

# Use of RAPD for the study of diversity within plant germplasm collections

PARMINDER S. VIRK\*, BRIAN V. FORD-LLOYD, MICHAEL T. JACKSON† & H. JOHN NEWBURY

School of Biological Sciences, University of Birmingham, PO Box 363, Birmingham B15 2TT, U.K. and †Genetic Resources Center, International Rice Research Institute, PO Box 933, 1099 Manila, The Philippines

As part of the development of a molecular toolkit for the study of diversity within large plant germplasm collections, RAPD technology has been applied to accessions of rice (*Oryza sativa*) obtained from the major world collection held at IRRI (the International Rice Research Institute) which supplies germplasm to breeders. Methods for the speedy extraction of DNA representative of a rice accession, its amplification by PCR to reveal reproducible products, and the analysis of the banding data using numerical techniques have been established. The biological meaningfulness of RAPD data has also been demonstrated by reference to previous work on classification and crossability.

**Keywords:** biodiversity, genetic resources, *Oryza sativa*, plant breeding, RAPD.

## Introduction

Major collections of important crop plants are held in international centres around the world. They are repositories of the biodiversity available for each species and are a valuable source of useful genes for plant breeders. Workers at many of the centres face problems in the efficient maintenance and exploitation of germplasm because of the large numbers of accessions that they hold. Two complementary solutions have been proposed to enhance the efficiency with which such collections are handled: the identification and removal of duplicates and the production of 'core' collections. In the latter case, a subset of germplasm, which might comprise up to 10 per cent of the total, would be selected so that it represented, with minimal redundancy, as much as possible of the diversity present within the whole collection. The 'core collection' accessions would be the first to be supplied in response to requests by breeders for material from the collection. Because of the smaller size, the core collection could also be the focus of back-up conservation material as it could be stored at other international sites and this is the principal rationale behind the development of a core collection at IRRI (Vaughan & Jackson, 1994). This core collection concept does not

involve a reduction in the size of the whole collection, but increases the efficiency of its management by prioritizing representational subsets of accessions for special attention (Frankel, 1984; Frankel & Brown, 1984; Brown, 1989a,b; Andersen & Fairbanks, 1990).

Establishment of a core collection based on genetic variation shown by accessions would obviously be advantageous, but clear and detailed assessments of diversity within germplasm collections are not usually available. Such information can be obtained using molecular markers which afford many benefits for identifying variation and estimating biological diversity. DNA-based markers have the obvious advantage of sampling the genome directly and RFLP analysis has been widely used for assessing variation in plants (Helentjaris *et al.*, 1985; Tanksley *et al.*, 1989). RFLP analysis has been used to distinguish between species of *Oryza* (Wang *et al.*, 1992) and particularly between *indica* and *japonica* types of *O. sativa* (Zhang *et al.*, 1992).

More recently, PCR-based marker systems have become available. One of these has been called RAPD (Random Amplification of Polymorphic DNA) (Williams *et al.*, 1990; Newbury & Ford-Lloyd, 1993). In this method, short oligonucleotides of arbitrary sequence are used singly to support the amplification of regions of the test plant genome and the amplification products are separated by gel electrophoresis; differences between genotypes are reflected as differ-

\*Correspondence.

ences in the banding patterns. Other reports (Fukuoka *et al.*, 1992; Yu & Nguyen, 1994) have shown that RAPD analysis can be used to detect variation within a restricted range of *O. sativa* germplasm including *indica* and *japonica* types, to identify suitable parents for linkage map construction, and for gene tagging for drought resistance.

The International Rice Germplasm Center (IRGC) at the International Rice Research Institute (IRRI) in The Philippines conserves more than 75 000 accessions of Asian rice (*O. sativa*). It provides perhaps the best system for evaluating the efficiency of molecular techniques for the analysis of diversity and, in the longer term, the establishment of core collections and identification of duplicates within large germplasm collections.

In this study, we have critically assessed stages in the process of analysis of variation within a major germplasm collection using RAPD technology. We have considered the production of DNA samples representative of accessions, optimization of PCR to provide reproducible results, and a range of numerical analysis techniques to collate the data. The results obtained demonstrate that this technology is effective in the analysis of diversity within rice germplasm, and there is clear evidence that the pattern of variation revealed mirrors that described by workers using other methods. The results are discussed in the context of the use of RAPD methodology for enhancing the organization of a large germplasm collection and particularly for developing standardized procedures for analysing genetic diversity on a large scale.

## Materials and methods

### Materials

Twelve accessions of *O. sativa* were obtained from the IRGC at IRRI (Table 1) representing ecogeographical diversity and variation based on previous isozyme analysis (Glazmann, 1987). Two accessions (IRGC 123 = Dee-geo-woo-gen from Taiwan and IRGC 5823 = Kra Suey, a diverse land race of unknown origin) were used to optimize amplification conditions for producing consistent RAPD banding patterns. Seeds were sown in 5.7 cm jiffy pots containing sterilized moist compost which were placed in a growth cabinet maintained at 27°C. Leaves were harvested from 10 healthy 2-week-old seedlings of each accession and pooled to provide 20 mg of material for DNA extraction (see below).

### DNA isolation

Two methods were employed. The first followed the protocol reported by Edwards *et al.* (1991) in which a leaf disc was punched out using the lid of a microfuge tube, and DNA was subsequently extracted in the same tube. The second method was a modification of the protocol of Gawel & Jarret (1991). A total of 20 mg of leaf tissue was collected by harvesting equal amounts from each of 10 seedlings within an accession. This was placed in a 1.5 mL microfuge tube, immersed in liquid nitrogen, and homogenized using a polypropylene pellet pestle mixer driven by a hand-held cordless motor

**Table 1** List of *Oryza sativa* accessions included in the study

Accession number	Name	Isozyme group*	Origin	Type†
IRGC 31	Tangkai Rotan	1	Malaysia	<i>Indica</i>
IRGC 63	Slo 16	2	India	<i>Indica</i>
IRGC 201	Doc Phung	1	Vietnam	<i>Indica</i>
IRGC 328	Azucena	6	Philippines	<i>Indica</i>
IRGC 393	SML 81 B	1	Surinam	<i>Indica</i>
IRGC 752	Kumbi	1	China	<i>Indica</i>
IRGC 1635	PR304	6	Puerto Rico	<i>Indica</i>
IRGC 2564	Takao Mochi	6	Japan	<i>Japonica</i>
IRGC 3408	Chirgua 1 I-22	6	Venezuela	<i>Japonica</i>
IRGC 5418	Sintane Diofor	1	Burkina Faso	<i>Indica</i>
IRGC 5496	Niao Yao	1	Taiwan	<i>Indica</i>
IRGC 6064	British Guiana 79	1	Guyana	<i>Indica</i>

\*Glazmann (1987) classification.

†IRGC classification based on morphological traits scored at IRRI.

(Kontes, USA.). The powder was then mixed with 700  $\mu\text{L}$  of preheated extraction buffer containing 4 per cent CTAB. The homogenate was incubated at 65°C for 30 min prior to adding 600  $\mu\text{L}$  of chloroform:isoamyl alcohol (24:1, v/v). After mixing by inversion for 5 min at room temperature the mixture was centrifuged at 12000  $g$  in a microfuge for 5 min, and the supernatant recovered and mixed with an equal volume of ice-cold isopropanol. The DNA was recovered as a pellet by centrifugation at 12000  $g$  for 5 min, washed with 300  $\mu\text{L}$  of 70 per cent ethanol, dried under vacuum and dissolved in 100  $\mu\text{L}$  of sterile distilled water. DNA concentration was monitored by subjecting samples to 0.7 per cent agarose gel electrophoresis in TBE buffer (Sambrook *et al.*, 1989) and by visual assessment of band intensities compared with lambda DNA standards.

#### Polymerase chain reaction

A set of 24 decanucleotides of arbitrary sequence were employed (Table 2): 22 were obtained from Operon Technologies (20 from Kit F and two from Kit C) and two (BFL01 and BFL05) were produced at Birmingham University by Alta Bioscience. Unless otherwise indicated, the 25  $\mu\text{L}$  reaction mix contained approximately 1 ng template DNA, 0.2  $\mu\text{M}$  concentration of a single decanucleotide, 200  $\mu\text{M}$  each of dNTPs (Pharmacia) and 0.5 units *Taq* polymerase in the incubation buffer provided by the manufacturer of the enzyme (Boehringer Mannheim). The mixture was overlaid by two drops of mineral oil. Amplification was achieved in a Hybaid Omnigene thermocycler programmed as follows: cycle 1, 5 min at 94°C; cycles 2–46, 30 s at 94°C, 1 min at 35°C and 2 min at 72°C; cycle 47, 5 min at 72°C. Amplification products were subjected to electrophoresis in 1.2 per cent agarose gels and detected by ethidium bromide staining, viewing under ultraviolet light and photographed using Polaroid 667 film. Molecular weights were estimated using a 1 kb DNA ladder (Gibco-BRL).

#### Data analysis

Only data generated from the detection of polymorphic fragments were analysed. Specific amplification products were scored as present (1) or absent (0) for each DNA sample and an index of genetic similarity or distance calculated ( $1 - F$ ).  $F$  values were initially calculated using Nei & Li's (1979) matching coefficient method:

$$F = 2 \times N_{AB1} / (N_A + N_B).$$

In this and later equations,  $N_A$  = the number of bands in accession A,  $N_B$  = the number of bands in accession

B,  $N_{AB1}$  = the number of bands present (scored 1) in both accessions A and B,  $N_{AB0}$  = the number of bands present upon amplification of some of the germplasm within this set of accessions, but not present in either accession A or B, and  $N_T$  = the total number of bands scored in the study. Later,  $F$  values were also calculated using other formulae:

- 1  $F' = N_{AB1} / (N_T - N_{AB0})$  (Jaccard's coefficient)
- 2  $F'' = (N_{AB1} + N_{AB0}) / N_T$  (Similarity coefficient)
- 3  $F''' = N_{AB0} / (N_T - N_{AB1})$ .

An agglomerative method of clustering accessions was employed utilizing the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (SAS, 1990), TWINSpan (Two-Way Indicator Species Analysis), which employs a contrasting method of classification based on a divisive clustering technique, was also used to analyse the data (Hill, 1979; Francisco-Ortega *et al.*, 1993).

## Results

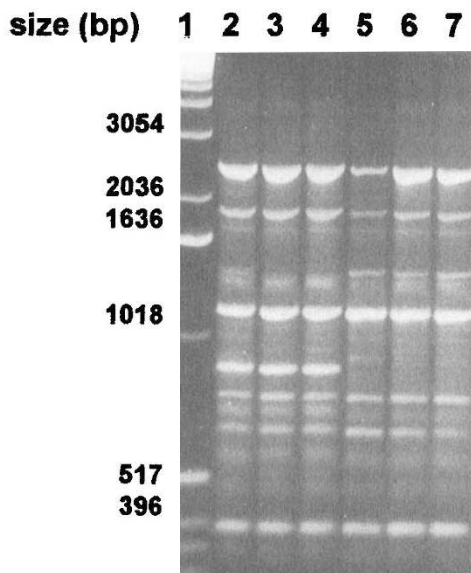
#### Optimization of RAPD protocol

For large-scale RAPD analyses, it is desirable to use a method for DNA isolation that allows fast extraction of DNA from a large number of samples while ensuring uniformity of yield and purity of samples and minimizing the risk of cross-contamination. The method described by Edwards *et al.* (1991) proved unsatisfactory for rice because it resulted in variable DNA yields between samples. When the modified extraction method of Gawel & Jarret (1991) was employed, the PCR protocol described in the Materials and methods resulted in reproducible patterns of amplification products when using a specific combination of accession and primer (Fig. 1).

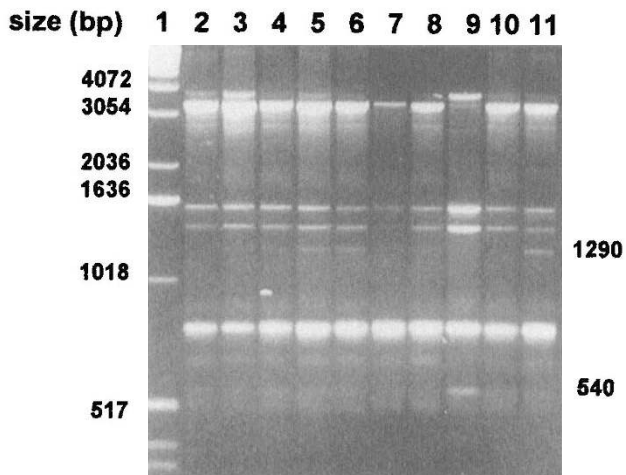
The sensitivity of RAPD PCR technology to changes in experimental parameters is well known (Munthali *et al.*, 1992) and we performed several experiments which both established a protocol yielding reproducible results and confirmed the importance of applying exactly the same protocol to ensure reproducibility. To optimize RAPD assay for rice we examined the effects of several parameters, namely, concentrations of magnesium ions, template DNA and primer, duration of time during the denaturation stage of the amplification cycle, different sources of polymerase enzyme and utilization of different thermocyclers (Caetano-Anolles, 1993; Macpherson *et al.*, 1993). In addition to the above reported factors we found that the use of a primer sequence synthesized by different suppliers and used at the same concentrations also resulted in slight banding changes (data not shown). Also certain primers produced reliable band-

**Table 2** Sequences and codes of random primers used and the number of monomorphic and polymorphic fragments amplified among 12 rice accessions

Primer	Sequence (5'-3')	No. of monomorphic fragments	Key to polymorphic markers and their sizes (bp)	
			Key	Size
OPF-01	ACGGATCCTG	3	1	1450
			2	1300
			3	900
OPF-02	GAGGATCCCT	2	4	2000
			5	1350
			6	1300
OPF-03	CCTGATCACC	3	7	1700
			8	1200
			9	750
OPF-04	GGTGATCAGG	2	10	650
			11	950
			12	900
OPF-05	CCGAATTCCC	2	13	700
			14	500
OPF-06	GGGAATTCGG	3	15	2200
			16	1900
			17	1850
			18	1700
			19	500
OPF-08	GGGATATCGG	3	20	800
OPF-09	CCAAGCTTCC	0	21	1600
			22	950
			23	700
			24	1100
OPF-12	ACGGTACCAG	2	25	1430
OPF-13	GGCTGCAGAA	5	26	925
			27	700
			28	650
			29	1560
OPF-14	TGCTGCAGGT	1	30	1530
			31	890
			32	870
			33	840
			34	750
OPF-15	CCAGTACTCC	2	35	1200
OPF-17	AACCCGGGAA	1	36	900
			37	800
			38	520
			39	2700
OPF-18	TTCCCGGGTT	0	40	1500
OPC-07	GTCCCGACGA	2	41	1000
OPC-15	GACGGATCAG	2	42	800
			43	700
			44	650
			45	1100
BFL-01	CGCTGTCCTT	1	46	700
BFL-05	GTCGCTGTCG	1	47	600
			48	500



**Fig. 1** Amplification products using DNA extracted from three individuals of each of two accessions of rice using primer OPF-01. Lane 1, 1 kb ladder: lanes 2-4 and 5-7 amplified using standard conditions (see Materials and methods) and DNA from accessions IRGC 123 and 5823, respectively.



**Fig. 2** Lane 1, 1 kb ladder. Lanes 2-10, amplification products following PCR directed by primer OPF-02 using DNA from nine individual rice plants of accession IRGC 5823. Lane 11, amplification products of DNA extracted from pooled leaf samples of nine individuals of accession IRGC 5823.

ing patterns only when used at a higher concentration than that reported in our standard protocol.

#### *Variation within accessions*

Rice accessions comprise a population of seeds originally collected from one site (e.g. a farmer's field)

usually from parts of the world in which modern varieties are not grown. Hence, it is quite possible that there is some genetic variation within an accession. To take account of this, an experiment was performed in which leaf samples from each of 10 individuals of one accession were used for individual DNA extraction, and DNA was also extracted from a pooled sample of 2 mg of leaf material from each of the ten individuals. An example of the results obtained following amplification directed by a decanucleotide is given in Fig. 2. Some differences between the banding patterns of individuals are apparent for this combination of accession and primer. Amplification of the pooled leaf sample DNA reveals all the bands that are monomorphic in the individuals. The fragment of 540 bp that is clearly present in only one individual is not represented in the amplification products of the pooled leaf samples. The 1290 bp product, which is clearly present in two individuals, is represented for the pooled sample. Taking these together with other data, it appears that the presence or absence of bands in the pooled sample that are not characteristic of all individuals is influenced by their frequency in the population. However, the minimum frequency required for a band to be visible within the pooled sample varies for different bands.

#### *Variation between rice accessions*

Twenty-four decanucleotides of arbitrary sequence (Table 2) were used to prime amplification of DNA sequences from 12 accessions of rice (Table 1). Reproducibility was checked by repeating PCRs at least twice on different occasions. The use of all primers led to amplification products, but primers OPF-17, -18 and OPC-07 gave banding patterns showing poor reproducibility and the use of primer OPF-7 resulted in a large number of faint bands when used at our routine concentration (0.2  $\mu\text{M}$ ). For all four of these primers, clear reproducible patterns were obtained when the primer concentration was increased from 0.2 to 0.8  $\mu\text{M}$ . Reproducible banding patterns were not obtained using primer OPF-19 even at the higher primer concentration; five primers (OPF-7, -10, -11, -16 and -20) generated only monomorphic bands for the rice accessions used here. Hence, data derived from the use of the latter six primers were not included in subsequent analyses. The remaining 18 primers yielded 83 reproducible marker bands of which 48 were polymorphic. The number of polymorphic bands varied between one and five per primer (with an average of 2.7) with amplification products varying between 300 and 2700 bp (Table 2). An example of RAPD variation detected using primers OPF-13 and



**Fig. 3** Amplification products obtained using DNA extracted from the 12 accessions given in Table 2 and employing primers OPF-13 (lanes 2–13) and OPF-14 (lanes 15–26). Lanes 1, 14 and 27, 1 kb ladder. \*For key to markers and their sizes see Table 2.

OPF-14 and DNA of 12 accessions is shown in Fig. 3. The amplification products obtained from OPF-13 were easily scorable. To facilitate scoring of some amplification products obtained using primer OPF-14 a longer electrophoretic run than that shown in Fig. 3 was required (data not shown).

#### *Analysis of relationships between accessions*

Cluster analysis (UPGMA method) was used to generate a dendrogram (Fig. 4). The analysis was carried out using the Nei & Li (1979) index and assuming amplification products of differing electrophoretic mobilities to be nonallelic and products with the same mobilities to be allelic. The dendrogram readily separated the 12 accessions into two major groups (X and Y) with one of the major groups being further divided into two minor ones (X1 and X2). The data were also subjected to three alternative methods for the calculation of similarity or distance between accessions before cluster analysis using an agglomerative method (see Materials and methods); these methods vary in the emphasis placed on the matching of bands present and bands absent and avoid the need to assume allelism between bands of the same mobility. No significant differences were observed between dendrograms produced using any of the four indices described; in each case the same accessions grouped together. In a further analysis, a divisive method of clustering based on initial ordination of material (TWINSPAN) was used to separate accessions. This analysis allows the identification of marker bands that are particularly diagnostic of each successive division of accessions. Results from the TWINSPAN analysis were in close agreement with those obtained using the agglomerative clustering

methods (Fig. 5). The re-ordered accession-by-marker data matrix obtained from this analysis is presented in Table 3. Accessions belonging to group Y were defined by 16 markers presented at the lower end of Table 3 (from 29 to 44). With regard to individual bands, markers 34 and 46 were present in all four *japonica* but none of the eight *indica* accessions whereas markers 38, 41 and 45 were present in all *indica* but none of the *japonica* accessions.

#### **Discussion**

Proven reproducibility of RAPD banding patterns is a prerequisite for the use of such data for the identification of variation and the estimation of diversity. Our results, in common with those reported by a large number of other workers (Newbury & Ford-Lloyd, 1993) demonstrate that, using a specific combination of template DNA and primer, reproducible results are obtained if an experimental protocol is properly tested and care is taken to avoid alteration of any of the experimental parameters. In addition to several reported factors affecting reproducibility of RAPD amplification our study has established that different sources and concentrations of primer influence banding patterns and that certain primers produce more reliable banding patterns containing more information when used at a higher concentration; this is probably because the genomic DNA used possesses an unusually high frequency of annealing sites for these primers so that their effective concentration for amplification is lowered in the reaction tube. This effect has also been noted in our laboratory using *Musa* DNA with the same primers (E. Howell, personal communication).

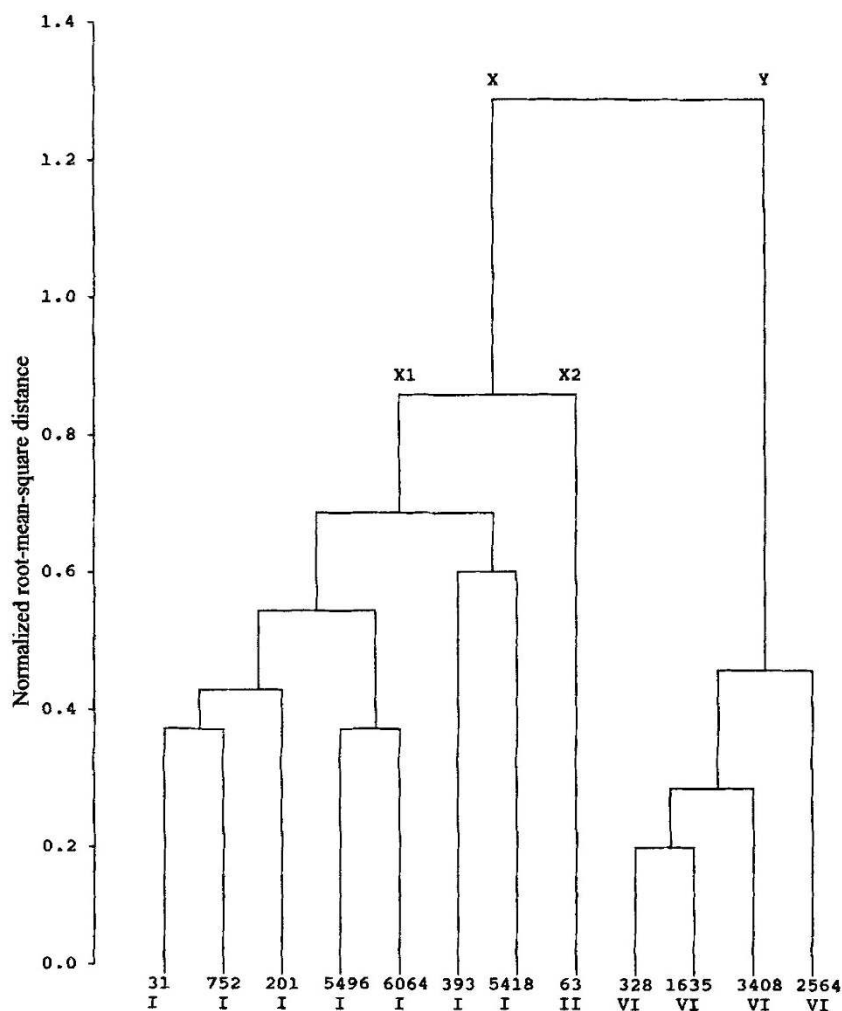


Fig. 4 Dendrogram generated by clustering using UPGMA analysis of 1-*F* values (based on Nei and Li's index) computed from pairwise comparisons of RAPDs between 12 accessions of *Oryza sativa*. The accession numbers are those of the IRGC collection (see Table 1). Roman numbers represent isozyme groupings of Glazmann (1987).

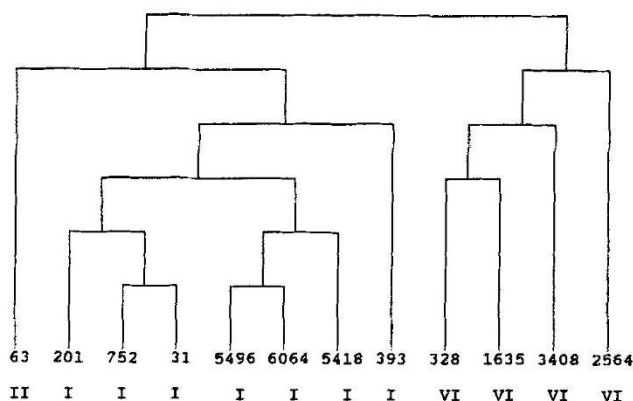


Fig. 5 Dendrogram of 12 accessions of *Oryza sativa* obtained after TWINSpan analysis. The accession numbers are those of the IRGC collection. Roman numerals refer to Glazmann's isozyme groups.

In all cases difficulties could be avoided or overcome following the identification of a potential source of artefact by rigorous testing for reproducibility, including the use of replicate DNA extracts from individual genotypes. Such testing is straightforward within a single laboratory but more complex when results obtained in two different laboratories are to be compared. We suggest that collation of specific RAPD data between laboratories can only be achieved with confidence following a demonstration that both collaborators obtain the same banding patterns using a test set of primers and defined genotypes.

*O. sativa* is predominantly autogamous in nature (<1 per cent outcrossing, G. Khush, personal communication; Butany, 1957) and hence little variation within accessions is expected. Nonetheless, within-accession variation could arise by natural cross-

**Table 3** Accession-by-marker data matrix generated from TWINSpan analysis of 12 accessions of *Oryza sativa*

Marker no.‡	Accession no.†
	565 132 27 4043645 60539619306 31216483584
31	011000000000
4	000000100000
6	000010000000
9	001111100000
15	010010000000
42	011111000000
40	001000110000
16	000000010000
36	011110110000
2	100000000000
18	100000000000
32	100000000000
47	100000000000
10	110000000000
24	110011010000
25	111110000000
30	111100000000
39	111111000000
38	111111100000
41	111111100000
45	111111100000
43	111111100001
11	111101100100
17	111111101100
23	111111111101
28	111111111110
48	111111111110
20	011111111111
22	011111111111
3	101111111111
8	111011111111
7	110111011111
29	000011111111
21	011000011111
26	010100011111
12	000010011111
14	000000000001
19	000000000001
1	000000010111
27	100000000111
5	000010001111
37	000000011111
33	000100001111
34	000000001111
46	000000001111
13	000000001101
35	000000001101
44	000000001110

1: present; 0: absent.

†For key to accessions see Table 1.

‡For key to markers see Table 2.

pollination which is influenced by floral morphology and environmental factors (Kato & Namai, 1987). It is also possible that some of the accessions are the result of the collection of mixtures of rice lines grown by some farmers. Wang & Tanksley (1989) found that 26 per cent of rice accessions exhibited polymorphism within five individuals of an accession using RFLP and they concluded that this was because of the presence of a mixture of pure-breeding lines within accessions. In our study, leaf samples were bulked from 10 seedlings in a pragmatic approach to take some account of any within-accession variation. Ideally, the patterns of amplification products from a large number of individuals would be obtained and scored but, given the large number of accessions present within most germplasm collections, this would make the application of molecular techniques impractical. The population of bands obtained when a pooled (ten) leaf sample is subjected to RAPD analysis is not always a summation of all the bands present in all of the individuals, presumably because of competition effects for the primer within the reaction tube. However, use of this sampling method must lead to a more representative population of bands for an accession than use of DNA from a single plant.

Variation between rice accessions was observed using RAPD technology with 18 of the 24 primers employed. While the ease with which variation was revealed in this study is influenced by the geographical diversity of the 12 rice accessions used, it is clear from other results obtained in our laboratory that variation between more closely related accessions is readily identifiable. One of the benefits of this initial study has been to identify primers which are likely to be efficient in revealing diversity in other sets of rice germplasm; for example in the Operon F kit, primers 3, 6, 13, 14 and 17 proved most useful since they all yielded four or five polymorphic bands when applied to the set of rice germplasm studied here.

With regard to the suitability of numerical techniques for the interpretation of RAPD data, we used two contrasting methods of clustering: agglomerative UPGMA and divisive TWINSpan. The former is a more popular method in which dendrograms reveal information based on similarity/dissimilarity between accessions and hence it is possible to identify readily closely-related accessions. The advantage offered by TWINSpan is that it provides a rearranged accession-by-marker data matrix based on statistical classification of both accessions and markers and this facilitates visual reinterpretation. Moreover, it identifies specific markers which contribute to each division of the dendrogram (Francisco-Ortega *et al.*, 1993). We recommend the use of both of these complementary methods of numerical analysis for RAPD data. The analyses of our RAPD banding data proved robust as



the same groupings of accessions were obtained by each method (i.e. using both divisive and agglomerative methods, and for the latter using a range of different methods for assessing genetic similarity or distance between accessions).

During the calculation of genetic similarity or distance using RAPD data, it has been assumed that comigrating amplified fragments are allelic. In RFLP analysis one can assume this with more confidence because a labelled probe has hybridized to both fragments under conditions of controlled stringency. For RAPD, information on sequence homology is limited to the 10-base termini of each amplification product. However, homology of comigrating RAPD products has been previously demonstrated for different species of *Glycine* and *Allium* (Wilkie *et al.*, 1993; Williams *et al.*, 1993). Indirect (but very significant) supporting evidence of allelism is derived from the conformity of taxonomic classifications based on RAPD data to those which may be widely accepted and based on more conventional approaches involving morphology, cytology and enzyme electrophoresis. For example, RAPD analysis has been shown to identify accepted relationships between genomes of *Musa* and *Brassica* (Demeke *et al.*, 1992; Howell *et al.*, 1994) and between species of *Stylosanthes* (Kazan *et al.*, 1993). We therefore feel that the assumption in our calculations that comigrating bands are allelic is not unrealistic. However, to confirm allelism conclusively the inheritance of band/s in the progenies of appropriate cross/es must be studied.

The classification of rice into the *japonica* and *indica* groups is important to breeders because of the difficulty in hybridization and recombination between these types. For breeders, the most useful classification into these groups is based on the isozyme work of Glaszmann (1987, 1988). His main groups I and VI corresponded to the *indica* and *japonica* types (in the sense of Oka, 1983) while Group II was an out-group from Group I. In our study, the 12 accessions were drawn from the much larger range of material analysed by Glaszmann and the pattern of variation revealed by the RAPD data is in strong agreement with the Glaszmann classification. The analyses grouped the 12 accessions based on 48 different band positions and the groups produced coincided directly with groups I, II and VI classified using isozyme data.

While Glaszmann's classification had shown concordance with previous groupings (Glaszmann, 1988) it is clear that identifications based on morphology alone can sometimes lead to different designations. For example, the lines Azucena and PR 304 have been classified as *indica* using morphological characters, whereas they behave as *japonica* type in crossing

studies (G. Khush, personal communication). These can also be classified as *japonica* following our RAPD analysis. Such discrepancies were apparent in a second experiment (data not shown) in which we used nine primers with 44 further accessions of *O. sativa* which had been classified as *indica* or *japonica* only on morphological grounds. Cluster analysis of the RAPD data divided these accessions into two major groups. All 31 of one group had been classified as *indica*; however, eight of the other group had been designated *japonica* or *javanica* while the other five had been designated *indica*. Whereas the RAPD classifications do not always correlate exactly with classifications based on morphology, they agree well with those based on isozyme analysis (Glaszmann, 1987, 1988) and on crossability data. This fact would make the utility of core collections defined by employing RAPD markers even more attractive both for the plant breeders as well as gene bank managers. We believe that this is strong evidence for the biological meaningfulness of the RAPD results.

These findings have an immediate practical application by rice breeders as the RAPD technology provides a fast method for classifying relatively uncharacterized accessions as *japonica* or *indica*. In a wider context, the speed, efficiency and reliability of the RAPD methodology, in conjunction with numerical analysis, render this technology particularly appropriate for the formulation of strategies for the effective management of germplasm collections in terms of duplicate identification, the estimation of diversity, monitoring genetic erosion and enhancing the use of those collections. This study represents part of the development of a package of procedures for studying agricultural biodiversity.

## Acknowledgements

This work was funded by the Overseas Development Administration, Plant Sciences Programme, Project Number R5059.

## References

- ANDERSEN, W. R. AND FAIRBANKS, D. J. 1990. Molecular markers: important tools for plant genetic resource characterization. *Diversity*, **6**, 51–53.
- BROWN, A. D. H. 1989a. The case for core collections. In: Brown, A. D. H., Marshall, D. R., Frankel, O. H. and Williams, J. T. (eds) *The Use of Plant Genetic Resources*, pp. 136–151. Cambridge University Press, Cambridge.
- BROWN, A. D. H. 1989b. Core collections: a practical approach to genetic resources management. *Genome*, **31**, 818–824.

- BUTANY, W. T. 1957. Natural crossing in rice. *Rice News Teller*, **5**, 18-21.
- CAETANO-ANOLLES, G. 1993. Amplifying DNA with arbitrary oligonucleotide primers. *PCR Methods Applic.*, **3**, 85-94.
- DEMEKE, T., ADAMS, R. P. AND CHIBBAR, R. 1992. Potential taxonomic use of Random Amplified Polymorphic DNA (RAPD)- a case study in *Brassica*. *Theor. Appl. Genet.*, **84**, 990-994.
- EDWARDS, K., JOHNSTONE, C. AND THOMPSON, C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucl. Acids Res.*, **19**, 1349.
- FRANCISCO-ORTEGA, J., NEWBURY, H. J. AND FORD-LLOYD, B. V. 1993. 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.
- FRANKEL, O. H. 1984. Genetic perspectives of plant germplasm conservation. In: Arber, W., Llimensee, K., Peacock, W. J. and Starlinger, P. (eds) *Genetic Manipulation: Impact on Man and Society*, pp. 161-170. Cambridge University Press, Cambridge.
- FRANKEL, O. H. AND BROWN, A. D. H. 1984. Current plant genetic resources- a critical appraisal. In: *Genetics: New Frontiers*, vol. IV, pp 1-11. Oxford and IBH Publishing Co., New Delhi.
- FUKUOKA, S., HOSAKA, K. AND KAMIJIMA, O. 1992. Use of random amplified polymorphic DNAs (RAPDs) for identification of rice accessions. *Jap. J. Genet.*, **67**, 243-252.
- GAWEL, N. J. AND JARRET, R. L. 1991. A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. *Pl. Mol. Biol. Rep.*, **9**, 262-266.
- GLASZMANN, J. C. 1987. Isozymes and the classification of Asian rice varieties. *Theor. Appl. Genet.*, **74**, 21-30.
- GLASZMANN, J. C. 1988. Geographic pattern of variation among Asian rice cultivars (*Oryza sativa* L.) based on 15 isozyme loci. *Genome*, **30**, 782-792.
- HELENTJARIS, T., KING, G., SLOCUM, M., SIEDENSTRANG, C. AND WEGMAN, S. 1985. Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. *Plant. Mol. Biol.*, **5**, 109-118.
- HILL, M. O. 1979. *TWINSPAN- a Fortran Program for Arranging Multivariate Data in an Ordered Two-way Table by Classification of the Individuals and Attributes*. Cornell University, Ithaca, New York.
- HOWELL, E. C., NEWBURY, H. J., SWENNEN, R. L., WITHERS, L. A. AND FORD-LLOYD, B. V. 1994. The use of RAPD for identifying and classifying *Musa* germplasm. *Genome*, **37**, 328-332.
- KATO, H. AND NAMAI, H. 1987. Floral characteristics and environmental factors for increasing natural outcrossing rate for F<sub>1</sub> hybrid seed production of rice (*Oryza sativa* L.). *Jap. J. Breed.*, **37**, 318-330.
- KAZAN, K., MANNERS, J. M. AND CAMERON, D. F. 1993. Genetic variation in agronomically important species of *Stylosanthes* determined using random amplified polymorphic DNA markers. *Theor. Appl. Genet.*, **85**, 882-888.
- MACPHERSON, J. M., ECKSTEIN, P. E., SCOLES, G. J. AND GAJADHAR, A. A. 1993. Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. *Mol. Cell Probes*, **7**, 289-293.
- MUNTHALI, M., FORD-LLOYD, B. V. AND NEWBURY, H. J. 1992. The random amplification of polymorphic DNA for fingerprinting plants. *PCR Methods and Applications*, **1**, 274-276.
- NEI, M. AND LI, W. H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.*, **74**, 5267-5273.
- NEWBURY, H. J. AND FORD-LLOYD, B. V. 1993. The use of RAPD for assessing variation in plants. *Plant Growth Reg.*, **12**, 43-51.
- OKA, H. I. 1983. The *indica-japonica* differentiation of rice cultivars- a review. *Proc. 4th Int. SABRAO Congress*, pp. 117-128.
- SAMBROOK, J., FRITSCH, E. F. AND MANIATIS, T. 1989. *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.
- SAS 1990. *SAS/STAT User's Guide*. Version 6, 4th edn, vol. 1. SAS Institute, Cary, NC.
- TANKSLEY, S. D., YOUNG, N. D., PATERSON, A. H. AND BONIERBALE, M. W. 1989. RFLP mapping in plant breeding: new tools for an old science. *Bio/Technology*, **7**, 258-264.
- VAUGHAN, D. A. AND JACKSON, M. T. 1994. The core as a guide to the whole collection. In: Hodkin, T. and Brown, A. D. H. (eds) *Proceedings of the International Workshop on Core Collections. IBPGR/CGN/CENARGEN Workshop*, Brasilia, 23-28, August, 1992. John Wiley (in press).
- WANG, Z. Y. AND TANKSLEY, S. D. 1989. Restriction fragment length polymorphism in *Oryza sativa* L. *Genome*, **32**, 1113-1118.
- WANG, Z. Y., SECOND, G. AND TANKSLEY, S. D. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theor. Appl. Genet.*, **83**, 565-581.
- WILLIAMS, J. G. K., KUBELIK, A. R., LIVAK, K. J., RAFALSKI, J. A. AND TINGEY, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.*, **18**, 6531-6535.
- WILLIAMS, J. G. K., HANAFEY, M. K., RAFALSKI, J. A. AND TINGEY, S. V. 1993. Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol.*, **218**, 704-740.
- WILKIE, S. E., ISAAC, P. G., AND SLATER, R. J. 1993. Random amplified polymorphic DNA (RAPD) markers for genetic analysis in *Allium*. *Theor. Appl. Genet.*, **86**, 497-504.
- YU, L.-X. AND NGUYEN, H. T. 1994. Genetic variation detected with RAPD markers among upland and lowland rice cultivars (*Oryza sativa* L.). *Theor. Appl. Genet.*, **87**, 668-672.
- ZHANG, Q., SAGHAI MAROOF, M. A., LU, T. Y. AND SIEN, B. Z. 1992. Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. *Theor. Appl. Genet.*, **83**, 495-459.