Theor Appl Genet (1998) 96: 602–611

J. Zhu · M. D. Gale · S. Quarrie
M. T. Jackson · G. J. Bryan

AFLP markers for the study of rice biodiversity

Received: 15 July 1997/Accepted: 29 October 1997

Abstract AFLP was used as a DNA fingerprinting technique in rice (*Oryza sativa* L.) germplasm analysis. The high efficiency and random coverage of AFLP markers were established. With only five combinations of primers and RFLP anchors, a framework linkage map was constructed. This map demonstrated that the AFLP markers from a limited number of primers were not confined to any particular regions or chromosomes in the rice genome. To analyse the biodiversity of 57 rice germplasm accessions, we examined 179 polymorphic AFLP markers generated from four primer combinations. Both principal component analysis and cluster analysis were used, and three groups were clearly identified which corresponded to genotypes of Isozyme Groups I, II and VI. The number of markers needed for robust classification of rice germplasm and the diversity between/within the groups was established.

Key words AFLP · Rice · Biodiversity · Germplasm · Mapping

Introduction

Cultivated rice (*Oryza sativa* L.) is one of the most important crops since it is the staple food of 50% of the world's population. Moreover, it is also an ideal model plant for the study of grass genetics and genome organisation due to its diploid genetics, relatively small genome size (C = 0.45 pg, Causse et al. 1994; 430 Mb, Kurata et al. 1994), significant level of genetic polymorphism (McCouch et al. 1988; Wang and Tanksley 1989), large amount of well-conserved genetically diverse material (approximately 100,000 accessions of rice germplasm worldwide) and the availability of widely collected, compatible wild species (Jena et al. 1992). In recent years, a linkage map of 1,383 DNA markers (Kurata et al. 1994), including 883 expressed sequences, combined with the recently observed high levels of conserved synteny between grass species (Moore et al. 1995) has made the genetics of rice even more attractive as a model for map-based gene cloning, genome organisation and evolution research, as well as plant breeding research.

Several molecular marker systems have been applied to genetic mapping and biodiversity studies of cultivated rice. Restriction fragment length polymorphism (RFLP) is the most conventional and has played the major part in genetic mapping, comparative mapping and classification of the species in the *Oryza* genus (Wang et al. 1992). However, RFLP detects relatively low amounts of genetic polymorphism and, in comparison with recently developed polymerase chain reaction (PCR)-based methods, is a somewhat labourious procedure. More informative and convenient PCR-based systems have been developed and used in biodiversity studies and genetic mapping, such as Simple Sequence Repeat (SSR) or microsatellite (Wu and Tanksley 1993), RAPD (Virk et al. 1995a) and anchored SSR (or Inter-SSR-PCR) (Zietkiewicz et al. 1994). Isozyme markers have also been used in a seminal study for the classification of 1,760 Asian cultivated rice varieties into six groups (Glaszmann 1987). Recently, the chromosome location of markers such as RAPD and ISSR-PCR was reported to be able to influence rice diversity assessments (Parsons et al. 1997).

Amplified Fragment Length Polymorphism (AFLP) is a relatively new DNA fingerprinting technique (Vos

Communicated by P. Langridge

J. Zhu (✉) · M. D. Gale · S. Quarrie · G. J. Bryan[1]
John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK
Fax: + 44-01603 502241
E-mail: zhu@bbsrc.ac.uk

M. T. Jackson
International Rice Research Institute, PO Box 933, 1099 Manila, Philippines

Present address:
[1] Crop Genetics Department, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

et al. 1995) which uses selective amplification of restriction fragments. This PCR-based method is able to generate complex banding patterns, DNA fingerprints, of up to at least 100 DNA fragments in each reaction. AFLP is, therefore, potentially very useful for the study of biological diversity, varietal identification and genetic mapping. Indeed, AFLP has been recently applied in gene isolation (Thomas et al. 1995), QTL analysis (Quarrie et al. 1997) as well as in other genetical and physical mapping studies (Becker et al. 1995; Mackill et al. 1996; Maughan et al. 1996). However, it was not very clear whether the AFLP markers from a limited number of primers were suitable for biodiversity analysis.

This paper describes experiments designed to test the efficiency and fidelity of the use of AFLP markers for surveying rice genetic diversity and for genetic mapping. The AFLP technique has been used to fingerprint rice accessions, and the results have been compared to those obtained by isozyme analysis and to those obtained by Mackill et al. (1996). The distribution of AFLP markers in the rice genome and the numbers of AFLP markers needed for the generation of stable groups have also been investigated.

Materials and methods

Genetic stocks

Approximately 90 plants from an F_2 mapping population, consisting of 120 individuals, derived from a cross between IR20 (*Indica*) and 63-83 (*Japonica*) were used for AFLP mapping. The seeds of a set of 57 rice accessions used in these experiments were from the International Rice Genebank Collection (IRGC) of the International Rice Research Institute (IRRI). These accessions, 14 of which had been analysed with isozymes, were taken from a random sample of approximately 180 rice accessions originating from south and southeast Asia, which in turn had been drawn from a larger set of material selected on the basis of the availability of accurate geographical information.

Preparation of genomic DNA from rice accessions

For each accession germplasm, approximately equal amounts of leaf tissue from 10 individual seedlings were pooled as described by Virk et al. (1995). A modified version of the CTAB method (Murray and Thompson 1980) was used for the isolation of rice genomic DNA. The liquid nitrogen-ground leaf material (1–3 g) was thoroughly mixed with 20 ml of preheated 65°C CTAB buffer [2.5% sorbitol, 1% *n*-lauryl sarcosine (Sarkosyl), 0.8% CTAB (hexadecyl trimethyl ammonium bromide, cetrimide), 4.7% NaCl, 0.8% EDTA (Ethylenediaminetetraacetic acid disodium salt), 1.3% TRIS-HCl, pH 8.5, 1% PVPP (polyvinylpolypyrrolidone, insoluble), 5 μg/ml RNase A (added to the CTAB buffer just before the extractions were performed)] and incubated at 65°C for 2 h. After two chloroform extractions, the DNA was precipitated and dissolved in 0.5 ml TE buffer.

Preparation of template DNA for AFLP reactions

The procedure was modified from the AFLP Protocol for Public Release (Zabeau and Vos 1993). *Mse*I and *Pst*I were used for the double digestion of template DNA. The *Pst* adapter and the *Mse* adapter (Table 1) were ligated to the ends of restriction fragments. Following adapter ligation the sequence complexity of the DNA mixture was reduced by selecting biotinylated DNA fragments using streptavidin-coated beads.

Non-selective preamplification

Before selective PCR, a pre-amplification was carried out to amplify the DNA fragments non-selectively. The reaction was the same as the preamplification (the first of the two successive selective amplification steps for selection with more than three selective nucleotides at each end) described by Zabeau and Vos (1993), except that a wider range of amount of template DNA (0.1–1.0 μg) was used, the primers (P00 and M00) contained none of selective nucleotides and the PCR programme was 30 cycles of 30 s at 94°C, 1 min at 56°C and 1 min at 72°C.

This non-selective amplification had a few advantages. It provided a more convenient template for subsequent selective PCR and obviated the need for resuspending the streptavidin-coated beads each time. The pre-amplified PCR product also provided an opportunity to check the quality and quantity of the PCR fragments in an agarose gel, and the amount of template for subsequent PCR could be adjusted easily by diluting the non-selective PCR product between 1 and 50 times.

Selective amplification

Three selective nucleotides (i.e. +3) on the *Pst* primer combined with two selective nucleotides (+2) on the *Mse* primer were selected for use. The sequence of the primers used are shown in Table 1. The pre-amplification product was normally diluted 20 times as a template for selective PCR. The *Pst* primer was end-labelled with [^{33}P]-γ-ATP using T4 polynucleotide kinase. The amplification reactions were performed essentially as outlined in the AFLP Protocol.

AFLP electrophoresis and autoradiography

AFLP products were electrophoresed on 5% denaturing polyacrylamide gels containing 7 *M* urea and 0.5 × TBE buffer. One volume of loading buffer (98% formamide, 10 m*M* EDTA pH 8.0, 0.1% bromophenol blue and xylene cyanol) was added to the samples. Each gel was run at 80–90 W for approximately 2 h. After the gel was dried on Whatman paper, AFLP products were visualised, either by autoradiography or by the use of a Molecular Dynamics PhosphorImager.

Table 1 Sequence of AFLP adapters and primers

Name	Sequence	
Mse-adapter I	5′-GACGATGAGTCCTGAG-3′	
Mse-adapter II	5′-TACTCAGGACTCAT-3′	
Pst-adapter I[a]	5′-CTCGTGACTGCGTACATGCA-3′	
Pst-adapter II	5′-CATCTGACGCATGT-3′	
M00	5′-GATGAGTCCTGAGTAA-3′	
P00	5′-AGACTGCGTACATGCAG-3′	
Other primers:	M12 = M00 + AC	M13 = M00 + AG
	P66 = P00 + GAT	P67 = P00 + GCA
P75 = P00 + GTA	P77 = P00 + GTG	P78 = P00 + GTT

[a] This primer was biotin labelled

RFLP probes

RFLP probes were from the collections of the Rice Genome Project, Japan (prefixes C, G, R, S, Y), Cornell University, USA (prefixes BCD) and the John Innes Centre (prefixes psr and ABA). They were selected according to previously published map locations (Causse et al. 1994; Kurata et al. 1994).

RFLP procedure

All procedures for restriction enzyme digestion, gel electrophoresis, Southern transfer, probe preparation and labelling and filter hybridization were performed as described by Sharp et al. (1988) and Devos et al. (1992).

Linkage analysis

MAPMAKER/EXP version 3.0B (Lincoln et al. 1992) was used for linkage analysis. The LOD threshold was 3.0, and the distance threshold was 50 cM with the Haldane function.

Data analyses

Each informative AFLP band was scored independently as 1 for "presence" and 0 for "absence". Shannon's information index (Maughan et al. 1996):

$$H_s = -\sum f_1 \ln(f_i)$$

was used to reveal the level of polymorphism where H_s is the phenotypic diversity value and f_i is frequency of an AFLP band, or a set of AFLP bands, in a specific group of rice accessions.

$$H_a = 1/n \sum H_s$$

is the average group diversity over n groups and

$$H_w = -\sum f \ln f$$

is the phenotypic diversity calculated from the whole collection tested. The intra- and inter-group diversity components, H_a/H_w and $(H_w–H_a)/H_w$, were compared.

Principal component analysis (PCA), based on sums-of-squares-and-products or correlation matrixes, was carried out using the Genstat programme (Genstat-5-Committee 1987). Similarity matrixes were generated based on the simple-matching coefficient (Sokal and Michener 1958) using the presence/absence data for individual AFLP fragments between pairs of rice accessions. Cluster analysis using the average-linkage (UPGMA) method was performed using the Genstat programme, with the similarity matrix as input data. Mantel tests were performed with the programme NTSYSpc, using similarity matrices generated from different subsets of the AFLP data.

Results

Linkage of AFLP markers

To investigate the distribution of AFLP markers in the rice genome, we constructed a framework linkage map of AFLP markers using approximately 90 plants from an F_2 intercross population generated from the IR20 (Indica, Isozyme Group I) × 63–83 (Japonica, Isozyme Group VI) cross. The F_2 population was produced by selfing a single F_1 rice plant from the cross between IR20 and 63/83. Thirty-six RFLP markers, mainly from the Japanese Rice Genome Project (Kurata et al. 1994), were used as anchors to determine the identity of linkage groups. These RFLP markers were selected to cover all rice chromosomes, based on published mapping information, and were evenly distributed in the genome, with the exception of chromosome 12, which contained only 1 of the selected RFLP markers (Fig. 1).

Five combinations of AFLP primers were used in the mapping experiment. They were termed P0278, P0277, P0276, P0178 and P0177, which corresponded to the combinations of primers M13 + P78, M13 + P77, M13 + P67, M12 + P78 and M12 + P77, respectively. Each primer combination generated between 80 and 115 clear fragments, of which between 7 and 25 were polymorphic. Overall, 20% of the AFLP fragments were polymorphic and, of these, approximately 80 were scored. Sixty-five AFLP markers were mapped to particular rice chromosomes, with the remainder not assigned to any known linkage group.

Nineteen linkage groups were identified and, with the help of single-copy RFLP anchors, the 12 rice linkage groups were constructed (Fig. 1). In this framework map, most of the AFLP fragments were linked to at least 1 RFLP "anchor", and hence their location in the linkage groups was determined. Four AFLP markers were tightly linked to each other, but not to any anchor probes. One of these markers, P017811C, was later on mapped chromosome 12 as the P017811/12 locus due to its linkage to probe C751 (Quarrie et al. 1997). Chromosomes 1, 3, 4, 5, 8, 10 and 11 were well-covered with markers. The AFLP markers in the framework linkage map were clearly not evenly distributed across the entire rice genome. However, the AFLP markers from this limited number of primer combinations were also not confined to any chromosome or chromosomal region. The combination of primers P78/M13 provided 18 markers which mapped to 8 of the 12 chromosomes. Similarly, the combination of P77/M13 generated 16 markers which mapped to 8 rice chromosomes. The least productive combination (P76 and M13) provided only 5 markers, which were located in 3 linkage groups. These results suggest that the AFLP markers were not confined to any particular regions of chromosomes, such as centromere or telomere or regions close to them. Instead, the numbers and locations of the polymorphic markers generated by different primer combinations were quite random and variable. Of particular significance for the study of rice biodiversity was the observation that the mapped AFLP markers were distributed throughout the rice genome.

Ten pairs of AFLP markers appeared to be allelic and could be mapped as co-dominant markers. The allelic relationship of these AFLP markers was

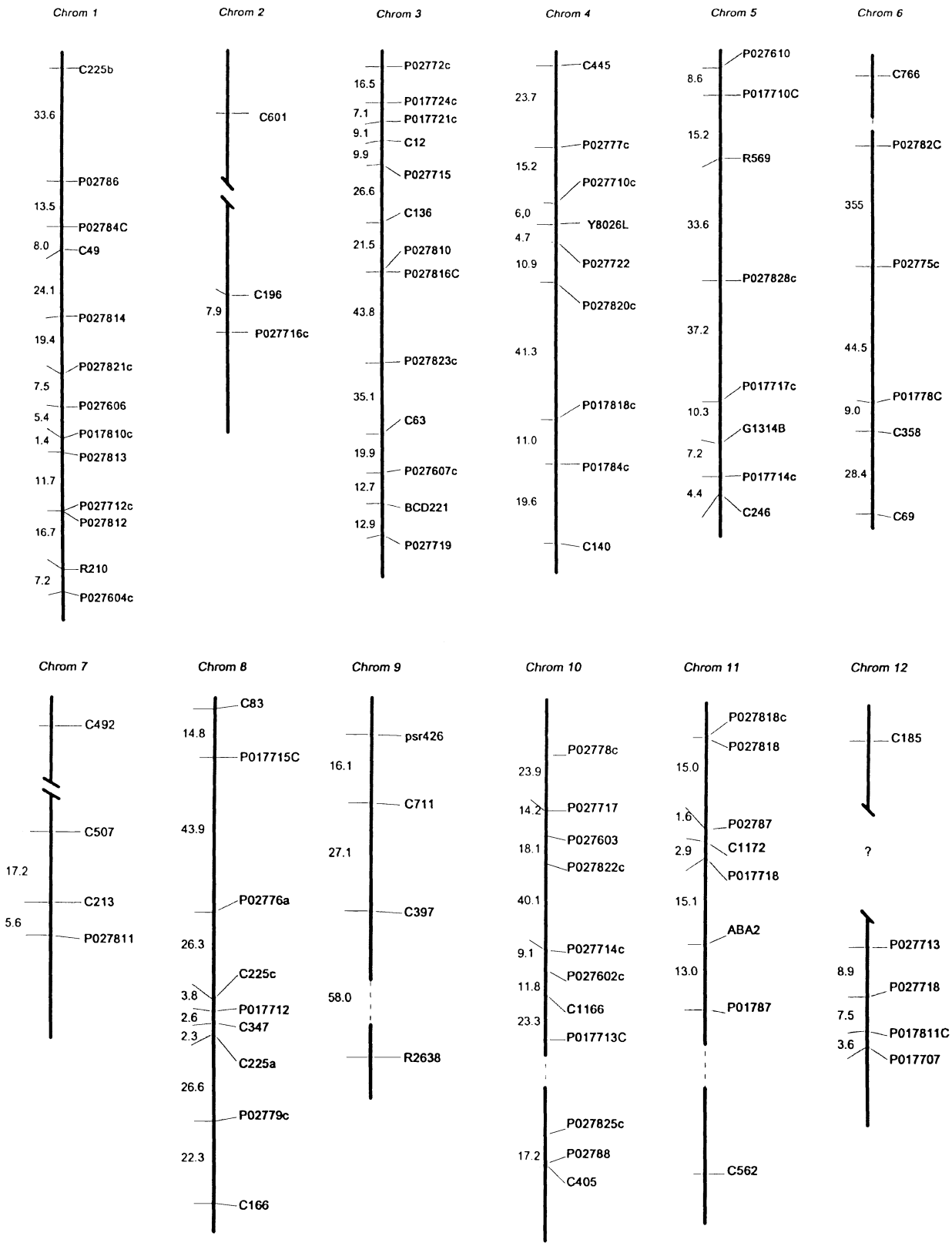

Fig. 1 The linkage map of the rice genome with AFLP and RFLP markers. The probes of the RFLP anchors were provided by the Rice Genome Program, Japan. Their location in the rice genome was described by Kurata et al. (1994). The names of all AFLP markers begin with *P0*. The *first two digits* indicate the *Mse*I end primers, the *third* and *fourth* show the *Pst*I end primers and the *following numbers* are the series number of the polymorphic AFLP fragments. The appendices *C* represent a codominant marker. The LOD threshold was 3.0, and the distance threshold was 50 cM. The unit of genetic distance labelled in the map is centiMorgans (cM)

suggested by their segregation characteristics, since each pair was never absent in the same individual of the F2 population, and the ratio of A : H : B tended to be close to 1 : 2 : 1. These loci were named only after 1 of the markers but a "C" (co-dominant) was appended (Fig. 1).

Level of AFLP in a rice germplasm collection

Four combinations of primers were used to analyse 57 rice accessions originating from south and southeast Asia. A total of 410 bands were generated, of which 179 (44%) were polymorphic. These markers could distinguish all 57 accessions, and the number of different bands between the closest 2 accessions was 3. Variation in the amount of template DNA, between 0.2 and 1.0 μg in this experiment, did not affect the intensity of the AFLP bands. The number of polymorphic markers from the four primer combinations ranged from 39 to 51, with an average of 45 markers per primer pair. The frequency of individual polymorphic AFLP fragments ranged from 0.02 (i.e. present in only 1 accession) to 0.98 (present in all but 1 of the 57 accessions) (Fig. 2).

Principal component and cluster analysis

Three groups of rice accessions were clearly defined by the first and second principal components, which represented 29.16% and 20.27% of the diversity in this sample, respectively (Fig. 3). Group C was well-separated from the other two groups, mainly on the basis of the first principal component. The second principal component separated group B from group A. The points within each group were very tightly clustered, and the three groups were separated with big gaps. Therefore, the first and second principal components were able to separate the 57 accessions into three clear groups.

The third principal component, which represented 4.26% of the variation, revealed a further sub-group by partitioning group B into two subgroups, although the division was not as clear as between groups A, B and C. The fourth and fifth principal components (3.58% and 3.26%, respectively) revealed diversity within group A but not within groups B and C. These results suggested that the third principal component mainly corresponded to diversity within group B, while the fourth and fifth components corresponded to diversity within group A.

Isozyme information obtained from IRRI for 14 rice accessions within this sample (unpublished data) suggests that groups A, B and C correspond to isozyme groups I, II and VI, respectively (Glaszmann 1987) (Table 2). Group I corresponds to the *Indica* subspecies and group VI to the *Japonica* subspecies.

The similarity matrix was used to cluster the data using the unweighted pair-group method with an arithmetic average (UPGMA) algorithm (Fig. 4). The level of dissimilarity between group C and the combination of groups A and B was around 50%. Groups A and B differed at the 37% level. The level of dissimilarity between the subgroups observed within groups A, B and C was 27%, 25% and 20%, respectively. Figure 4 showed that the intermediate accessions identified in PCA were also separated from the group B in cluster analysis, but the relations between these intermediate accessions and the other groups could not be revealed using the latter method. The main subgroups within groups A and B were observed in both cluster analysis and PCA of the first five components (data not shown).

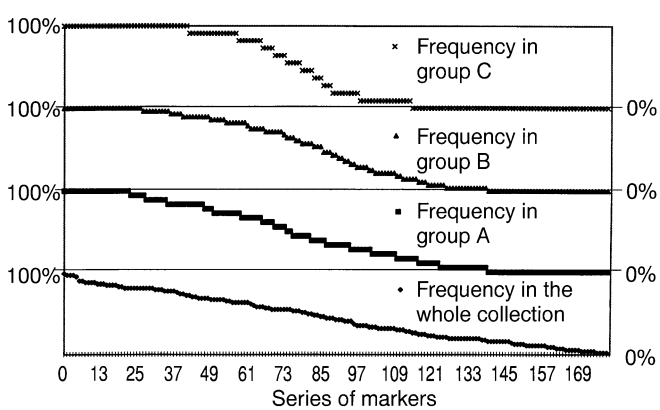

Fig. 2 The frequency of the 179 AFLP fragments in the collection and in the groups of the collection (identified in Fig. 3) were sorted and shown independently. Many markers were polymorphic in the collection but not polymorphic within the groups of the collection.

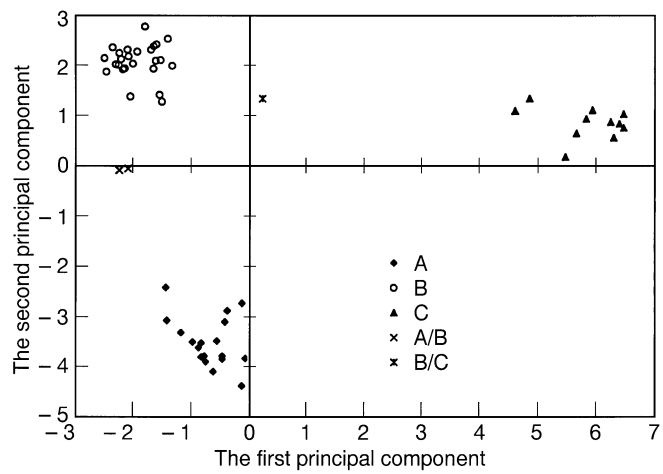

Fig. 3 The plot of PCA of 57 rice accessions with 179 polymorphic AFLP markers. *A*, *B* and *C* represent the groups A, B and C. *A*/*B* and *B*/*C* show the intermediate accessions in this plot.

Table 2 Grouping of the 57 rice accessions from south and south-east Asia

Name	Accession number	Group	Source	Isozyme
Uba Murali	IRGC 25928	A	Bangladesh	
Bashiraj	IRGC 26298	A	Bangladesh	
Ashmber	IRGC 27522	A	Bangladesh	
Bazail	IRGC 27526	A	Bangladesh	
Hnanwa	IRGC 33118	A	Myanmar	
Asu	IRGC 62154	A	Bhutan	
Ray Jazaykayz	IRGC 62181	A	Bhutan	
Ray Nabja	IRGC 26184	A	Bhutan	
Ngaja	IRGC 64917	A	Bhutan	
Balasuriya A	IRGC 66509	A	Sri Lanka	
Girresa	IRGC 66512	A	Sri Lanka	I
Guru Muthessa	IRGC 66513	A	Sri Lanka	I
Cut Keureusek	IRGC 66540	A	Indonesia	
Fache Dogo	IRGC 66547	A	Indonesia	
IE mata cicem	IRGC 66562	A	Indonesia	
Kedot	IRGC 66568	A	Indonesia	
Lekat Lembayung	IRGC 66585	A	Indonesia	
Neli	IRGC 66603	A	Indonesia	I
Molladigha	IRGC 25906	A/B[a]	Bangladesh	
Sampatti	IRGC 25920	A/B[a]	Bangladesh	
Aus Jhari	IRGC 25833	B	Bangladesh	
Benaful	IRGC 25839	B	Bangladesh	
Benamuri	IRGC 25840	B	Bangladesh	II
Dhalashaita	IRGC 25851	B	Bangladesh	II
Dumai	IRGC 25852	B	Bangladesh	
Jabarshail	IRGC 25865	B	Bangladesh	
Jhum Fulbadam	IRGC 25868	B	Bangladesh	II
Lakhsnikajal	IRGC 25885	B	Bangladesh	
Pankhiraj	IRGC 25911	B	Bangladesh	
Sultanjata	IRGC 25924	B	Bangladesh	
Noroi	IRGC 31611	B	Bangladesh	
Boteshawar	IRGC 34682	B	Bangladesh	
Lakhi Puri	IRGC 34712	B	Bangladesh	
Mery	IRGC 34722	B	Bangladesh	
Serety	IRGC 34732	B	Bangladesh	
Bawoi	IRGC 34737	B	Bangladesh	
Chikon Shoni	IRGC 64771	B	Bangladesh	
Dharia	IRGC 64773	B	Bangladesh	
Kal Shoni	IRGC 64780	B	Bangladesh	
Kat Gimi	IRGC 64781	B	Bangladesh	
Moshur	IRGC 64789	B	Bangladesh	II
Narikel Jhupi	IRGC 64792	B	Bangladesh	II
Pakhoil	IRGC 64793	B	Bangladesh	II
Shada	IRGC 64796	B	Bangladesh	
Sreeramur Shaita	IRGC 64799	B	Bangladesh	
Panniti	IRGC 66527	B/C[a]	Sri Lanka	
Mimidam	IRGC 25897	C	Bangladesh	
Mimidim	IRGC 25898	C	Bangladesh	
Walanga	IRGC 27502	C	Indonesia	
Npe837	IRGC 38697	C	Pakistan	
Dagpa Bara	IRGC 64887	C	Bhutan	VI
Dumja Kaap	IRGC 64890	C	Bhutan	VI
Maap	IRGC 64911	C	Bhutan	
Mandasherpo	IRGC 64913	C	Bhutan	VI
Tschinanangka	IRGC 64935	C	Bhutan	
Podi Niyan Wee	IRGC 66529	C	Sri Lanka	VI
Pulut Gantung	IRGC 66612	C	Indonesia	VI[a]

[a] The classification was not certain for the particular accession

Phenotypic diversity

Phenotypic diversity values (H_s) calculated for the four primer pairs ranged from 10.15 to 12.35, with an average value of 11.39 (Table 3). There appeared to be large differences in amounts of diversity within the three groups defined by the clustering/PCA analyses. The *Indica* group showed the highest level of diversity,

Table 3 The number of AFLP markers and the Shannon's information index

Primers	Number of bands	Number of polymorphic bands (*Np*)						Phenotypic diversity (*H*)					
		Indica	Group B	*Japonica*	Shared	Total	*Np*/*N*	*Indica*	Group B	*Japonica*	H_w	*H-intra*	*H-inter*
P66/M13	100	28	24	19	9	43	43%	6.31	5.01	3.95	12.00	0.42	0.58
P75/M13	104	29	24	22	12	46	44%	7.47	5.82	5.10	11.07	0.53	0.47
P77/M12	104	26	25	16	7	39	38%	6.13	4.94	3.74	10.15	0.49	0.51
P77/M13	106	34	40	16	8	51	48%	9.07	6.82	2.85	12.35	0.55	0.45
Mean	104	29	28	18	9	45	43%	7.25	5.65	3.91	11.39	0.50	0.50
Total	418	118	117	72	36	181							

Notes: *H-intra* and *H-inter* represent the intra-group and inter-group diversity components, respectively

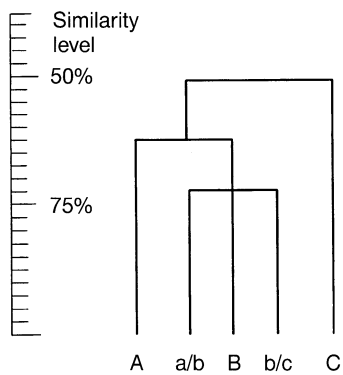

Fig. 4 This dendrogram was based on the cluster analysis of 57 rice accessions with 179 AFLP markers. The *A*, *B* and *C* were identified by PCA. *a*/*b* and *b*/*c* represent the intermediate accessions in PCA, and they were also separated from group B in the cluster analysis

while the *Japonica* group showed the lowest. Partitioning of the phenotypic diversity into within-group and between-group components showed that the AFLP variation was equally divided between the two components (Table 3).

Consistency of analysis with subsets of AFLP markers

The robustness of the clustering/PCA results and the number of AFLP markers needed for reliable groupings was investigated. Based on the linkage studies, the AFLP markers from each primer combination were more or less randomly distributed in the rice genome. Therefore, subsets of markers generated by each primer combination represented random subsets of the full set of data.

The four combinations of primers generated four basic subsets of markers, which were I (P77 + M13), II (P77 + M12), III (P66 + M13) and IV (P75 + M13) and which provided 51, 39, 43 and 46 markers, respectively. From the four basic subsets, larger subsets could be constructed by combining any two or three basic subsets. The results of PCA showed that any of the subsets, alone or in combination, could clearly identify the three groups of rice accessions which were defined by the whole set of markers. However, when fewer markers were used, the groupings became looser. Use of the smallest basic subset, II, containing 39 markers, did not provide clear separation between groups A and C.

Table 4 Comparison of average correlations between similarity matrices using subsets of AFLP data of different size. Values are r-values using the Mantel test programme in NTSYSpc. Note that the value in brackets for two prime pairs is the average of the three possible orthogonal data sets. The *n* represents the number of pair-wise in the comparison

	Comparison with same size of data set	*n*	Comparison with whole data set (i.e. 4 primer pairs)	*n*
One primer pair	0.71	6	0.88	4
Two primer pairs	0.91 (0.85)	15 (3)	0.96	6
Three primer pairs	0.97	6	0.99	4

Mantel tests were used to compare similarity matrices generated using different subsets of the AFLP markers. Correlations between the matrixes generated from different subsets of AFLP markers are shown in Table 4. These results showed that the AFLP data generated from two, or at most three primer combinations appeared to be sufficient for robust similarity estimates for this set of material and that there was little advantage in generating sets of data from four or more primer combinations. However, a larger set of data would allow more orthogonal comparisons of the effects of using subsets of the data. For example, in this analysis, it was impossible to compare orthogonal data sets comprising data from three of the four AFLP primer combinations, since the four possible such data sets contained data from two primer pairs in common, which inflates the correlations between the data sets.

Comparison of classification using AFLP or isozyme data

Only 14 of the 57 accessions used were associated with isozyme grouping information. Based on this informa-

tion, the AFLP groups A, B and C corresponded to isozyme groups I, II and VI, respectively. This correspondence was confirmed in a further study on 48 rice accessions which were analysed with both AFLP and isozyme markers (data was not presented).

Resolution of AFLP markers for discrimination of varieties

All 57 accessions surveyed could be discriminated using the 179 AFLP markers. The highest similarity between any 2 accessions was 98% and five pairs of rice accessions showed this level of similarity. The 2 accessions in each pair with this high level of similarity differed by 3 or 4 AFLP markers. This high power of resolution was also retained when only two primer combinations were used. At least 1 AFLP fragment difference could be found between the most similar accessions using any two combinations of AFLP primers.

Group-specific markers

A study of the frequency of the markers in the groups defined by clustering and PCA showed that, among the 57 rice accessions, 8 markers had a frequency of 100% in group C but were absent from both groups A and B. Conversely, 5 markers were absent from group C but fixed in both groups A and B. Furthermore, many other markers were present at a high frequency in one group but absent in others. These group-specific markers have been confirmed in larger rice accession samples (> 300 accessions, unpublished data).

Discussion

Low sensitivity of AFLP to variations in template DNA concentration

AFLP banding patterns were not particularly sensitive to the initial concentration of rice template DNA. This observation is consistent with Vos et al. (1995). Furthermore, the AFLP bands tended to behave similarly when reaction conditions were varied. This high "uniformity" of AFLP banding patterns render it easy to detect problems with individual samples or reactions, so reducing the possibility of erroneous scoring caused by variation in band density. This gives AFLP a considerable advantage over other PCR-based DNA fingerprinting methods such as RAPD. In our experiments, DNA mixtures composed of up to 10% of DNA from 1 accession mixed with that of another genetically distinct accession sample could not be detected by AFLP markers (data was available but not shown). This insensitivity of AFLP to DNA mixtures decreases the potential for erroneous scoring due to sample contamination but also decreases the potential for detecting polymorphism in bulked DNA samples.

Efficiency and genome coverage of AFLP

Markers from only five primer combinations covered most arms of the rice chromosomes except chromosome 9, and one arm of chromosome 2, 7 and 12. Furthermore, these AFLP markers appeared not to be confined to any particular region(s) of the rice chromosomes, although it is not known yet whether they would cluster when more AFLP markers are accumulated. Therefore, on this evidence, it seems likely that AFLP markers have wide coverage throughout the rice genome. This wide coverage was important for the use of AFLP markers in biodiversity studies (Parsons et al. 1997), particularly in the map-based AFLP fingerprinting study (unpublished data).

Suitability of AFLP for germplasm analysis

Comparison of the groupings obtained with different subsets of AFLP markers suggests that the markers from any one of the primer combinations, which can generate more than 30 polymorphic markers, could be sufficient for the classification of the major groups of cultivated rice accessions. However, when more markers are used, there are fewer ambiguities. The 'transferability' of mapped AFLP markers between crosses or into germplasm collections was also investigated, and 'map-based' AFLP fingerprinting has been carried out. The aim of this approach was to estimate genetic diversity using markers whose position in the genome is known. This would allow the determination of the levels of genetic diversity in different regions of the rice genome. If markers used for germplasm analysis were simultaneously mapped on a segregating population, a polymorphism "profile" of the genome could be generated, which would lead to more precise germplasm evaluation studies and plant breeding research.

This study clearly demonstrates the efficiency of using the AFLP technique for rice germplasm analysis. The costs of performing AFLP appeared high, particularly if [^{33}P] is used for detection. However, compared with other marker systems, such as microsatellites, the development costs are negligible, and a competent laboratory worker, familiar with PCR technology, could generate AFLP data within a few weeks. Moreover, because of the incomparably high speed with which fingerprinting data was accumulated using AFLP markers, the cost per marker was relatively low for germplasm analysis. Our studies suggest that data from only two or three AFLP primer

combinations (82-140 markers) would be sufficient to estimate accurately the similarities between any two of the major rice groups. Furthermore, little advantage would be gained from generating larger sets of data. In practice this means that any accession can be fingerprinted using only two or three tracks on a sequencing gel. This observation is limited to the material under study, and clearly more markers are needed for more closely related varieties, such as cultivar collections, or for pedigree analysis. However, our results suggest that, if this result is generally applicable, AFLP may well offer a very cost-effective way of fingerprinting large germplasm collections. Moreover, the AFLP data reported here produce classifications that are consistent with those based on isozyme studies (Glaszmann 1987). A more extensive and rigorous comparison of different molecular techniques for rice germplasm analysis was carried out (unpublished data). An AFLP analysis using a smaller number of rice accessions has recently been reported (Mackill et al. 1996). This study contains an analysis of 147 polymorphic AFLP fragments, from 17 AFLP primer combinations, an average of only eight to nine fragments per primer pair. However, these authors largely used a $+3/+3$ AFLP fingerprinting strategy, compared to the $+3/+2$ strategy employed in this study, so that about five times the number of polymorphic markers from each primer combination were observed in our experiments.

More recently, the AFLP procedure has been simplified and the process for purification of the adapter-ligated template DNA has been omitted (Vos et al. 1995). Therefore, the AFLP technique has become more convenient and less expensive than the version which was used here (Zabeau and Vos 1993). We have since noted no difference between the quality and reliability of the fingerprints using the two procedures (data not presented).

Grouping of *O. sativa*

It might be widely accepted that the cultivated rice in Asia falls into six groups, Isozyme Groups I–VI (Glaszmann 1987), in which Isozyme Group I and VI correspond to *Indica* and *Japonica* groups respectively. The accessions analysed in this experiment were from a collection selected randomly from the International Rice Genebank, but this collection only included the accessions from south and south-east Asia. This would be the reason why the accessions of Isozyme Groups III, IV and V were not included in this collection and the temperate *Japonica* cultivars were also excluded. The grouping of the 57 accessions has been confirmed by another experiment with larger number of accessions (up to 400 accessions). The consistency of grouping with AFLP and isozyme markers has also been demonstrated with more rice accessions (unpublished data). However, it has to be noted that the conclusion regarding grouping in this paper is limited to the accessions within south and south-east Asia.

Comparison of AFLP with other marker systems

The theoretical number of combinations or "haplotypes" of the 179 polymorphic markers generated by four AFLP primer combinations is 7.7×10^{53}. This is equivalent to the number of allelic combinations from 77 SSR loci that had an average of five alleles, or 54 loci with ten alleles each. While, the codominant nature of microsatellite (SSR) markers and the almost 100% transferability of map position makes them preferable to AFLP, more studies are required to compare the relative merits of the two types of approaches. It is perhaps prudent to consider SSRs as the logical replacement for RFLP, especially for linkage studies, and AFLP as a more robust and productive replacement for RAPD technology. SSRs can be used to provide codominant 'anchor' markers for mapping studies, but the development and application costs may hinder their application in the large numbers needed to study a large germplasm collection. It is likely that strategies utilising a combination of the two marker techniques will prevail in the coming years.[1]

Acknowledgements We thank Graham Moore and his group for preparing the DNA of the F_2 population, Dave Laurie for reconfirming the AFLP scoring in linkage analysis, James Brown for helpful discussion on statistical matters, Pauline Stephenson for kind help on experimental techniques and Robert Koebner for helpful discussion and comments on the manuscript. We also acknowledge researchers of the Rice Genome Program of Japan for supplying mapped RFLP probes. This study was a part of the Plant Sciences Research Programme supported by the Overseas Development Administration, UK. We also thank our collaborators in this programme in Birmingham University, John Newbury, Bryan Ford-Lloyd and Parminda Virk for the important discussions and collaboration. Some of the authors are supported by the Biotechnology and Biological Sciences Research Council, UK.

References

Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gel Genet 249:65–73

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

Devos K, Atkinson M, Chinoy C, Liu C, Gale M (1992) RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye. Theor Appl Genet 83:931–939

Genstat-5-Committee (1987) Genstat 5: reference manual. Oxford University Press, New York

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 74:21–30

[1] All data are available when required

Jena KK, Khush GS, Kochert G (1992) RFLP analysis of rice (*Oryza sativa* L.) introgression lines. Theor Appl Genet 84 : 608–616

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

Lincoln S, Daly M, Lander E (1992) Constructing genetic maps with MAPMAKER/EXP 3.0, 3rd edn. Whitehouse Institute Technical Report, Whitehouse Institute, Cambridge, Mass.

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

Maughan PJ, Saghai Maroof MA, Buss GR, Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. Theor Appl Genet 93 : 392–401

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76 : 815–829

Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution. Curr Biol 5 : 737–739

Murray M, Thompson W (1980) The isolation of high molecular weight plant DNA. Nucleic Acids Res 8 : 4321–4325

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

Quarrie S, Laurie D, Zhu J, Lebreton C, Seikhodskii A, Steed A, Witsenboer H, Calestani C (1997) QTL analysis to study the association between leaf size and abscisic acid accumulation in droughted rice leaves and comparisons across cereals. Plant Mol Biol 35 : 155–165

Sharp PJ, Kreis M, Shewry PR, Gale MD (1988) Location of β-amylase sequences in wheat and its relatives. Theor Appl Genet 75 : 286–290

Sokal R, Michener C (1958) A statistical method for evaluationg systeatic relationships. Univ Kan Sci Bull 38 : 1409–1438

Thomas C, Vos P, Zabeau M, Jones D, Norcott K, Chadwick B, Jones J (1995) Identification of amplified restriction fragment polymorphism (AFLP) markers tightly linked to the tomato *CF-9* gene for resistance to *Cladosporium fulvum*. Plant J 8 : 785–794

Virk PS, Ford-Lloyd BV, Jackson MT, Newbury HJ (1995a) Use of RAPD for the study of diversity within plant germplasm collections. Heredity 74 : 170–179

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

Wang ZY, Second G, Tanksley SD (1992) Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor Appl Genet 83 : 565–581

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

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application 92402629. 7; Publication number EP 0534858 A1

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