Molecular Breeding **3:** 115–125, 1997.

Contrasting genetic diversity relationships are revealed in rice (*Oryza sativa* L.) using different marker types

Beverley J. Parsons[1], H. John Newbury[1,*], Michael T. Jackson[2] and Brian V. Ford-Lloyd[1]
[1]*School of Biological Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK (*author for correspondence; fax UK+121-414-5925);* [2]*International Rice Research Institute, PO Box 933, os Baños, Manila, Philippines*

Received 21 July 1995; accepted 22 October 1996

Key words: diversity, genetic mapping, *Oryza sativa*, RAPDs, rice, inter-simple sequence repeats

Abstract

Genetic variation between samples of *Oryza sativa* from 19 localities in Bangladesh and Bhutan was assessed using two PCR-based molecular marker systems: RAPD (random amplification of polymorphic DNA) and ISSR-PCR (inter-simple sequence repeat polymerase chain reaction). Employing RAPD, a set of 14 decanucleotides of arbitrary sequence directed the amplification of 94 reproducible marker bands, 47 (50%) of which were polymorphic. In addition, a set of 9 ISSR primers were used to direct amplification of 71 PCR products, 40 (56%) of which were polymorphic. Multivariate analyses of the two PCR-based molecular marker data sets provided evidence that the patterns of variation correspond with the classification described by Glaszmann [9] using isozyme analysis. Subtle differences in the relationships revealed between rice groups using the two types of PCR-based marker led to investigations of their map positions using an intraspecific doubled haploid mapping population. The observation that the chromosomal locations of markers can influence diversity assessments is presented and the significance of this is discussed.

Introduction

Rice is the world's most important staple crop, and more than half the world's population, mostly in Asia, depends upon it directly for food. It has also become a model organism for genome analysis, having a diploid chromosome number of 24 and the smallest genome size of all major crop plants of 430 Mb [1]. Causse *et al.* [3] and Kurata *et al.* [17] have independently developed saturated RFLP maps of *Oryza*, resulting in a combined total of over 2000 mapped markers. Rice is now one of the most densely mapped crop species with an average DNA to cM ratio of 250–300 kb/cM. Recently, it has been possible to use gene dosage analysis of 211 RFLP markers to reveal the chromosomal locations of all 12 rice centromeres [26].

Oryza sativa is reported to be one of the most polymorphic crop species based on molecular data [38]. Variation within *O. sativa* is extensive, due to its adaptation to a wide range of geographical and ecological niches and climatic regimes. Ecogeographic differentiation of Asiatic rice has produced a number of types, initially recognised on the basis of their morphology, referred to as *indica*, *japonica* and *javanica* [4]. Rice breeders exploit the high degree of heterosis that results from crosses between these types, leading to the development of improved varieties [32, 40]. Glaszmann [9, 10] developed a valuable classification system, based on 15 polymorphic isozyme loci encoding 8 enzyme activities, for resolving rice germplasm into 6 varietal groups, characterised by opposing multilocus types, with varying environmental and macrogeographic distributions. This method of classification is highly respected by rice breeders and shows clear concordance with previous classifications. Due to difficulties in hybridisation and recombination between *japonica* and *indica* rices, breeders currently rely on isozyme classification to aid the selection of suitable parents to cross. Since the discovery of wide compatibility varieties (WCVs) which yield fertile F1 off-

spring when crossed to either *indica* or *japonica* varieties, heterosis resulting from these crosses is becoming increasingly exploited in rice [42]. The development of techniques which allow fast, reliable and precise classification of potential breeding material is therefore essential.

Using RFLP analysis, *indica* rice has been shown to be genetically more diverse than *japonica* rice, as a result of extensive genetic differentiation between major portions of the genome [41]. After the introduction of the polymerase chain reaction [23], numerous alternative DNA marker systems have been developed, including an array of multiple arbitrary amplicon profiling (MAAP) approaches [2], such as random amplification of polymorphic DNA (RAPD) [34, 35], which have proven valuable in rice genetic diversity studies [7, 20, 39]. Virk *et al.* [29] employed RAPD technology as an efficient means of determining levels of diversity within a subset of germplasm sampled from the rice collection maintained in the International Rice Gene Bank at IRRI, and also for the designation of duplicates within this collection [30].

It has been suggested that the information value of MAAP analyses can be enhanced if, rather than utilising an arbitrary sequence, the primer instead recognises interspersed repetitive elements which are found ubiquitously in eukaryotic genomes. Simple sequence repeats (SSRs) can be defined as relatively short runs of tandemly repeated DNA, whose repeat length is not greater than 6 bp [27]. Using oligonucleotide probes specific for SSRs, the abundance and level of polymorphism of these sequences has been investigated with the objective of identifying rice genotypes [21]. SSRs are highly variable in terms of the number of repeated units per locus, but they are relatively stable with respect to their genomic localization [19, 33], which is considered random, with no positional bias or clustering to specific chromosomal regions [36].

Zietkiewicz *et al.* [43] have developed an alternative system which encompasses both the broad taxonomical applicability of RAPDs, where no prior knowledge of the analysed genome is required, and has the advantage of simultaneously targetting multiple genomic SSR loci. The approach employs oligonucleotides based on a SSR, anchored at either the 5′ or 3′ end with 2–4 purine or pyrimidine residues, to initiate PCR amplification of genomic segments flanked by inversely orientated, closely spaced repeated sequences. Inter-simple sequence repeats (ISSR) primers anchored at their 3′ ends have been exploited for the classification of popcorn and dent corn lines into distinct heterotic gene pools [15] and for assessing genetic diversity and genome origins in the genus *Eleusine* [24].

In our study, the levels of genetic diversity have been determined between 19 rice cultivars sampled from two regions with distinct differences in their geography and climate: Bangladesh (generally tropical) and Bhutan (generally temperate). The rice material encompasses four of the six isozyme groups classified by Glaszmann [9], namely groups I, II, V and VI. Group I corresponds to *indica* rice and is predominant in tropical and subtropical regions that comprise the major portion of rice-growing areas of the world. Group II encompasses rices observed exclusively in south and west Asia and are especially frequent along the Himalayas. Group V is a highly polymorphic group that includes a diversity of high-quality rice varieties spanning the geographic range from Iran to Myanmar. Group VI includes both *japonica* and *javanica* rices with the former being found in temperate areas and high elevations in south and southeast Asia.

Diversity within this rice material has been assessed using both RAPD and ISSR-PCR and the patterns of relationship revealed have been compared. Following the diversity analyses, we have investigated the chromosomal map locations of sets of molecular markers. A relationship has been observed between the location of certain markers, the distribution of these markers among the 19 rice accessions, and hence the influence that they have on the subsequent classification of rices. The importance of molecular marker choice in diversity assessments and varietal classification is discussed.

Materials and Methods

Diversity survey

Plant material

An ecogeographically diverse set of 19 germplasm accessions of *O. sativa*, comprising 14 from Bangladesh and 5 from Bhutan, were obtained from the International Rice Gene Bank at IRRI. Their origins are shown in Table 1. Plants were grown in a glasshouse at Birmingham, maintained at 27 °C, with a 17 h daylength and 80% relative humidity for 2 weeks. Two mg of leaf material were sampled from each of 10 seedlings, yielding 20 mg of material for DNA isolation. The DNA extraction method adopted was a scaled

down version of a 4% CTAB procedure [8] as described by Virk *et al.* [29]. On average, 1.65 μg of DNA was obtained from 20 mg of fresh leaf tissue. After agarose gel electrophoresis, DNA concentration was calibrated on the basis of ethidium bromide fluorescence using defined amounts of DNA as a standard [25]. DNA was diluted to a final concentration of 2 ng/μl in sterile distilled water.

RAPD protocol

A set of 14 decanucleotide primers (Operon C, F and K kit; see Table 2) of arbitrary sequence were used to initiate amplification of PCR products. The reaction mix, with a total volume of 25 μl, contained ca. 5 ng DNA, 0.4 μM of each decanucleotide, 200 μM of each dNTP (Pharmacia) and 1 U *Taq* polymerase, maintained in a 1$\times$ incubation buffer (Boehringer Mannheim), containing a total of 2.5 mM magnesium chloride (Laboratory Services). Samples were covered with 35 μl of mineral oil and processed, using a Hybaid Omnigene thermocycler, through 47 simulated tube-controlled temperature cycles as follows: cycle 1, 2 min at 95 °C; cycles 2–3, 30 s at 95 °C, 1 min at 37 °C and 2 min at 72 °C; cycles 4–5, 30 s at 95 °C, 1 min at 35 °C and 2 min at 72 °C; cycles 6–46, 30 s at 94 °C, 1 min at 35 °C and 2 min at 72 °C; cycle 47, 5 min at 72 °C. On completion of the thermal cycling, 4 μl of gel loading buffer was mixed with the 25 μl of reaction mix and 10 μl aliquots of PCR amplification products were subjected to 1.2% agarose gel electrophoresis in 1$\times$ TBE buffer [25] and were detected by ethidium bromide staining, viewed by fluorescence under UV light. A 1 kb DNA ladder (Gibco-BRL) was positioned on either side of the gel as size standards for MW estimation of PCR products. Gels were photographed using a digital imaging system (Flowgen). All reactions were performed at least twice to check consistency.

ISSR-PCR

A range of primers, namely $(GGC)_5$, $TA(GGC)_5$, $GTAC(GGC)_5$, $(GGC)_5AT$ and $(GGC)_5ATG$, were initially synthesized (AltaBioscience, University of Birmingham) to determine the optimum design for amplification of ISSR sequences. Nine subsequent oligonucleotides were synthesized based on the three core repeats, $(GGC)_5$, $(AGC)_5$ and $(AAG)_5$, anchored at their 3′ end by a selection of dinucleotides (see Table 2), and these were used to screen the diverse set of rice accessions. The specificities of all anchored SSR oligonucleotides were optimised in a total volume of 25 μl by titrating against primer concentration (varying from 0.2 to 0.8 μM), template DNA concentration (ranging from 1 to 10 ng), concentration of magnesium ions across a range of 1.5–3.5 mM, and varying annealing temperature and thermal cycling procedures. Optimal conditions were selected on the basis of the maximum number of reproducible and easily resolvable PCR products initiated by each primer. The thermal profile applied was a slightly adjusted version of the RAPD protocol employed previously. Annealing temperatures varied for each primer relative to the base composition and 35 amplification cycles were performed. Amplification products were subjected to 1.4% agarose gel electrophoresis in 0.5$\times$ TBE buffer [25]. Staining of amplified fragments and image capture were as described for RAPD. Consistency of ISSR-PCR products was checked by performing all reactions at least twice.

Data analysis

Specific PCR products that were reproducible in successive amplifications were selected and marker bands were defined by their molecular weights, estimated from the size standards. Amplification products were scored as present (1) or absent (0) for all 19 accessions. Indices of similarity were calculated using Jaccard's coefficient, to estimate relationships between accessions. An agglomerative technique utilising the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) computed using NTSYS-pc was used to group accessions [22]. In addition, Two-Way Indicator Species Analysis (TWINSPAN), based on a divisive method of clustering, was employed to analyse the data generated by RAPD and ISSR-PCR markers [13]. The relationships between the 19 accessions were portrayed graphically in the form of dendrograms.

Mapping analysis

Plant material

A doubled haploid population of 135 plants was derived from the cross IR64 (an improved *indica* variety) $\times$ Azucena (a traditional *japonica* variety). Development of this population is described by Huang *et al.* [14] who supplied a subset of 60 doubled haploid individuals used in this investigation. Plant growth conditions and the DNA extraction protocol was as described above.

Table 1. Oryza sativa accessions used in this study.

Arbitrary reference number	IRGC accession number	Variety name	Isozyme group	Origin
1	IRGC 26523	Shoru	V	Dacca
2	IRGC 37362	Badali	II	Chittagong
3	IRGC 64767	Bholanath	I	Mymensingh
4	IRGC 66766	Aus Meghi	II	Jessore
5	IRGC 66770	Bamura	I	Sylhet
6	IRGC 66794	Hizuli	II	Chaudanga
7	IRGC 66821	Pathor Bhati	II	Dinajpur
8	IRGC 77225	Boya	I	Kurigram
9	IRGC 77232	Danga Boro	I	Panchagarh
10	IRGC 77241	Godai	I	Rajbari
11	IRGC 77286	Paizra	I	Laxmipur
12	IRGC 77303	Til Kapor	V	Natore
13	IRGC 27510	Bhaturi	II	Khulna
14	IRGC 77231	Dabbor	I	Comilla
15*	IRGC 62155	Aulum	VI	Tongsa
16*	IRGC 62161	Bondilip	VI	Wandipodang
17*	IRGC 62167	Dagpazamo	I	Lhuntshi
18*	IRGC 64891	Dumza	VI	Paro
19*	IRGC 67836	Kali Dhan	VI	Chirang

[1] Reference numbers (1–19) listed here are those used in all subsequent figures. Fourteen accessions originate from Bangladesh; the five indicated with an asterisk originate from Bhutan. IRGC = International Rice Gene Bank Collection; this is a unique number for each accession held within the germplasm collection at IRRI.

DNA markers

A total of 135 RFLP loci have previously been assigned to a framework map currently being used for QTL mapping (Huang, pers. commun.). To establish the extent of parental polymorphism revealed using PCR-based markers, IR64 and Azucena were screened using 12 RAPD and 8 ISSR primers which were previously employed in the diversity survey. Polymorphic markers were then scored across the mapping population of 60 doubled haploid individuals. Five ISSR markers AM5-900, AM6-980, AM7-500, AM9-2700 and AM9-980 were used as probes in Southern analyses to confirm the homology of bands co-migrating in the diverse material and the mapping population.

Construction of linkage map

Two point linkage analysis was performed using Mapmaker/EXP(3.0) [18] to determine the recombination fraction (ϑ) and the significance of linkage using LOD score. Pairs of linked markers were identified with LOD >4.0 and $\vartheta < 0.4$. PCR-based markers were integrated into the RFLP framework using three point analysis with a LOD score >2.0, with additional markers being assigned to intervals using the 'try' command of the Mapmaker program.

Results

In order to examine the reliability of the RAPD protocol on the rice test material, amplifications were carried out using template DNA obtained following three independent extractions from individuals of each of three accessions of rice. The results, obtained using two primers (data not shown), demonstrated that although some minor irreproducible bands were amplified, by using those PCR products which are consistent in successive PCR amplifications, RAPD bands can act as reliable markers in assessing variation in the genotypes of rice accessions. In all subsequent RAPD work, DNA was obtained from 10 pooled individuals per accession because it has been shown [29] that this method of sampling provides a better representation of an accession than using DNA from a single plant only.

Table 2. RAPD and ISSR-PCR primer codes, sequences and optimized annealing temperatures. The numbers of monomorphic and polymorphic fragments amplified using each primer across a range of 19 rice accessions (see Table 1) are indicated.

Primer code	Sequence (5′-3′)	Annealing temperature (°C)	Number of polymorphic bands	Number of monomorphic bands
RAPD				
OPK - 01	CATTCGAGCC	35	2	5
OPK - 02	GTCTCCGCAA	35	3	1
OPK - 04	CCGCCCAAAC	35	4	5
OPK - 06	CACCTTTCCC	35	4	5
OPK - 07	AGCGAGCAAG	35	3	3
OPK - 08	GAACACTGGG	35	5	2
OPK - 10	GTGCAACGTG	35	4	3
OPK - 11	AATGCCCCAG	35	3	4
OPK - 12	TGGCCCTCAC	35	3	2
OPK - 15	CTCCTGCCAA	35	4	5
OPK - 16	GAGCGTCGAA	35	3	0
OPK - 17	CCCAGCTGTG	35	4	2
OPF - 13	GGCTGCAGAA	35	2	6
OPC - 14	TGCGTGCTTG	35	3	4
ISSR-PCR				
AM-1	$(GGC)_5AT$	58	2	3
AM-2	$(AAG)_5GC$	38	5	5
AM-3	$(AAG)_5TG$	38	2	3
AM-4	$(AAG)_5CC$	40	7	4
AM-5	$(AGC)_5CA$	57	2	6
AM-6	$(AGC)_5GG$	51	6	5
AM-7	$(GGC)_5TA$	60	2	4
AM-8	$(AGC)_5GA$	53	3	7
AM-9	$(AAG)_5CG$	40	2	3

Polymorphisms were revealed in the banding patterns across the set of 19 accessions. Using a total of 14 decanucleotide primers, 94 marker bands were identified of which 47 (50%) were polymorphic. Table 2 lists the codes and sequences of the primers used to generate these 94 products, summarising the total number of monomorphic and polymorphic fragments amplified. Bands were scored as present or absent for each accession and the data were subjected to multivariate analysis. Figure 1 is a dendrogram generated by UPGMA revealing patterns of relatedness of the 19 cultivars. There is a clear correspondence with the enzymatic classification described by Glaszmann [9], with isozyme groups I, II, V and VI clearly identifiable.

In a study using ISSR-PCR, preliminary primer design experiments were performed using the $(GGC)_5$ SSR core. 5′-anchored primers and core repeats with no anchoring nucleotides generated smeared, indistinct amplification profiles. However, clearly resolv-

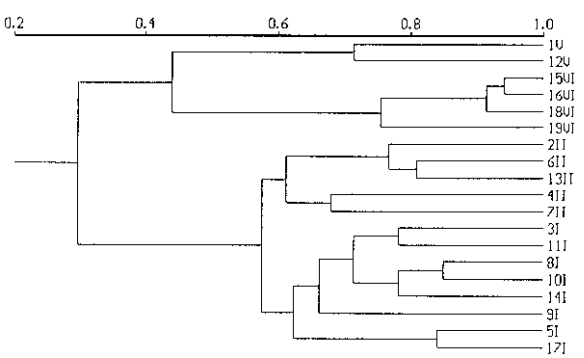

Figure 1. Dendrogram generated by UPGMA cluster analysis [22] of 1–F values based on Jaccard's coefficient determined using 47 RAPD bands showing clustering of 19 rice accessions into corresponding isozyme groups (I, II, V and VI). Accessions 1–14 originate from Bangladesh; accessions 15–19 originate from Bhutan.

able fragments were amplified using the 3′-anchored $(GGC)_5$ primer, so additional ISSR primers were synthesized based on this design, using the core repeats $(GGC)_5$, $(AGC)_5$ and $(AAG)_5$ with two anchoring nucleotides. In an attempt to further enhance amplification stringency, the effect of using longer 3′-anchored primers was investigated by synthesising oligonucleotides with seven core repeats instead of five. No differences were observed between the amplification profiles generated with primers of five or seven core repeats and the same 3′-anchoring nucleotides. Template and magnesium ions were used in the same concentrations as for RAPD and for the majority of oligonucleotides a primer concentration of 0.4 μM was optimal. In contrast to RAPD primers, which have a G+C content of 60–70%, the ISSR primers employed in this study spanned a greater range with G+C contents ranging from 35–88%. As a consequence, careful optimisation of primer annealing temperature was required in order to generate reproducible amplification profiles. The optimal annealing temperature for each ISSR primer are listed in Table 2.

Figure 2a illustrates the amplification profiles generated across the 19 accessions using primer $(GGC)_5$TA. Nine primers were used to initiate amplification of 40 (56%) polymorphic and 31 monomorphic products, ranging in size from 340 to 3080 bp. Amplification products were scored as present or absent and subjected to the same multivariate analyses as before. It is clear from Fig. 2a that, in addition to the four markers indicated, there are some polymorphic markers which have been ignored. A conservative approach to marker scoring was employed to minimise the chance of misscoring markers and any bands which were unstable in multiple PCR runs were excluded from the analysis. Very closely migrating markers which could not be accurately resolved by agarose gel electrophoresis were also not scored. The relationships among the 19 cultivars are shown in Figure 3. Once again, accessions resolve into 4 main clusters, corresponding to Glaszmann's isozyme groups I, II, V and VI.

Both RAPD and ISSR markers reliably differentiate rice accessions into their respective isozyme groups. However, whereas the RAPD classification, in agreement with Glaszmann's isozyme classification [9], identifies a close genetic relationship between group I (*indica*) and group II accessions, the classification generated from ISSR data indicates that group II is more closely related to groups V and VI (*japonica*). This important difference in the relationship of group II accessions to other rice groups was independently

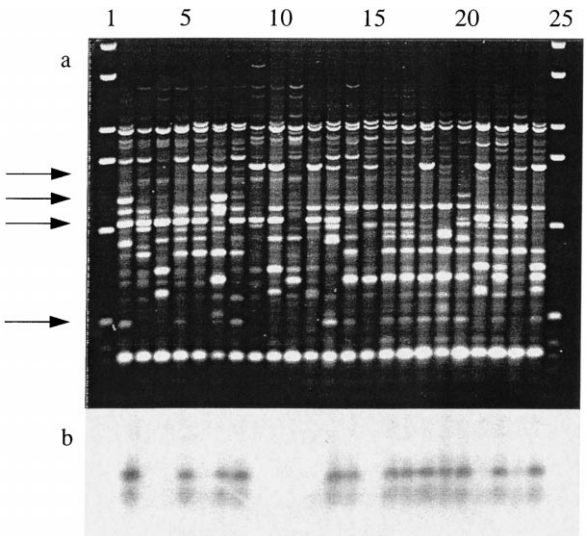

Figure 2. **a.** Agarose gel electrophoresis of ISSR-PCR products amplified using primer AM7 ($(GGC)_5$TA). Lanes 1 and 25 contain 1 kb ladder size standards. Lanes 2–20 contain DNA amplified from rice accessions 1–19 (see Table 1). Lanes 21–24 contain DNA amplified from IR64, Azucena and two representatives of the mapping population respectively. Scored polymorphic markers are indicated with arrows. *b.* ISSR fragment AM7-500 (lowest arrow) was used as a hybridisation probe on a Southern blot of the above gel illustrating the homology of co-migrating markers in both the diverse material and material from the mapping population.

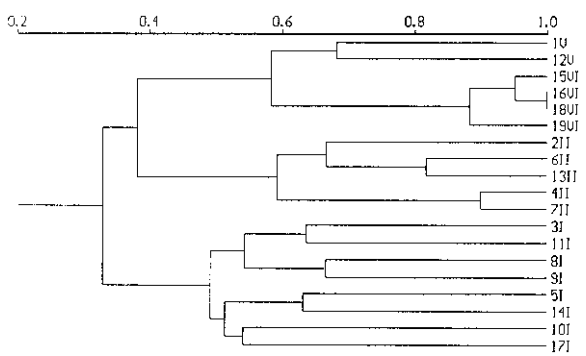

Figure 3. Dendrogram generated by UPGMA cluster analysis [22] of 1–F values based on Jaccard's coefficient determined using 40 ISSR-PCR bands showing clustering of 19 rice accessions (1–19) into corresponding isozyme groups (I, II, V and VI). Accessions 1–14 originate from Bangladesh; accessions 15–19 originate from Bhutan.

observed in an analysis of 48 different rice accessions using RAPD and ISSR-PCR (Virk, pers. commun.). In order to determine whether this difference in classification is related to the chromosomal location of the marker loci, mapping was carried out using a set of

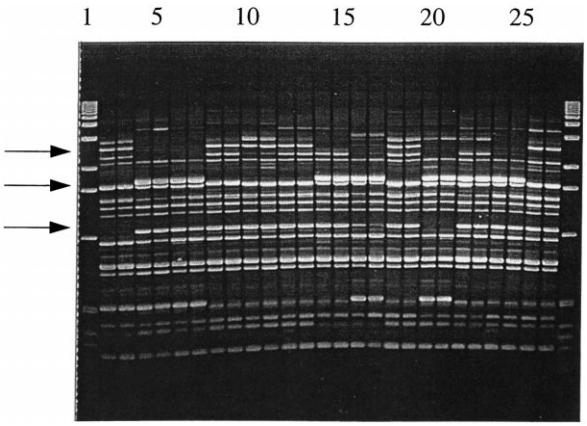

Figure 4. Agarose gel electrophoresis of RAPD PCR products amplified using primer OPK-08. Molecular weights of size standards shown in lanes 1 and 28 are indicated in base pairs. Lanes 2–27 contain duplicate amplifications of 13 doubled haploid progeny. Scored polymorphic markers are indicated with arrows.

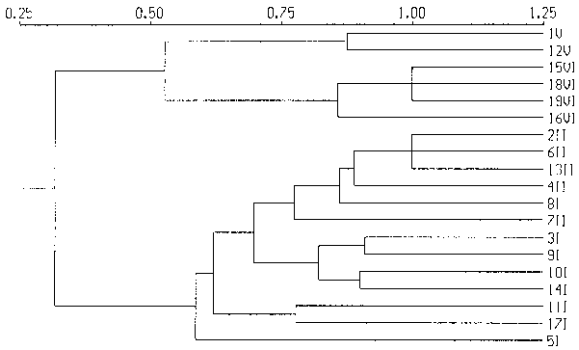

Figure 5. Dendrogram generated by UPGMA cluster analysis [22] of 1–F values for 19 rice accessions, based on Jaccard's coefficient determined using 16 markers located distal to centromeres ($>$ 15 cM from centromeres). Glaszmann's isozyme classifications (I, II, V and VI) are indiated.

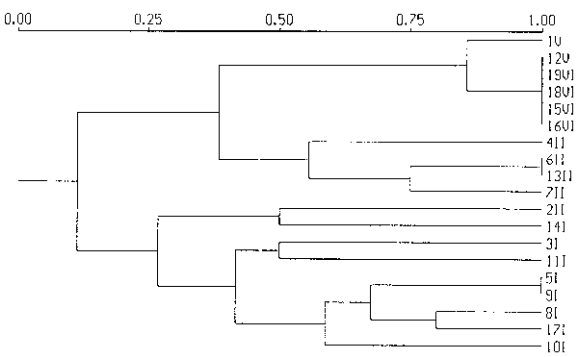

Figure 6. Dendrogram generated by UPGMA cluster analysis [22] of 1–F values for 19 rice accessions, based on Jaccard's coefficient determined using 10 centromeric markers ($<$ 15 cM from centromeres). Glaszmann's isozyme classifications (I, II, V and VI) indiated.

doubled haploid progeny produced following an *indica* $\times$ *japonica* cross.

In total, 26 of the PCR-based markers used in the initial diversity study could be mapped to chromosomal locations; these comprised 14 RAPD and 12 ISSR markers. An example of the segregation of RAPD markers among the doubled haploid progeny is shown in Fig. 4. A further 12 RAPD and 8 ISSR markers that were either not clearly amplified or were difficult to score during screening of the diverse material segregated clearly in the mapping population; in total therefore, 46 PCR-based markers were added to the RFLP framework. Overall, the RAPD markers spanned 9 chromosomes and the ISSR markers spanned 10, suggesting that both techniques sample across the rice genome with no obvious bias towards one or a few chromosomes.

Of the 26 PCR-based markers which were both assigned to chromosomal positions and used in the diversity screen, 16 could be considered as distal, located more than 15 cM from the centromere. The remaining 10 markers were proximal to the centromere. In total, four out of the 14 RAPD markers (29%) were classified as centromeric, whilst six out of 12 ISSR markers (50%) resided within 15 cM of the centromere. After classifying the PCR-based markers as either proximal or distal to the centromere, further cluster analyses were performed on the set of 19 rice accessions. When the distal markers were used to differentiate between accessions, the classification obtained was in good agreement with Glaszmann's isozyme classification [9], and the main division separated accessions in isozyme groups I and II from those in groups V and VI, as shown in Fig. 5. In contrast, when cluster analysis was performed using the ten centromeric markers, group II accessions were shown to be genetically more similar to the rices of groups V and VI (see Fig. 6). The limited sample size of markers mapping proximal to the centromere had a slight affect on the stability of this cluster analysis such that accession 2 (isozyme group II) clustered with accessions in isozyme group I.

The effect of sampling variation was further tested by randomly generating sub-samples of the 26 markers. Cluster analysis of the 19 accessions was performed on the basis of three pairs of randomly generated sub-samples, comprising 16 and 10 markers respectively. In all six cases, regardless of whether

the random sample comprised 16 or 10 PCR-based markers, isozyme group II accessions clustered with those in group I, with isozyme groups V and VI grouping together. These results indicate that the genomic location of molecular markers influences the pattern of relationships revealed in genetic diversity studies.

As an alternative approach, TWINSPAN numerical analysis was employed to facilitate the identification of key markers which are responsible for the separation of different groupings in a dendrogram [6]. When applied to the ISSR marker data set, comprising 40 markers, six ISSR markers (AM5-680, AM5-900, AM6-980, AM7-500, AM2-1560, and AM6-1600) were found to be responsible for the unexpected association of group II rices with groups V and VI rather than with group I. This is the result of the pattern of distribution of these six markers across the 19 accessions and, for convenience, they are referred to as UD (unexpected distribution) markers. Three of these UD ISSR markers could not be mapped since they were not polymorphic between the parents used to produce the doubled haploid population. The remaining three (AM5-900, AM6-980 and AM7-500) were mapped near to the centromeres of chromosomes 5 and 7 (Fig. 7). For each of these three key bands, homology of the comigrating fragments scored across the diverse accessions and the mapping progeny was confirmed by Southern analysis using gel-purified, radiolabelled amplified UD fragments as probes (see Fig. 2b).

Discussion

Two techniques for the detection of genetic variation, RAPD and ISSR-PCR have been compared directly for their ability to generate useful polymorphic bands in accessions of *O. sativa*. Slightly higher levels of polymorphism were revealed using ISSR (56%) compared to RAPD (50%). However, careful optimisation of the former technique was necessary to minimise background amplification of non-specific products; because of the differing G+C contents of the ISSR primers, different annealing temperatures were used. Although previous reports [5, 43] presented clear amplification profiles generated using 5′-anchored ISSR primers, in this study the most successful primers were anchored at their 3′ ends by two bases. The 5′-anchored primers and core repeats without anchoring nucleotides lacked the specificity required to amplify distinct PCR products. By priming within the microsatellites themselves, these oligonucleotides generated a range of non-specific ISSRs of variable lengths, resulting in smeared amplification profiles. Specificity of primer annealing resides in the first eight nucleotides from the 3′ terminus [2]. By anchoring oligonucleotides at their 3′ ends, the enhanced specificity reduces the number of sequences which have homology to the primer, limiting the number of potential amplifiable products, resulting in a 'fingerprint' of distinct marker bands, instead of a poorly resolvable 'smear'. 3′ anchoring of ISSR primers has also been used successfully by several other groups working with plants [15, 24, 28].

Both RAPD and ISSR markers reliably differentiate rice accessions into their respective isozyme groups. Since the latter correlate well with classifications based on crossability, this demonstrates the biological meaningfulness of classifications produced using the PCR-based markers. However, whereas the RAPD classification, in agreement with Glaszmann's isozyme classification [9], identifies a close genetic relationship between group I (*indica*) and group II accessions, the classification generated from ISSR data indicates that group II is genetically more closely related to groups V and VI (*japonica*).

In our study, the six ISSR markers with unexpected distributions (AM5-680, AM5-900, AM6-980, AM7-500, AM2-1560, and AM6-1600) do not appear to exhibit any common structural elements; they were produced using six different primers with three different core repeats (AAG, AGC and GGC). By their nature, ISSR markers span short chromosomal regions between microsatellites (SSRs) and hence one might expect to sample a different subset of the genome from that accessed using RAPD markers. This is not apparent in the general distribution of ISSR and RAPD markers across chromosomes. Both marker types span the majority of chromosomes and map both proximal to and distal to the centromeres, although more ISSR markers map near to centromeres.

Whilst the overall pattern of relationships of the rice groups revealed using the RAPD marker data had shown group II rices closely related to group I, the distribution of one of the 47 RAPD markers (K8-1000), like the 6 ISSR markers above, indicated that group II rices were more closely related to groups V and VI. Interestingly, this UD RAPD marker was mapped near to the centromere of chromosome 7. Figure 7 illustrates that all mapable UD markers (3 ISSR and 1 RAPD) are located close to the centromeres of chromosomes 5 and 7.

In spite of the relatively low resolution of centromere mapping because of the technical diffi-

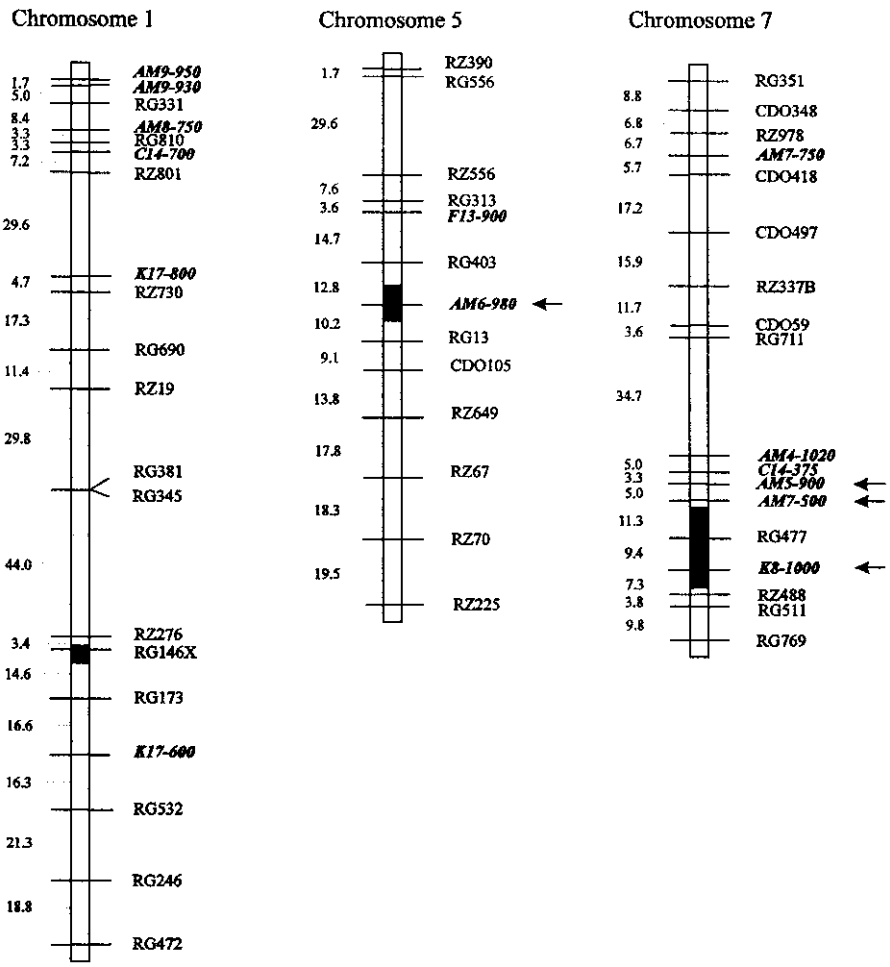

Figure 7. Molecular marker maps of chromosomes 1, 5 and 7. RFLP markers (mapped by Huang *et al.* [14]) are shown in standard text. Markers coded with RG, RZ and CDO were provided by S. Tanksley, Cornell University, USA. RAPD markers and ISSR markers are shown italicised and emboldened. UD (unexpected distribution) markers located on chromosomes 5 and 7 are indicated by arrows. Map distances, calculated using the Kosambi function [16], are indicated in centimorgans (cM). Centromeric locations are taken from Singh *et al.* [26] by aligning their RFLP framework with the RFLP framework used in this study; centromeres are located within the areas indicated as solid black regions.

culties involved in this process, it is clear that the three mappable ISSR markers with an unexpected distribution across the accessions studied are located close to the centromeres of chromosomes 5 and 7; furthermore, the only RAPD marker that shows the same unexpected distribution is also located close to the centromere of chromosome 7. It is not clear whether the fact that the location of these key markers on only two of the 12 chromosomes is significant since the numbers of markers involved are low, so that it has not been possible to generate chromosome specific analyses. It is indeed striking that all four mappable markers with unexpected distribution across the rice accessions are located close to centromeres. However, this effect on classification is not a characteristic of all markers that are located close to centromeres. Eight other markers used in this study map within 15 cM of the centromeric regions of chromosomes 2, 3, 7, 8 11 and 12; the distribution of these markers across the diverse accessions varies but has no clear effect on the classification of group II rices with respect to the other groups.

It is well known that centromeric regions tend to contain large amounts of repeated sequence (heterochromatic) DNA [11] and this may account for the increased representation of centromeric markers among the ISSR data set. Gupta *et al.* [12] used ISSR primers with tetranucleotide repeats and no anchoring nucleotides and was able to map markers close

to the centromeres of maize chromosomes. However, this does not explain the altered pattern of distribution of centromeric markers across diverse rice accessions. Rice classification based only upon markers mapping within 15 cM of a centromere results in a different pattern of relationships to that produced using markers mapping further along the chromosome arms. This suggests that, with regard to alleles at marker loci, group II rices are similar in their centromeric regions to groups V and VI (at least for two chromosomes) but for the more distal regions of the chromosomes they are more closely related to group I. This could mirror the pattern of chromosomal translocations that may have occurred during the differentiation of rice group II.

These results indicate that the detailed pattern of relationships within *O. sativa* germplasm is dependent on the type of marker employed. Intra-specific relationships observed using RAPD, isozymes and RFLP data are all similar [9, 31, 39]. We have shown that classification of rice accessions based on ISSR-PCR data differs in an important respect. Variation in ISSRs results from differences in the sequences between SSR loci. Analysis of rice material has also been reported using data generated from length polymorphisms in the SSR loci themselves, a consequence of variation in the number of repeated core units. Xu *et al.* [37] reported agreement between classifications based on RFLP data and data based on morphological and phenotypic characterisation, but poorer correspondence with classification based on microsatellite data. Yang *et al.* [38] demonstrated that microsatellites represent an efficient marker system for discriminating between closely related individuals and material with a narrow genetic base. However, microsatellites were less effective in determining major subgroup divisions.

The results of this study suggest that the chromosomal locations of molecular markers influence the patterns of relationships revealed in diversity studies. With the rapid development of plant molecular marker maps, the availability of information which will help to select markers well distributed throughout the genome is increasing. This wealth of information must not only be exploited for current agricultural advances but should also be used to help gain an insight into the genetic resources which will be incorporated into future cultivars.

Acknowledgements

Funding for the development of the ISSR marker protocols was provided by the ODA (Grant R5059). We are grateful for the technical assistance of Fay Hughes for help with the ISSR marker work and Amita Juliano for the isozyme classifications.

References

1. Arumuganathan K, Earle ED: Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208–218 (1995).
2. Caetano-Anollés G: DNA fingerprinting: MAAP: a versatile and universal tool for genome analysis. Plant Mol Biol 25: 1011–1026 (1994).
3. Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD: Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138: 1251–1274 (1994).
4. Chang TT: The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. Euphytica 25: 425–441 (1976).
5. Charters YM, Robertson A, Wilkinson MJ, Ramsay G: PCR analysis of oilseed rape cultivars (*Brassica oleracea* L. ssp. *oleifera*) using 5′-anchored simple sequence repeat (SSR) primers. Theor Appl Genet 92: 442–447 (1996).
6. Francisco-Ortega J, Newbury HJ, Ford-Lloyd BV: Numerical analyses of RAPD data highlight the origin of cultivated tagasaste (*Chamaecytisus proliferus* ssp. *palmensis* in the Canary Islands. Theor Appl Genet 87: 264–270 (1993).
7. Fukuoka S, Hosaka K, Kamijima O: Use of random amplified polymorphic DNAs (RAPDs) for identification of rice accessions. Jpn J Genet 67: 243–352 (1992).
8. Gawel NJ, Jarret RL: A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. Plant Mol Biol Rep 9: 262–266 (1991).
9. Glaszmann JC: Isozymes and classification of Asian rice varieties. Theor Appl Genet 74: 21–30 (1987).
10. Glaszmann JC: Geographic pattern of variation among Asian native cultivars (*Oryza Savita* L.) based on fifteen isozyme loci. Genome 30: 782–792 (1988).
11. Grady DL, Ratliff RL, Robinson DL, McCanlies EC, Meyne J, Moyzis RK: Highly conserved repetitive DNA sequences are present at human centromeres. Proc Natl Acad Sci USA 89: 1695–1699 (1992).
12. Gupta M, Chyi Y.-S., Romero-Severson, Owen JL: Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. Theor Appl Genet 89, 998–1006 (1994).
13. Hill MO: TWINSPAN – a Fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Cornell University, Ithaca, NY (1979).
14. Huang N, McCouch S, Mew T, Parco A, Guiderdoni E: Development of an RFLP map from a doubled haploid population in rice. Rice Genet Newsl 11: 134–137 (1994).
15. Kantety RV, Zeng X, Bennetzen JL, Zehr B: 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 (1995).
16. Kosambi DD: The estimation of map distances from recombination values. Ann Eugenet 12: 172–175 (1944).
17. 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, Kirihaya T, Hayasaka K, Miyao A, Monna L, Zhong HS, Tamura Y, Wang ZX, Momma T, Umehara Y, Yano M, Sasaki T, Minobe Y: A 300 kilobase interval genetic map of rice including 883 expressed sequences. Nature Genet 8: 365–372 (1994).
18. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L: MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174–81 (1987).
19. Litt M, Luty J: Hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within cardiac muscle actin gene. Am J Hun Genet 44: 397–401 (1989).
20. Mackill DJ: Plant genetic resources: classifying *japonica* rice cultivars with RAPD markers. Crop Sci 35: 889–894 (1995).
21. Ramakishana W, Lagu MD, Gupta VS, Ranjekar PK: DNA fingerprinting in rice using oligonucleotide probes specific for simple sequence repetitive DNA sequences. Theor Appl Genet 88: 402–406 (1994).
22. Rohlf FJ: NTSYS-pc Numerical Taxonomy and Multivariate Analysis System Version 1.70. State University of New York, Stony Brook, NY (1992).
23. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N: Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230: 1350–1354 (1985).
24. Salimath SS, de Oliveira AC, Godwin ID, Bennetzen JL: Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. Genome 38: 757–763 (1995).
25. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbour Laboratory Press, Cold-Spring Harbor, NY (1989).
26. Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS: Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.) Proc Natl Acad Sci USA 93: 6163–6268 (1995).
27. Tautz R: Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl Acids Res 17: 6463–6471 (1989).
28. Tsumara Y, Ohba K, Strauss SH: Diversity and inheritance of inter-simple sequence repeat polymorphisms in Douglas fir (*Pseudotsuga menziessi*) and sugi (*Cryptomeria japonica*). Theor Appl Genet 92: 40–45 (1996).
29. Virk PS, Ford-Lloyd BV, Jackson MT, Newbury HJ: Use of RAPD for the study of diversity within plant germplasm collections. Heredity 74: 170–179 (1995).
30. Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd BV: The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor Appl Genet 90: 1049–1055 (1995).
31. Wang ZY, Tanksley SD: Restriction fragment length polymorphism in *Oryza sativa* L. Genome 32: 1113–1118 (1989).
32. Wang G, Castiglione S, Zhang J, Fu R, Ma J, Li W, Sun Y, Sala F: Hybrid rice *Oryza sativa* L.: identification and parentage determination by RAPD fingerprinting. Plant Cell Rep 14: 112–115 (1994).
33. Weber JL, May PE: Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44: 388–396 (1989).
34. Welsh J, McClelland M: Fingerprinting genomes using PCR with arbitrary primers. Nucl Acids Res 18: 7213-7218 (1990).
35. Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV: DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res 18: 6531–6535 (1990).
36. Wu K, Tanksley SD: Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241: 225–235 (1993).
37. Xu Y, Olufote X, Chen WD, Park WD, Beachell HM, Dilday RH, Gotto M, McCouch SR: Evaluation of allelic diversity within and between rice cultivars using microsatellite and RFLP markers. Abstract 49, Plant Genome IV Conference, San Diego (1996).
38. Yang GP, Saghai Maroof MA, Xu CG, Zhang Q, Biyashev RM: Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 24: 187–194 (1994).
39. Yu LX, Nguyen HT: Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa* L.). Theor Appl Genet 87: 668–672 (1994).
40. Zhang Q, Gao YJ, Yang SH, Ragab RA, Saghai Maroof MA, Li ZB: A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor Appl Genet 89: 185–192 (1994).
41. Zhang Q, Saghai Maroof MA, Lu TY, Shen BZ: Genetic diversity and differentiation of indica and japonica rice detected by RFLP analysis. Theor Appl Genet 83: 495–499 (1992).
42. Zheng K, Qian H, Shen B, Zhuang J, Lin H, Lu J: RFLP-based phylogenetic analysis of wide compatibility varieties in *Oryza sativa* L. Theor Appl Genet 88: 65–69 (1994).
43. Zietkiewicz E, Rafalski A, Labuda D: Genome fingerprinting by simple sequence repeats (SSR) – anchored PCR amplification. Genomics 20: 176–183 (1994).