CCCGGATCC(**CT**)9	59.3	65.8	67	8	88.1	150–1500	44.1
	CCCGGATCC(**GT**)9	59.3	65.8	47	11	76.6	210–1500	61.2
	CCCGGATCC(**CA**)9	59.3	65.8	55	10	81.8	200–800	49.5
tri-nt	GATC(**CGG**)9	93.5	82.2	17	7	58.8	690–1200	76.7
	GATC(**TGG**)9	64.5	70.3	31	14	54.8	350–1400	71.4
	GATC(**GAG**)9	64.5	70.3	16	9	43.8	690–1200	79.8
	GATC(**CAG**)9	64.5	70.3	0	n.a.	n.a.	n.a.	n.a.
	GATC(**ATC**)10	35.3	59.8	31	26	16.1	700–1500	94.3
	GATC(**TTG**)10	35.3	59.8	21	7	66.7	200–1500	62.6
	GA(**TCT**)10	34.4	58.5	15	2	86.7	350–1400	27.9
	GATC(**ATT**)10	5.9	47.7	0	n.a.	n.a.	n.a.	n.a.
tetra-nt	GATC(**GAGG**)7	71.9	73.8	4	0	100.0	500–1200	13.9
	GATC(**GTAG**)7	50.0	64.9	30	0	100.0	200–1400	25.4
	GATC(**AGTG**)7	50.0	64.9	23	2	91.3	180–1500	38.2
	GATC(**TTCG**)8	50.0	66.6	3	1	66.7	240–1200	9.9
	GATC(**TCTG**)8	50.0	66.6	28	20	28.6	200–1400	88.8
	GATC(**GACA**)7	50.0	64.9	0	n.a.	n.a.	n.a.	n.a.
	GATC(**GATC**)7	50.0	64.9	0	n.a.	n.a.	n.a.	n.a.
	GATC(**TTAG**)8	27.8	57.5	0	n.a.	n.a.	n.a.	n.a.
	GATC(**TATC**)7	28.1	55.9	0	n.a.	n.a.	n.a.	n.a.
	GATC(**TATG**)8	27.8	57.5	0	n.a.	n.a.	n.a.	n.a.
	GATC(**GATA**)7	28.1	55.9	0	n.a.	n.a.	n.a.	n.a.
penta-nt	GA(**TCTAG**)6	40.6	61.0	0	n.a.	n.a.	n.a.	n.a.
	GAT(**CTATG**)6	39.4	61.0	0	n.a.	n.a.	n.a.	n.a.
	GATC(**TTTTC**)6	23.5	54.9	0	n.a.	n.a.	n.a.	n.a.

[a] The motif of the primer is highlighted in bold
[b] McConaughy equation taken from Biochemistry 8 : 3289-3295 (1969): $T_m = 60 + 41(G + C)/L - 500/L$ (where L = length of primer)
[c] Size range in nucleotides, based on 1% agarose and 4% polyacrylamide gels
[d] Average Jaccard's genetic similarity index across all variety comparisons

Data analysis

Polymorphic bands from the ISSR and AFLP analysis were scored qualitatively for presence (1) or absence (0) in each variety. The genetic associations between varieties were evaluated by calculating the Jaccard similarity coefficient for pairwise comparisons based on the proportion of shared bands produced by the primers. Similarity matrices were generated using the "simqual" subprogram of the software NTSYS-PC (Numerical Taxonomy System) (Rohlf 1992) and were compared by calculating the product-moment correlation (r) between matrices and the Mantel (Daniel's) test statistic (Z) of significance (Manly 1997) using the "mxcomp" subprogram. The similarity coefficients were also used for cluster analysis of the varieties, performed using the "sahn" subprogram of NTSYS-PC, and to build dendrograms by the unweighted pair-group method with arithmetic average (UPGMA).

Results

ISSR fingerprint complexity is based on the anchoring of the primers and the repeat motif

ISSR bands could be clearly distinguished on 4% denaturing polyacrylamide gels (Fig. 1). PCR amplification using the six 3′-anchored dinucleotide repeat ISSR primers produced varying numbers of DNA fragments depending on their SSR motif (Table 1). The primers that were based on the poly(GA) motif produced significantly more bands on average (57.3 ± 4.9) than the primers based on the poly(GT) motif (12.6 ± 7.1) (paired *t*-test, $P = 0.006$). The same 4.5-fold difference

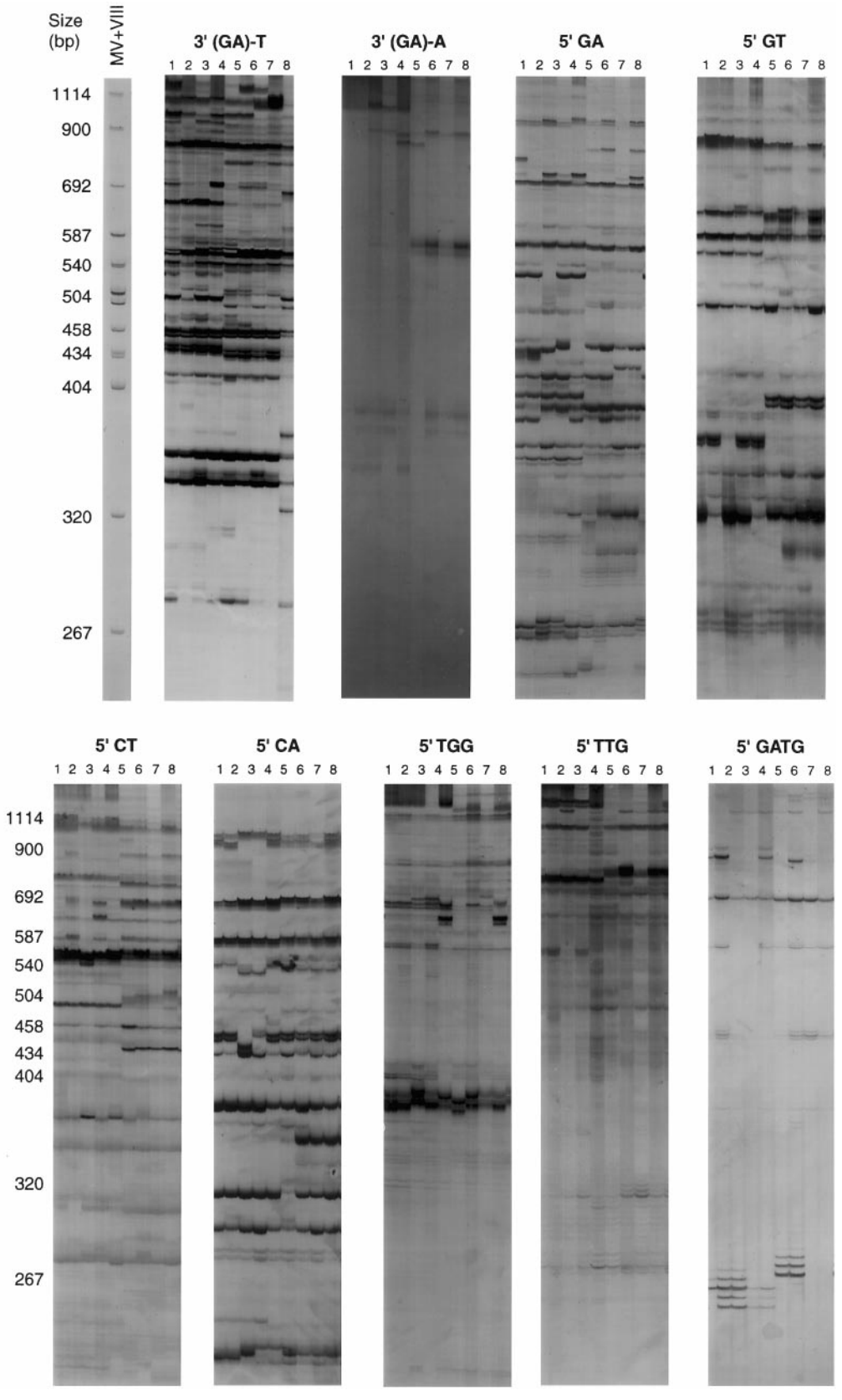

Fig. 1 Inter-simple sequence repeat (ISSR) banding profile on silver-stained 4% polyacrylamide gels for 8 numbered rice varieties [*lanes 1* 'IR64', *2* 'IR42', *3* 'IR36', *4* 'DGWG' (*Indicas*), *5* 'Palawan', *6* 'Moroberakan', *7* 'Rexmont' and *8* 'Lemont' (*Japonicas*)] for representative 3′-anchored primers with the GA and GT motifs, and for the 5′-anchored primers with di-nucleotide motifs (GA, GT, CT and CA), tri-nucleotide motifs (TGG and TTG) and tetra-nucleotide motif (GATG). Size standard (*left panel*) is the combination of molecular-weight markers V and VIII (Boehringer-Mannheim), and the size is indicated in base-pairs

in the numbers of bands was observed with automated fluorescent or radioactive (ρ^{33}) detection of ISSR bands (data not shown) as with detection by silver staining (Fig. 1). The number of bands produced with the different selective nucleotides (A, C or T) varied more with the poly(GT) primers than with the poly(GA) primers. Amplification using the four 5′-anchored dinucleotide repeat ISSR primers produced an average of 58.7 ($\pm$ 9.4) bands over all the genotypes. No significant differences were observed between the number of bands produced by the complementary primer pairs, poly(GA) and poly(CT) or poly(GT) and poly(CA), nor between the average of these pairs.

The size of the bands amplified by dinucleotide-based primers ranged from 100 bp to 1500 bp with a peak concentration around 700 bp, although the specific pattern varied depending on the primer (Fig. 1). The individual bands produced by 5′-anchored dinucleotide ISSR primers showed variable intensity and stuttering, a common feature of SSR amplification caused by strand slippage during DNA replication (Dow et al. 1995). The bands produced by the 3′-anchored ISSR primers were easier to score than those produced by the 5′-anchored primers because they were of a sharper intensity with less stuttering.

The eight trinucleotide, ten tetranucleotide and three pentanucleotide motif ISSR primers did not amplify as many bands as the best dinucleotide motif primers (Table 1). Tri- and tetra-nucleotide-containing primers produced on average 16.4 ($\pm$ 11.9) and 8.8 ($\pm$ 12.7) bands, respectively, while none of the pentanucleotide primers showed any amplification. The average number of bands produced by ISSR primers with motif classes was negatively correlated ($r = -0.911$, $P < 0.005$) with the number of nucleotides in the repeating unit of the motif.

The success in amplifying bands was more variable with ISSR primers that contained tri- or tetra-nucleotide repeats and depended greatly on the specific sequence of the motif. For example, the ISSR primers containing trinucleotide motifs produced 15 or more bands each, except for those based on poly(ATT) and poly(CAG), which produced no bands under the PCR conditions used in this study. Among the tetranucleotide motif ISSR primers, those that were GC-rich (containing 50% or more guanidine and cytosine bases) amplified an average of 14.7 bands, while those that were AT-rich produced no bands at all.

The possibility that these differences in amplification were due to the optimum annealing conditions of the oligonucleotides used as ISSR primers (Panaud et al. 1995) was evaluated by using a range of temperatures for annealing – from 45° to 55°C for AT-rich primers and from 50° to 65°C for GC-rich primers – without modifying the other conditions or times of the PCR program. Because the ISSR primers used in this study were relatively long (27–31 bases), the predicted melting temperatures (T_m) were 55°C or higher for all of the AT-rich primers and 65°C or higher for all the GC-rich primers (Table 1). The primers with 75% or greater AT content (i.e. those based on ATT, TATG, TTAG, TATC, GATA and TTTTC) gave no amplification at either 45°C or 55°C annealing temperatures. Meanwhile, the ISSR primers with greater than one-half GC content produced the same reproducible bands at either 50°C or 65°C annealing temperatures. Larger motif ISSR primers that would dimerize were not used except for one based on poly(GATC) which formed a ladder of amplification products differing in 4 bp that was assumed to be template-independent amplification (data not shown).

The size range of bands produced by specific tri- and tetra-nucleotide ISSR primers was variable and depended on the overall number of bands produced. Primers amplifying many bands produced them over a wide range of molecular weights, visible as a fingerprint of fragments down the length of the polyacrylamide gel, while primers amplifying few bands produced mainly bands more than 700 bp long (Fig. 1).

Frequency of ISSR polymorphism in the variety analysis

The ISSR technique produced enough polymorphic bands to distinguish all 59 varieties from each other based on a single PCR reaction. When all of the ISSR primers were considered together, the rate of polymorphism averaged 78.0% across all the varieties (Table 1). The 3′-anchored primers based on poly(GA) and poly(GT) motifs produced high average polymorphism rates of 95.9% and 100%, with low average similarity indices of 32.4% and 24.6%, respectively. Among the 5′-anchored primers, the dinucleotides as a class had the highest average polymorphism rate (84.4% $\pm$ 6.4), the trinucleotides had the lowest (54.5% $\pm$ 23.6), with the tetranucleotides being intermediate (77.3% $\pm$ 30.5). The average similarity index was negatively correlated ($r = -0.836$, $P < 0.001$) with the level of polymorphism across all the ISSR primers. Correspondingly, the tetranucleotides had a lower average similarity index (35.3% $\pm$ 31.9) than either the dinucleotides (48.5% $\pm$ 9.5) or trinucleotides (68.8% $\pm$ 22.6).

Among the individual 5′-anchored primers, several based on di- or tetra-nucleotide motifs (GA, GAGG, GATG and AGTG) produced over 90% polymorphic bands. While the 5′-anchored dinucleotide primers produced high rates of polymorphism, the complementary primers based on poly(GA) and poly(CT) detected significantly more polymorphisms on average (89.5% $\pm$ 2.0) than the complementary primers based on poly(GT) and poly(CA) (79.2% $\pm$ 3.7) (paired t-test, $P = 0.037$). The rates of polymorphism and levels of similarity among the individual tri- and tetra-nucleotide ISSR primers were much more variable than

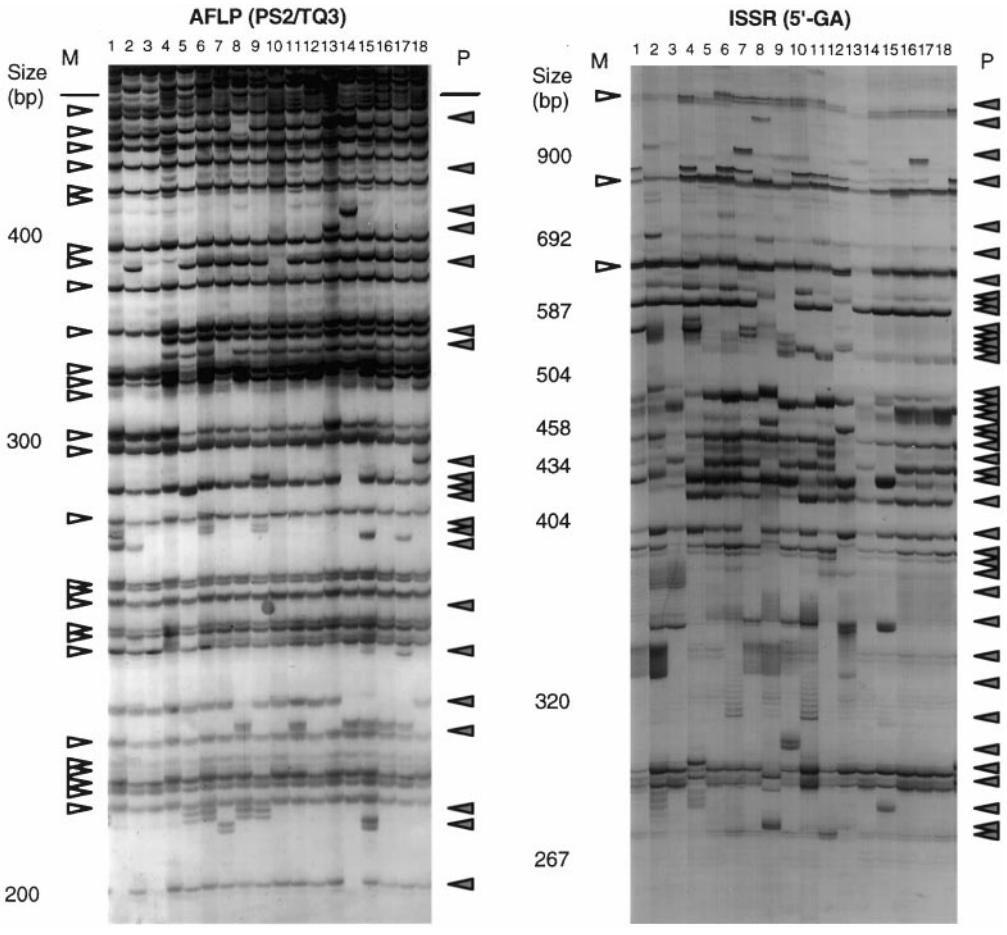

Fig. 2 Comparison of the banding profiles of a subset of 18 varieties for AFLPs, using the primer pair PS2/TQ3, and for ISSR, using the 5′-anchored (GA) motif primer. Bands were resolved on 4% polyacrylamide gels and detected with silver staining. *Arrows* mark PCR fragments which are polymorphic (*right* of panel) or monomorphic (*left* of panel) for the varieties shown. The molecular weights (in base-pairs) of the size standard used are indicated to the *left* of the gels

among the dinucleotide primers, ranging from 16.1% for the poly(ATC) primer to 100% for the poly(GATG) and poly(GAGG) primers.

The clearest ISSR bands were scored for presence or absence among the varieties and used for the cluster analysis. The complete data set was based on a total of 473 bands (370 polymorphisms), consisting of 132 bands (129 polymorphisms) for the 3′-anchored dinucleotide primers and 131 (116), 125 (62) and 85 (63) bands (polymorphisms) for the 5′-anchored di-, tri- and tetra-nucleotide primers, respectively. The data sets for each individual and each class of ISSR primers were analyzed separately and in combination.

AFLP fingerprint pattern and level of polymorphism

The AFLP technique produced bands which were of generally uniform intensity and which had none of the stutter bands characteristic of 5′-anchored ISSRs (Fig. 2). This was important because of the large number of bands per lane and the short separation between bands. PCR amplifications based on the *Pst*I primer PS2 (with 2 selective nucleotides) and *Taq*I primers (with 3 selective nucleotides) produced many easily readable bands in rice compared to amplification based on the *Pst*I primers PS1 and the *Taq*I primer TQ1 (with 1 and 3 selective nucleotides each, respectively), which produced many faint bands that were difficult to score. The primer combinations PS2/TQ2 and PS2/TQ3 produced 98 and 106 bands, respectively. Peak fragment concentration on PAGE gels appeared to be at approximately 700 bp, regardless of the primer pair used. However, because of the density of closely spaced bands in AFLP gels, it was difficult to individually distinguish bands above 750 bp with 4% PAGE gels.

Approximately half of the AFLP bands were monomorphic across all the varieties for the enzyme combination *Pst*I/*Taq*I used in this study. We used a *Pst*I digestion in our AFLP analysis hoping that it would sample the same part of the genome as the RFLP markers used in the previous diversity study of the same varieties (G. Second, personal communication) which were derived from *Pst*I-digested DNA. Two primer combinations (PS2/TQ3 and PS2/TQ2) were required to produce enough polymorphic bands to uniquely distinguish each of the 59 varieties tested. The

Table 2 Correlation of similarity matrices[a] between ISSR motif classes and between 5′- and 3′-anchored, GA- and GT-based primers

		(A)	(B)	(C)
Between Classes				
3′ Di-nt	(A)	–	–	–
5′ Di-nt	(B)	0.801***	–	–
5′ Tri-nt	(C)	0.424***	0.505***	–
5′ Tetra-nt	(D)	0.448***	0.447***	0.414***
Between 5′ *and* 3′				
5′-GA	(A)	–	–	–
3′-GA	(B)	0.733***	–	–
5′-GT	(C)	0.532***	0.526***	–
3′-GT	(D)	0.009	0.000	0.000

Significant correlations at $*P = 0.05$, $**P = 0.01$ and $***P = 0.001$ levels based on the Mantel Test statistic

[a] Similarity matrices based on Jaccard coefficient and all pair-wise comparisons between 59 rice varieties

combination of PS2 and TQ3 primers produced 75 readable bands in the range of 100–600 bp, of which 36 (48.0%) were polymorphic, while the combination of PS2 and TQ2 primers produced 50 readable bands in the same size range, of which 26 (52.0%) were polymorphic. The complete AFLP data set was based on a total of 125 bands, of which 62 were polymorphic.

Diversity analysis and correlation of genetic profiles

The fingerprinting patterns were used to determine genetic similarity matrices, which were then compared to each other. Significant correlations were found between the similarity matrices generated for all four classes of ISSR primers (Table 2). The highest correlation was found between similarity matrices for the combined data sets based on 3′- and 5′-anchored dinucleotide motif primers, while the correlations between similarity matrices derived from di-, tri- or tetra-nucleotide-based primers were lower but still significant. Within the motif classes some individual ISSR primers were more divergent than others. Notably, although the similarity matrices of 3′-anchored poly(GT)-A and poly(GT)-C primers were correlated, they were not significantly correlated with those of the other dinucleotide-based primers. Among the 5′-anchored dinucleotide primers the correlations were significant, while among the individual 5′-anchored tri- and tetra-nucleotide primers they were low or not significant (data not shown). The similarity matrices for the complete AFLP and ISSR data sets were significantly correlated ($r = 0.411$, $P < 0.001$), as were the two primer combinations used in AFLP analysis ($r = 0.324$, $P < 0.001$).

Cluster analyses were used to group varieties and to construct dendrograms based on the ISSR (Fig. 3) and AFLP (Fig. 4) data sets. The Jaccard similarity index scale is consistent between the two figures to facilitate comparison of the estimates of relatedness between varieties. With both methods, none of the variety comparisons yielded 100% similarity. However, the AFLP analysis was less able to discriminate between varieties and showed higher average similarity (73.3%) than the ISSR analysis (55.5%). Both genome scanning methods produced a classification of the germplasm that was comparable to that based on isozymes (Glaszmann 1987) or single-locus RFLP markers (Wang and Tanksley 1989).

The genotypes belonging to either of the two main branches representing the *Indica* (isozyme group I) and *Japonica* (isozyme group V and VI) subspecies of cultivated rice could clearly be differentiated. The improved varieties from IRRI, 'IR8', 'IR24', 'IR36', 'IR42' and 'IR64', were all identified as *Indica* genotypes. As expected, two IRRI sister lines (IR56381A and IR56381B) were very similar. Another cluster below the main *Indica* branch of the ISSR tree corresponded to some of the varieties from India ('BP176', 'CO18', 'DA9', 'JC120', 'JC117' and 'SLO17'). Below the *Japonica* branch, subgroups could be discerned with both genome scanning methods. Varieties belonging to isozyme group V ('Basmati 370', 'Pankhari 203', 'Dom Sofid') formed a distinct cluster that included 'JC157', which has an unknown isozyme status. Three long-grain varieties from the southern USA ('Texas Patna', 'Lemont' and 'Rexmont') clustered together. This group was classified near other tropical *Japonicas* from isozyme group VI, including traditional cultivars from the Philippines ('Azucena', 'Binulawan' and 'Palawan') and from Thailand ('Hawm Om' and 'Dam'). The temperature *Japonicas* from Japan and China ('Aichiasahi', 'Shan Kiu Ju', 'Ta Hung Ku', 'Haifugoya' and 'Taichung 65') also clustered together, but separately from the tropical *Japonicas*. Because of the choice of varieties used in this study, the *Indica* and *Japonica* subspecies appeared equally diverse. An intermediate cluster between the two subspecies consisted of the south Asian varieties 'Aswina', 'Chinsurah Boro II', 'Dular', 'DV85', 'Nakhi', 'Bamoia 341', 'Jhona 349' and 'Fr13a', which have previously been classified as members of isozyme group II and III (Glaszmann 1987).

Discussion

ISSR analysis predicted the frequency and polymorphism of SSR motifs

The number of bands produced by ISSR primers with different di-, tri- and tetra-nucleotide repeat motifs correlated well with what is known about microsatellite frequency in the rice genome. The types of simple sequence repeats found in rice cDNA and genomic libraries have been characterized during the

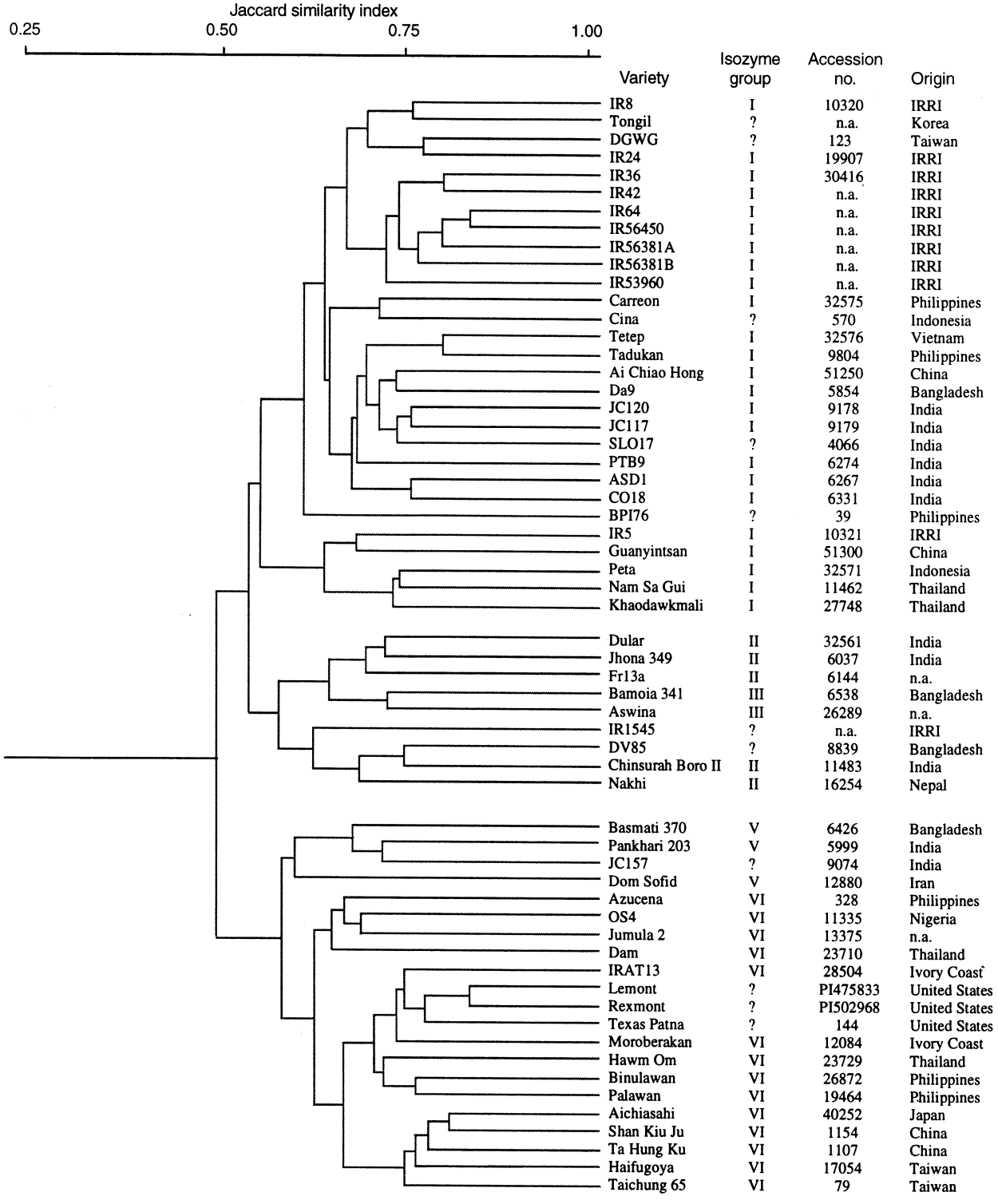

Fig. 3 Dendrogram derived from a UPGMA cluster analysis using the Jaccard's similarity coefficient based on 370 polymorphic ISSR bands showing the associations among 59 rice varieties with their isozyme group designations, accession numbers from the International Rice Germplasm Collection and geographical origin. *n.a.* not available

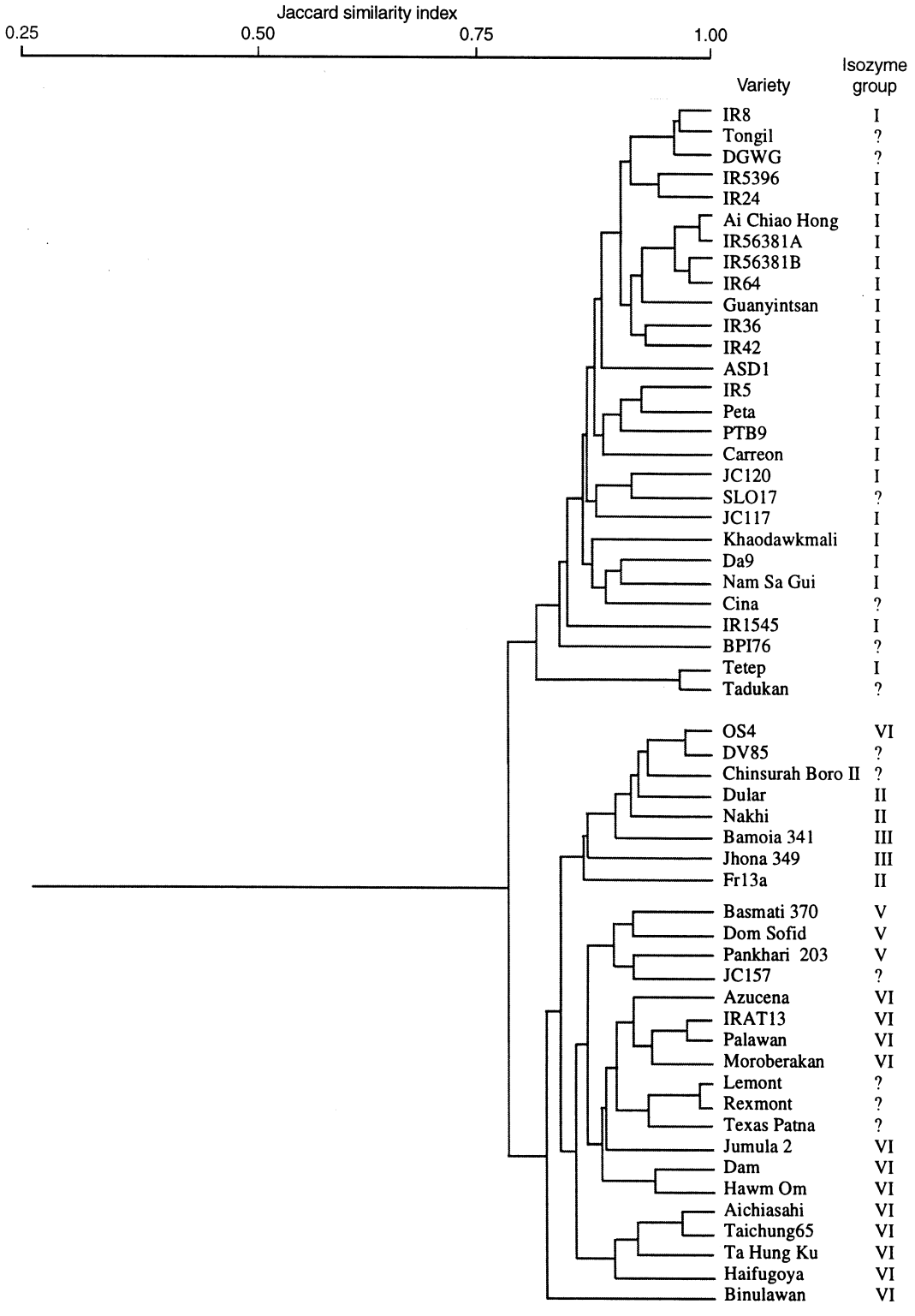

Fig. 4 Dendrogram derived from a UPGMA cluster analysis using the Jaccard's similarity coefficient based on 62 polymorphic bands generated by AFLP analysis showing the associations among 54 rice varieties and their isozyme group designations (I, II, III, V or VI)

construction of framework microsatellite maps (Chen et al. 1997; Panaud et al. 1996; Wu and Tanksley 1993) and the screening of expressed sequences in databases (Akagi et al. 1996a; Y. G. Cho, personal communication). From these studies it is evident that dinucleotide simple sequence repeats are more common in the rice genome than any of simple sequence repeats with larger units. Predictably, dinucleotide motif primers were more amenable to ISSR analysis in rice, while primers based on the more infrequent, trinuclotide and tetranucleotide simple sequence repeats produced fewer bands. Indeed, pentanucleotide repeats were so dispersed or uncommon in the rice genome that no ISSR bands were produced by primers containing this motif. In comparison, dinucleotide motifs made the best ISSR primers in wheat (Nagaoka and Ogihara 1997), regardless of whether the genotype was diploid, tetraploid or hexaploid; while in maize, primers with trinucleotide and tetranucleotide motifs were also valuable for ISSR analysis (Gupta et al. 1994; Kantety et al. 1995).

The poly(GA)-containing 3′-anchored ISSR primers produced on average almost five times as many bands as those with the poly(GT) motif. This difference may reflect the low frequency of the poly(GT) motif in the rice genome or a lack of clustering of these microsatellites. Wu and Tanksley (1993) found a 3:1 greater abundance of GA versus GT microsatellites during hybridization of a 15-kb insert library. They estimated that GA blocks and GT blocks occurred individually every 225 kb and 480 kb, respectively, in the rice genome and that they were rarely clustered together. Recent results based on the screening of a small insert genomic library supports the conclusion that poly(GT) is significantly less abundant than poly(GA) in rice (S. Temnykh, personal communication). Many other plant species also have a greater abundance of microsatellites with the GA motif compared to the GT motif (Depeiges et al. 1995; Morgante and Olivieri 1993). This can explain why in wheat poly(GA)-based ISSR primers were more useful and produced twice as many bands as poly(GT) primers (Nagaoka and Ogihara 1997).

The 5′-anchored poly(GT) primer produced more bands than the 3′-anchored poly(GT) primer. This might be expected because the 5′-anchored primer lacks selective nucleotides at the critical 3′ end. However, this was not the case for the poly(GA) primers, which produced equivalent numbers of bands using 5′- or 3′-anchored primers. One explanation for these anomalous results is that PCR amplification with 5′-anchored primers imposes selection for long simple sequence repeats that bind along the entire length of the primer ($\geq$ 9 repeat units, or 18 bp), while amplification with 3′-anchored primers would not impose selection for repeat length. Therefore, the use of 3′-anchored primers, while providing an overview of the overall frequency of a motif, may detect a higher frequency of short simple sequence repeats than 5′-anchored primers. The average number of repeating units is thought to be lower in poly(GT) simple sequence repeats than in poly(GA) simple sequence repeats (S. Temnykh, personal communication). This would explain why the poly(GA) and poly(GT) primers produced similar numbers of bands when anchored at the 5′ end but different numbers of bands when anchored at the 3′ end. It would be interesting to study whether 3′-anchored ISSR primers based on complementary reverse sequences, such as poly(GA) and poly(CT) or poly(GT) and poly(CA), also detected these differences.

Poly(AT) dinucleotide repeats are suggested to be very frequent in rice (Akagi et al. 1997) and are thought to be among the most abundant motifs in plant species (Depeiges et al. 1995; Morgante and Olivieri 1993; Lagercrantz et al. 1993). However, neither poly(AT) nor poly(GC) microsatellites were targeted in this study because ISSR primers based on these motifs are self-annealing, due to sequence complementarity, and would form dimers during PCR amplification.

Among the trinucleotide-based ISSR primers, those containing the following repeat motifs, TGG, ATC, TTG, CGG, GAG, TCT, CAG and ATC, produced sequentially fewer bands in rice. The ISSR amplification based on the poly(CGG) motif was not as productive as might have been expected from the results of Zhao and Kochert (1992), which suggested that this motif was common in the rice genome. In a study by Parsons et al. (1997), in which ISSR primers were used with poly(AAG), poly(CGG) and poly(CAG) trinucleotide motifs anchored at the 3′ end by 2 selective nucleotides produced 5–11 bands in rice, similar to results reported here based on 1 selective nucleotide. The relative abundance of trinucleotide motifs appears to be different in rice than in other plant species (Depeiges et al. 1995; Kantety et al. 1995; Nagaoka and Ogihara 1997).

The ten tetranuclotide motif ISSR primers tried were very variable in their ability to amplify bands: those with poly(GATG), poly(AGTG) and poly(TCTG) motifs produced many bands, those with poly(GAGG) and poly(TTCG) produced few bands, and those with the poly(GACA) motif or the A-T rich motifs (TTAG, TATC, TATG or GATA) produced no bands. We were interested in some of these tetranucleotide repeats because of reports of their high frequency and clustered distribution in peas (Lu et al. 1996) maize (Gupta et al. 1994) and rice (Ramakishana et al. 1994). However, Southern hybridization experiments (unpublished work, this laboratory) support the view that these motifs are uncommon in the rice genome.

In the case of infrequent SSR motifs, the conditions used for PCR amplification, such as the duration of the extension step, can be modified to favor the production of larger ISSR bands (Parsons et al. 1997). However longer PCR products can only be analyzed on agarose gels stained with ethidium bromide, a system which has both lower resolving power and DNA sensitivity than

the silver-stained polyacrylamide gels used in this and other ISSR studies (Charters et al. 1996; Fang and Roose 1997).

Trinucleotide repeat ISSR primers as a class produced the lowest rates of polymorphism and the highest similarity index among all the rice varieties analyzed. In contrast, ISSR analysis with dinucleotide-based primers provided the highest rates of polymorphism observed for any multi-locus marker system in rice. In maize, on the other hand, trinucleotide ISSR primers based on poly(GCT) and poly(AGC) motifs revealed more polymorphisms than dinucleotide primers based on poly(CA) or poly(GT) motifs (Kantety et al. 1995). Although all the dinucleotide motif ISSR primers produced high levels of polymorphism, some variability was observed. The 5′-anchored dinucleotide-based ISSR primers detected more monomorphic bands than the 3′-anchored primers. Among the 5′-anchored dinucleotide-based primers the poly(GT) and poly(CA) motifs detected lower rates of polymorphism than the poly(GA) and poly(CT) motifs. These results agree with diversity surveys conducted with microsatellites containing these motifs in which poly(GT) loci tend to have 30–50% fewer alleles than poly(GA) loci (Y. G. Cho personal communication). Since rice microsatellite loci are often made up of compound perfect repeats of two or more different motifs (Chen et al. 1997) which can vary in size independently for each different allele (Panaud et al. 1996), changes in the composition or size of compound simple sequence repeats could be an important source of ISSR polymorphisms. The polymorphism at ISSR loci detected with 5′-anchored primers can arise from the elimination of a flanking simple sequence repeat that leads to the loss of a priming site, the expansion and contraction of the flanking simple sequence repeat or insertion/deletions between the priming sites that lead to a change in the size of the ISSR allele (Zietkiewicz et al. 1994). Therefore, the lower rate of polymorphism at ISSR loci detected with the poly(GT) motif might result from their location in more static regions of the genome than ISSR loci detected with the poly(GA) motif.

Patterns of rice diversity and predictions of variety relationships

The patterns of diversity revealed by the different types of ISSR primers were similar in the majority of comparisons as indicated by the significant correlations between most similarity matrices. This means that most of the ISSR primers, especially members of the 5′-anchored dinucleotide class, could be freely interchanged with no loss in the power to distinguish varietal diversity. However, the similarity matrices of ISSR primers that produced fewer than 10 polymorphic bands were not correlated with those of the other, more productive ISSR primers. Therefore, while only one 5′-anchored dinucleotide based primer was sufficient to differentiate among and classify the rice varieties, several of the tri- and tetra-nucleotide-based ISSR primers would need to be used in combination to achieve the same precision for the cluster analysis of rice varieties. These results suggest that the use of various ISSR primers would enable the assay of much of the genetic diversity of rice.

The genome scanning techniques used in this study proved useful for the classification of rice accessions. The cluster analysis performed for the ISSR and AFLP data sets easily differentiated the isozyme groups determined by Glaszmann (1987). In comparison, RFLP or random amplified polymorphic DNA (RAPD) markers can distinguish *Indica* and *Japonica* subspecies of cultivated rice but detect only modest amounts of genetic variability within these gene pools (Ishii et al. 1995; Mackill 1995; Wang and Tanksley 1989; Yu and Nguyen 1994; Zhang et al. 1992). In this study, the ability to distinguish between closely related cultivars within a subspecies was greater for ISSR than for AFLP analysis.

The overall predictions of rice variety relationships generated from the fingerprinting with either the ISSR or the AFLP techniques were similar, as indicated by the correlations between the genetic similarity matrices. However, the cluster representing isozyme groups II and III was more closely identified with *Japonica* varieties in the AFLP analysis and with the *Indica* varieties in the ISSR analysis. When cluster analysis was performed for each of the classes of ISSR primers, the dinucleotide motif ISSR loci identified isozyme group II varieties as an *Indica* sub-cluster, while the trinucleotide ISSR loci identified this group as a *Japonica* sub-cluster. Parsons et al. (1997) found a similar divergence for isozyme group II when using centromeric versus distal markers in their cluster analysis. If different markers were distributed unevenly, it could explain the split identity of isozyme group II in this study.

Members of isozyme group II were a mixture of traditional landraces and improved lines from the Indian subcontinent that have been categorized as "Aus" and "Boro" rices; these are considered to be part of the *Indica* subspecies (Glaszmann 1987). This isozyme group, along with group III (Deepwater rices) and V (Basmati rices) from South Asia, exhibit many distinguishing characteristics, including unusual grain quality traits, deepwater growth habit, resistance genes with novel biotype or race specificities and tolerances to hot or cold weather. These unique traits, as well as the molecular differences commented upon in this work, suggest that these rice groups may represent mixtures of *Indica* and *Japonica* genomes possibly highlighted with introgressions from wild rice species. ISSR markers diagnostic for these varietal differences could be used to facilitate the deployment of these traits in rice breeding.

In this study, the ISSR technique was successfully applied to two important aspects of genetic analysis. First, it was shown that this technique can be used to study the frequency of simple sequence repeats in a given genome, without the time-consuming and costly steps of library construction, DNA hybridization, subcloning or sequencing typical of previous methods of estimating frequency. Data about frequency are crucial for deciding which microsatellite motifs to target for marker development (Panaud et al. 1996). Second, ISSR fingerprinting was found to be a rapid, efficient and cost-effective, whole-genome scanning technique for rice that detected more polymorphisms than the AFLP technique and had the added advantages of requiring only one primer and a single PCR amplification step to carry out. ISSR fingerprinting of rice genotypes could be used to characterize the large numbers of rice accessions held in national and international germplasm centers, to prioritize the collection and conservation of diversity and to manage breeding populations effectively.

Acknowledgements Thanks to G. Second for providing DNA samples for the 59 rice genotypes, to M. Sorrells and S. Temnykh for critical reading of the manuscript and to Carole Morehouse for help with formatting. This work was supported by a Plant Science Center (USDA-DOE-NSF-Cornell Biotechnology Program) fellowship to MWB.

References

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

Akagi H, Yokozeki Y, Inagaki A, Nakamura A, Fujimura T (1996b) A co-dominant DNA marker closely linked to the rice nuclear restorer gene, *Rf-1*, identified with inter-SSR fingerprinting. Genome 39:1205–1209

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1997) Highly polymorphic microsatellites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor Appl Genet 94:61–67

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

Charters YM, Robertson A, Wilkinson MJ, Ramsay G (1996) PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. *oleifera*) using 5′ anchored simple sequence repeat (SSR) primers. Theor Appl Genet 92:442–447

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

Cho YG, Blair MW, Panaud O, McCouch SR (1996) Cloning and mapping of variety-specific rice genomic DNA sequences: amplified fragment length polymorphisms (AFLP) from silver stained polyacrylamide gels. Genome 39:373–378

Cho YG, McCouch SR, Kuiper SR, Pot J, Kang MR, Groenen JTM, Eun MY (1997) Integrated map of AFLP, SSLP and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). Theor Appl Genet 97:370–380

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

Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable (GA/CT)n microsatellites in burr oak, *Quercus macrocarpa*. Theor Appl Genet 91:137–141

Fang DQ, Roose ML (1997) Identification of closely related citrus cultivars with inter-simple sequence repeat markers. Theor Appl Genet 95:408–417

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

Goodwin ID, Aitken EAB, Smith LW (1997) Application of inter simple sequence repeat (ISSR) markers to plant genetics. Electrophoresis 18:1524–1528

Gupta M, Chyi YS, Romero-Severson J, Owen JL (1994) Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. Theor Appl Genet 89:998–1006

Ishii T, Brar DS, Second G, Khush GS (1995) Nuclear genome differentiation in Asian cultivated rice as revealed by RFLP analysis. Jpn J Genet 70:643–652

Jackson MT (1997) Conservation of rice genetic resources: the role of the International Rice Genebank at IRRI. Plant Mol Biol 35:61–67

Kantety RV, Zeng X, Bennetzen J, Zehr BE (1995) Assessment of genetic diversity in dent and popcorn (*Zea mays* L.) inbred lines using inter-simple sequence repeat (ISSR) amplification. Mol Breed 1:365–373

Lagercrantz U, Ellegren H, Kakanuga T (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucelic Acids Res 21:1111–1115

Lu J, Knox MR, Ambrose MJ, Brown JKM, Ellis THN (1996) Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. Theor Appl Genet 93:1103–1111

Mackill DJ (1995) Classifying *Japonica* rice cultivars with RAPD markers. Crop Sci 35:889–894

Mackill DJ, Zhang Z, Redona ED, Colowit PM (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. Genome 39:969–977

Maheswaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC, Huang N (1997) Polymorphism, distribution and segregation of AFLP markers in a doubled haploid rice population. Theor Appl Genet 94:39–45

Manly BFJ (1997) Randomization, bootstrap and montecarlo methods in biology, 2nd edn. Chapman and Hall, New York, pp 172–201

McConaughy BL, Laird CD, McCarthy BJ (1969) Nucleic acid reassociation in formamide. Biochemistry 8:3289–3294

Morgante M, Olivieri AM (1993) PCR-amplified microsatellite markers in plant genetics. Plant J 3:175–182

Nagaoka T, Ogihara Y (1997) Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. Theor Appl Genet 94:597–602

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

Parsons BJ, Newbury HJ, Jackson MT, Ford-Lloyd BV (1997) Contrasting genetic diversity relationships are revealed in rice (*Oryza sativa* L.) using different marker types. Mol Breed 3:115–125

Ramakishana W, Lagu MD, Gupta VS, Ranjekar PK (1994) DNA fingerprinting in rice using oligonucleotide probes specific from simple repetitive DNA sequences. Theor Appl Genet 88:402–406

Rohlf FJ (1992) NTSYS-PC: Numerical taxonomy and multivariate analysis system version 1.70. State University of New York, Stony Brook N.Y.

Salimath SS, De-Oliviera AC, Godwin ID, Bennetzen JL (1995) Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. Genome 38 : 757–763

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23 : 4407–4414

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

Yang W, De-Oliveira AC, Godwin I, Schertz K, Bennetzen JL (1996) Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums. Crop Sci 36 : 1669–1676

Yu LX, Nguyen HT (1994) Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa* L.). Theor Appl Genet 87 : 668–672

Zhang QF, Maroof MAS, Lu TY, Shen BZ (1992) Genetic diversity and differentiation of *Indica* and *Japonica* rice detected by RFLP analysis. Theor Appl Genet 83 : 495–499

Zhao X, Kochert G (1992) Clusters of interspersed repeated DNA sequences in the rice genome (*Oryza*). Genome 36 : 944–953

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20 : 176–183