Heredity **76** (1996) 296–304 *Received 27 June 1995*

Predicting quantitative variation within rice germplasm using molecular markers

PARMINDER S. VIRK*, BRIAN V. FORD-LLOYD, MICHAEL T. JACKSON†, HARPAL S. POONI, TOMAS P. CLEMENO† & 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

Diverse Asian rice (*Oryza sativa*) germplasm has been used to identify associations between various quantitative traits and RAPD molecular markers using multiple regression analysis. This has allowed us to predict for other samples of germplasm their performance for traits such as culm length and number, days to flowering, grain width, and panicle and leaf length using only RAPD marker data. Such predictive capability is possible because of the availability of extensive diversity held in genebanks, and can be used in the future to facilitate the exploitation of that biodiversity. More specifically the methodology could facilitate crop improvement by rapid ideotype prediction. For the mapping and isolation of QTLs (genes controlling quantitative traits) the method would provide information to guide the selection of parental material for hybridization and markers expected to show linkage to QTLs. It may also be possible that these associations could lead the way towards marker-assisted selection during breeding programs. In the future, this demonstration of association between markers and easily measured traits could also be extended to the study of important adaptive traits, such as stress tolerance, found either within germplasm collections or in natural populations.

Keywords: genebank, genetic resources, molecular markers, quantitative characters, RAPD, rice.

Introduction

Plant genetic resources, which represent crop biodiversity, have been actively conserved for four decades in 'genebanks' across the globe. The Consultative Group on International Agricultural Research (CGIAR) centres maintain about 500 000 accessions of more than 30 crops, while the United States Plant Germplasm System stores 380 000 samples of over 8000 plant species. The reasons for conserving the world's biodiversity as a global resource are varied. They range from the moral and aesthetic to the practical, where future needs and values are perceived to be unpredictable, and where the impact of the disappearance of any component of an ecosystem is currently uncertain. The need for the conservation of crop biodiversity has its basis in agricultural demands, with priority being given to maintaining germplasm which may enhance crop improvement either immediately or in the future.

*Correspondence.

The genebank at the International Rice Research Institute (IRRI) has a collection of more than 80 000 samples of rice germplasm (Jackson & Huggan, 1993) comprising mostly landrace varieties, breeding lines and commercial varieties of *Oryza sativa* but also including landrace varieties of *O. glaberrima* and all 20 wild species of the genus *Oryza*. Since 1973, over 740 000 packets of rice have been distributed throughout the world for use in applied research, and this germplasm has contributed to improvements in many characteristics of new rice varieties (Jackson, 1994). Pressure on germplasm distribution will increase over the next 30 years as plant scientists strive to meet the demands for increased rice production. The management of such collections is difficult simply because of their vast size, and there is a clear requirement for the development of procedures which utilize fast and reliable methods for the measurement of diversity in order to facilitate the organization and prioritization of the germplasm resources (Virk *et al.*, 1995b). However, even more important is the need to be able to identify rapidly and efficiently the most

appropriate rice material for use in basic and applied research, including crop improvement.

In recent years there has been an explosion of new DNA-based marker methods utilizing the polymerase chain reaction (PCR). One of the simplest and most widely used is Random Amplification of Polymorphic DNA (RAPD) (Williams *et al.*, 1990; Welsh & McClelland, 1990). RAPD technology has been used successfully for measuring diversity in plants and the patterns of variation observed have been shown to resemble closely those obtained using more classical characters (Howell *et al.*, 1994; Virk *et al.*, 1995a). In the present study, we have used RAPD on highly diverse accessions of rice in order to determine associations between the DNA marker(s) and quantitative traits and then we have utilized these associations to predict quantitative traits in other germplasm.

Materials and methods

Quantitative evaluation

Initially, 200 accessions from the South and South East Asian germplasm collection maintained at the IRRI Genetic Resources Center were selected at random. These 200 accessions were then evaluated at Los Baños in the Philippines in a randomized plot experiment, with two replicate blocks which were grown during the dry season (November, 1993–May, 1994). Quantitative data were collected on 10 representative plants of each accession and each plant was scored for 10 traits; viz., culm number, culm length (cm), culm diameter (mm), grain length (mm), grain width (mm), leaf length (cm), leaf width (cm), days to 50 per cent flowering, panicle length (cm) and seedling height (cm), essentially following International Board for Plant Genetic Resources – International Rice Research Institute (IBPGR–IRRI) descriptors for rice (IBPGR–IRRI Rice Advisory Committee, 1980).

In order to assemble material that represented the range of diversity found within the initial 200 accessions, the quantitative data were subjected to cluster analysis. Prior to analysis the data were standardized to zero mean and unit variance, because various traits were measured on very different scales. Agglomerative hierarchical clustering was performed on the squared Euclidean distance matrix utilizing the unweighted pair group method using arithmetic averages (UPGMA). This method was chosen because of its simplicity and widespread use, and because it is the most appropriate for situations where all accession means are subjected to the same sampling error (Rohlf, 1992). This then led to the selection of 48 accessions which were extracted from the resultant dendrogram (Fig. 1) by stratification. Table 1 shows the countries of origin and geographical locations of collection of these accessions.

DNA extraction and PCR analysis

Seedlings of the 48 accessions were raised in jiffy pots for 2–3 weeks in order to obtain leaf material for DNA extraction. A sample of fresh leaf (2 mg) was taken from each of 10 randomly selected seedlings and then mixed before co-extraction to obtain a DNA sample representative of each accession (Virk *et al.*, 1995a). For the RAPD analysis, the total reaction volume was 25 μL containing 5 ng DNA, 200 μM of each dNTP, 0.4 μM decanucleotide primer (supplied by Operon), 1 unit of Taq polymerase, 2.25 mM $MgCl_2$ and 1 × reaction buffer comprising 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8.8), and 0.1 per cent Tween 20. The mixture was overlaid by with 45 μL mineral oil and the amplification was performed in a thermocycler (Hybaid–Omnigene) programmed as follows: one cycle of 95°C for 2 min; two cycles of 95°C for 30 s, 37°C for 1 min and 72°C for 2 min; two cycles of 94°C for 30 s, 37°C for 1 min and 72°C for 2 min; 41 cycles of 94°C for 30 s, 35°C for 1 min and 72°C for 2 min; and finally one cycle of 72°C for 5 min. Ten microlitres of the amplified products were subjected to electrophoresis on a 1.4 per cent agarose gel cast in 1 × TBE and run in 0.5 × TBE at 200 V for 2.5–3.0 h. The electronic image of the ethidium bromide-stained gel was captured using a Flowgen IS500 imaging system and the bands were scored from the image displayed on the monitor. The seven Operon primers used in the RAPD analysis were C–03, C–06, C–08, C–10, C–14, F–13 and K–11. These primers were selected from a survey of over 100 decamers screened with six diverse accessions of rice, only on the basis that they revealed a large number of polymorphic bands.

Data analysis

Associations between the RAPD markers and the phenotypic means of accessions for the various quantitative traits were established using the multiple regression approach. Each quantitative trait was treated as a dependent variable and the various RAPD marker genotypes (scored as 1 for presence and 0 for absence) as independent variables. The analysis was based on the model:

$$Y = a + b_1 m_1 + b_2 m_2 + \ldots + b_j m_j + \ldots + b_n m_n + d + e$$

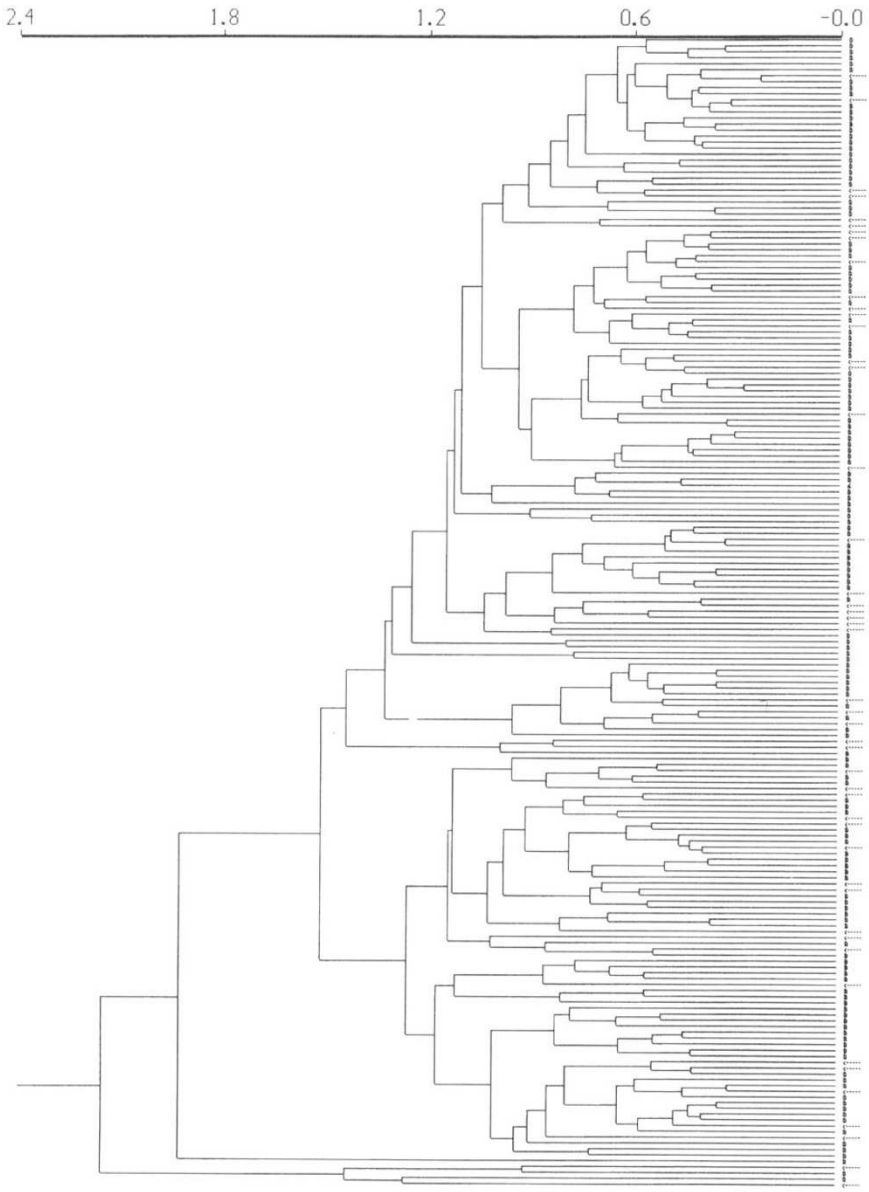

Fig. 1 Dendrogram resulting from cluster analysis of standardized morphological data from 200 accessions of rice using Squared Euclidean Distance and UPGMA clustering and from which the 48 diverse accessions (identified by bars at root of dendrogram) were selected.

which related the variation in the dependent variable (Y = accession means for a quantitative trait) to a linear function of the set of independent variables m_j, representing the RAPD markers. The b_j terms are the partial regression coefficients that specify the empirical relationships between Y and m_j, d represents the between accessions residual which is left after regression and e is the random error of Y that includes environmental variation. This method provided maximum likelihood estimates of relationships between individual quantitative traits and various markers which were then used to identify the most significant components (markers) of the best fitting multiple regression equation and to test its goodness of fit following Mather & Jinks (1982) and Draper & Smith (1981). The maximum r^2 improvement (MAXR) option of the PROC REG of the SAS statistics package (SAS, 1990) was used to determine the most appropriate model. Initially, one variable (i.e. marker) models were assessed and the marker with the highest r^2 value was identified. Then the second variable was added and the best two-variable model was selected using the usual criterion of the largest r^2. The model fitting was continued until all significant variation in Y relative to replicate variation was exhausted. A further condition was imposed

Table 1 The locations from within South and South East Asia from which accessions of rice were originally collected. The International Rice Germplasm Centre (IRGC) uses a unique number to identify each accession

Country	IRGC numbers of accessions	Range of latitudes (N)
Indonesia	66540, 66603, 66612, 66669 66678	1°7′ to 5°0′
Malaysia	71493, 71501, 71515, 71517 71537, 71544, 71545, 71578 71596, 71646	5°2′ to 6°6′
Sri Lanka	66513, 66529	6°59′ to 7°32′
Philippines	67436	14°15′
Thailand	78245, 78250, 78253, 78259 78270, 78275, 78276	14°8′ to 19°9′
Vietnam	78357	20°60′
Bangladesh	25840, 25851, 25868, 64789 64792, 64793, 66787, 66791, 66817, 77210, 77264, 77272, 77279	22°40′ to 25°38′
India	67480, 74716, 74720, 74773	22°50′ to 26°10′
Pakistan	73090	26°22′
Bhutan	64887, 64890, 64913, 67848	27°30′ to 27°48′

on the model, whereby a new variable was retained in the model only if it was significant at the 5 per cent level.

The multiple regression analyses were carried out on 40 accessions, after we had randomly removed eight from the original 48 which were to be used for testing the efficiency of the predictions. The larger sample was used in the multiple regression analysis so that the regression equations could attain sufficient statistical accuracy. Regression parameters thus estimated were used to predict the mean score of each of the eight accessions and the observed and expected values were compared using Student's *t*-test.

Although the above analysis demonstrates the predictive power of the model unambiguously, the efficiency of utilizing the information that is available from the data can be improved with the use of the 'leave out one at a time' method where multiple regression analysis is applied to 47 of the 48 accessions at a time and the 48th is predicted. This procedure allowed for the fitting of 48 models for each trait, one for each accession, and the variance of the 48 residual deviations provided the standard error of the difference between the observed and predicted values, as shown in Table 6.

Results

Altogether 63 polymorphic RAPD bands which showed consistent results over two independent runs were scored for the purposes of the present analysis (see Table 2 for further details). Figure 2 shows examples of the marker bands obtained by following the protocol presented in the Materials and methods section.

Analysis of variance revealed that the variation between the accessions was highly significant for all the quantitative traits (results not shown). Six of the 10 traits (culm number, culm length, days to 50 per cent flowering, leaf length, panicle length and grain width) which represented different facets of morphological variation in rice were selected for further analysis and a summary of some useful statistics on these traits is presented in Table 3. Tests of non-normality revealed that the mean scores of the 48 accessions were normally distributed for all the traits. Correlations between the accession means were significant for seven pairs of traits and all of these correlations were positive except the one between culm number and grain width (Table 3).

The number of bands that showed significant associations with the six traits varied from 12 for culm number and culm length to 32 for grain width (Tables 4a,b). Further, the multiple regression accounted for all the significant variation (r^2 varied from 0.89–0.99) in all the quantitative characters and only a negligible portion (d) of the quantitative variation remained unassociated with the markers. Various characters were observed to show marked differences in their association with the RAPD markers. For example, five of the 63 markers (nos 1,

Table 2 Key to polymorphic RAPD markers and the oligonucleotides used in the present study on rice

	Polymorphic markers			Polymorphic markers	
Primer	Key	Size	Primer	Key	Size
OPC-03	1	325	OPC-14	36	325
	2	350		37	390
	3	375		38	410
	4	475		39	460
	5	775		40	520
	6	2500		41	530
OPC-06	7	340		42	600
	8	355		43	650
	9	460		44	950
	10	600		45	1150
	11	1020		46	1500
	12	1600		47	1350
OPC-08	13	280	OPF-13	48	360
	14	340		49	400
	15	350		50	420
	16	500		51	340
	17	625		52	530
	18	950		53	690
	19	1050		54	880
	20	1000		55	1000
	21	2200	OPK-11	56	220
OPC-10	22	350		57	480
	23	400		58	520
	24	500		59	650
	25	520		60	1200
	26	600		61	1500
	27	625		62	2100
	28	700		63	2800
	29	750			
	30	720			
	31	950			
	32	490			
	33	1450			
	34	2200			
	35	2800			

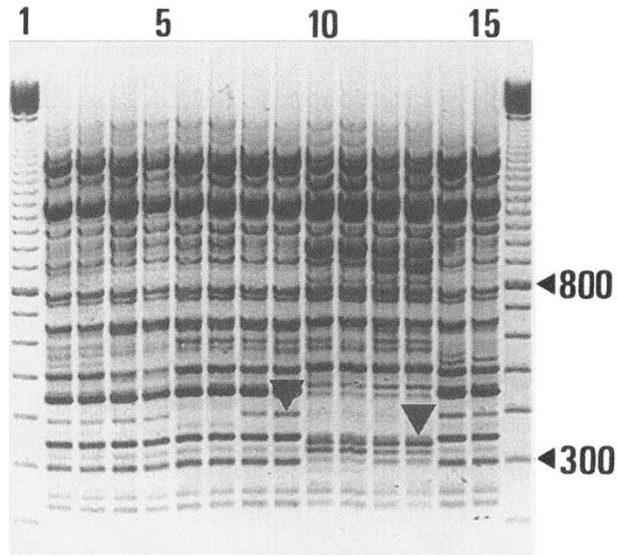

Fig. 2 Amplification products obtained using rice genomic DNA and primer OPF–13. Lanes 1 and 16 are 100 bp molecular weight markers, and the remainder are amplification products from DNA of rice accessions from Malaysia, as follows: 71515, lanes 2–3; 71517, lanes 4–5; 71537, lanes 6–7; 71544, lanes 8–9; 71545, lanes 10–11; 71578, lanes 12–13; 71596, lanes 14–15. The results show the reproducibility of the PCR for each accession and diversity between accessions. The bands arrowed (OPF–13–400 and OPF–13–340) are markers which, as shown by subsequent regression analysis, explained the variation in performance for quantitative traits (see Table 5).

7, 35, 52 and 56 in Table 2) did not associate critically with any of the six quantitative traits. At the other extreme, only one marker (OPC–08–950, marker no. 18) was critically associated with all the traits except culm number. Of the 57 remaining markers, 18 showed correlation with one, 19 with two, 16 with three and 4 with four traits, respectively.

The multiple regression analysis is capable of identifying those markers which show particularly strong associations with QTLs and for further study to demonstrate genetic linkages. However, to minimize the detection of false positives more stringent criteria had to be employed in determining the level of correlation between marker(s) and the quantitative traits. For example, the significance of a marker association was tested conservatively by multiplying its observed probability by k, the total number of markers from which the one is selected. Marker(s) with adjusted probability of $P<0.05$ were identified for this purpose and are presented in Table 5. Interestingly, one marker (OPF–13–400) explains about 24 per cent of the variation for both culm length and days to 50 per cent flowering which in turn showed significant positive correlation (0.67), suggesting that these two traits might involve one or more QTLs in common (QTLs with pleiotropic effects).

More interesting, however, were the results of our predictions of the quantitative scores of eight accessions which were excluded from the multiple regression analysis for this purpose (Table 6). Out of the 48 trait/accession combinations for which we have made predictions, in four cases the observed scores

Table 3 Overall mean, standard deviation, the maximum and minimum scores and correlations among the famly means of the 48 accessions of rice used in the present study. Measurements of culm, panicle and leaf lengths are given in cm; grain widths are in mm

	Culm no.	Culm length	Days to 50% flowering	Panicle length	Leaf length	Grain width
Characters						
Mean	32.00	101.59	83.69	25.77	53.88	3.03
SD	11.92	22.79	16.86	3.94	9.92	0.41
Max. score	71.70	145.00	124.44	32.10	77.00	4.20
Min. score	11.20	47.70	56.00	17.00	25.10	2.41
Correlations						
Culm number		−0.05	0.02	−0.16	0.10	−0.44**
Culm length			0.66**	0.69**	0.78**	0.04
Days to 50% flowering				0.62**	0.51**	0.01
Panicle length					0.53**	0.26
Leaf length						−0.07
Grain width						

**$P<0.01$.

Table 4a Regression coefficients, Fisher's test of the goodness of fit of the model and r^2 value for culm number, culm length and leaf length of rice accessions

Culm number		Culm length (cm)		Leaf length (cm)	
Marker†	Regression coefficient	Marker	Regression coefficient	Marker	Regression coefficient
5	4.97	2	−17.33	14	4.47
13	6.09	6	−16.78	15	8.04
14	−16.27	18	10.56	18	8.55
26	−6.62	19	27.96	19	10.61
33	−12.79	20	18.63	20	10.59
37	−5.61	26	−27.47	26	−13.93
41	26.43	28	8.39	28	10.52
42	−11.01	30	−16.85	30	−8.33
46	−29.98	48	−25.28	32	8.77
53	5.82	49	15.92	33	−4.56
55	−10.81	61	−26.23	39	9.18
60	−12.59	63	−24.55	41	18.60
				43	6.80
				44	9.20
				49	12.84
				61	−5.86
Intercept:	95.97		147.50		30.03
F	0.63 (27,47‡)		1.16 (27,47)		1.04 (23,47)
r^2	0.90		0.91		0.89

Each regression coefficient is significant at the $P<0.05$ level.
†See Table 1 for key to markers.
‡The analysis was performed on 40 accessions but the replication × accession mean squares of the whole experiment was used as error in the F-test and therefore has 47 d.f.

Table 4b Regression coefficients, Fisher's test of the goodness of fit of the model and r^2 value for days to 50 per cent flowering, panicle length and grain width of rice accessions

Days to 50% flowering		Panicle length		Grain width	
Marker†	Regression coefficient	Marker	Regression coefficient	Marker	Regression coefficient
2	−7.23	2	−0.62	3	−0.18
3	−8.86	6	−3.36	4	−0.48
5	−7.14	8	5.07	6	−0.21
9	−9.91	10	−5.73	8	−0.13
10	11.50	11	−5.46	9	−0.49
16	−18.97	12	−2.64	11	−1.23
17	5.20	17	−5.23	14	0.13
18	3.54	18	3.63	16	−0.38
19	16.43	21	−4.31	17	−0.31
21	−10.97	23	4.65	18	−0.52
22	25.60	24	−1.81	20	−0.07
24	−4.69	26	−2.67	21	−0.26
25	−15.28	27	5.96	22	−0.31
26	−28.51	29	−8.29	24	−0.44
27	33.54	31	1.87	25	0.42
28	7.14	37	−3.00	29	−0.63
31	−3.75	38	−2.27	31	0.30
32	−6.29	40	0.52	34	0.58
33	13.31	50	−1.47	36	−0.46
34	−8.17	53	−1.13	37	0.22
37	−22.54	54	−2.37	38	−0.72
39	14.66	55	−6.59	39	−0.24
42	10.09	61	−3.77	42	−0.45
45	−16.63	62	−3.28	44	−0.19
49	21.90	63	−2.60	47	−0.33
54	−19.65			48	−0.33
55	−20.25			49	−0.12
57	−27.18			51	−0.20
61	−22.16			53	0.24
				54	−0.09
				58	0.21
				59	−0.53
Intercept:	150.15		53.01		7.25
F	0.32 (10,47‡)		0.83 (14,47)		0.11 (7,47)
r^2	0.99		0.98		0.99

Each regression coefficient is significant at $P<0.05$.
†See Table 1 for key to markers.
‡See footnote to Table 4a.

differed significantly from the predicted. But, in these four cases the significance was marginal and the rank correlation between the observed and predicted values of each trait was high (mean correlation = 0.85), the lowest being 0.74 for culm number and the highest 0.95 for grain width. In other words, association of quantitative variation with molecular markers has provided good predictions of the performance of these accessions. This is ideal for the purpose of identifying useful accessions prior to assessing their phenotypic performance in the field. Moreover, the regression model was even able to predict the number of culms of accession IRGC 78275, whose score actually falls outside the range observed among the 40 accessions used in making the predictions.

In the case of 'leave out one at a time' analysis, altogether seven predictions differed significantly

from the observed. These failures constitute only about 2.4 per cent of the 288 accession/trait combinations to which models were fitted and in no case were more than two failures noted for any trait. As up to 5 per cent of the predictions can in theory be expected to differ from the observed values, there is little doubt that our predictions are highly reliable.

Discussion

Molecular markers are increasingly being used in marker-assisted selection programs (Stomberg *et al.*, 1994). Both theoretical and experimental studies have shown that marker-assisted selection can be highly effective for producing improved genotypes. However, the success of such selection programmes depends exclusively on the extent of genetic linkage between markers and the relevant loci such as QTLs. Although such studies are invariably based on materials derived from planned crosses there is no reason why the same principles cannot be applied to either natural populations or genetic resources generally, assuming that similar associations are observed between marker loci and the various allelomorphic forms of QTLs, and that the basis of these is in fact genetic linkage. In this case there is no reason why the regression technique, which is generally used to predict the response to selection, cannot be used to identify the best genotypes within a segregating population or a collection of landraces/varieties, as we have shown in the present study.

In addition to linkage, there could be other causes of the association which we have found between dimorphic markers and polymorphic diversity. These include linkage disequilibrium involving chance associations resulting from correlated allele frequencies in small samples. However, linkage between markers and QTLs is most likely to be responsible for the present results, although this can only be demonstrated unambiguously by analysing populations derived from single crosses between two parents. If true, it appears that linkage between alleles at QTLs and at marker loci has been conserved throughout the period of diversification of rice germplasm in South and South East Asia.

If genetic linkage is the main cause of the associa-

Table 5 RAPD marker bands that explain the largest proportion of variation in performance for each of the six quantitative traits measured in rice

Trait	Marker	r^2 (%)
Culm number	OPC-08-340	49.84
	OPC-10-600	5.78
Culm length	OPF-13-400	24.25
Days to 50% flowering	OPF-13-400	23.84
Panicle length	OPK-11-2800	36.05
	OPC-08-2200	6.17
Leaf length	OPC-08-1000	19.85
Grain width	OPF-13-340	29.70
	OPC-14-410	5.30

Table 6 The observed and predicted performances of eight accessions of rice for each of the six quantitative traits

			Accession numbers							
Traits		SE†	64792	64890	66603	67480	67848	71596	74716	78275
Culm number	Observed	5.75	22.70	19.67	38.10	41.00	34.70	22.50	43.40	11.22
	Predicted		18.83	19.06	40.86	30.06	37.37	20.19	36.02	14.10
Culm length (cm)	Observed	10.02	110.50	85.67*	88.60	69.20	94.30	112.10	68.60	69.22
	Predicted		115.48	105.93*	98.27	82.11	88.14	106.31	63.59	79.70
Days to 50% flowering	Observed	2.54	76.80	67.00	94.40	65.40	72.70*	91.20	59.40	74.50
	Predicted		73.88	67.61	93.46	68.20	79.58*	89.53	59.97	77.96
Panicle length (cm)	Observed	1.56	25.78	22.30	25.80	17.40	26.20	28.40	19.20	26.10
	Predicted		27.14	19.37	27.43	16.90	27.14	28.57	20.15	24.20
Leaf length (cm)	Observed	5.74	56.20	39.00	52.80	51.80	42.10	56.00	51.50	46.56
	Predicted		54.21	40.38	49.04	57.59	36.47	64.25	45.74	41.32
Grain width (mm)	Observed	0.08	3.26*	3.34	2.50	2.46	2.91	3.00*	2.89	3.92
	Predicted		3.09*	3.43	2.45	2.57	3.00	3.17*	2.79	3.94

†Standard error (see text for further details) of difference (Predicted − Observed).
*Significantly different at $P<0.05$.

tions, then one benefit of obtaining information about molecular markers and quantitative traits could be for use in more efficiently selecting putative parents for producing populations to map QTLs for a particular trait. As an example from our data, if we wanted to map QTLs for culm length, two accessions (namely IRGC 71646 and 74720) not only represent phenotypic extremes for this trait but are also polymorphic for the key target bands and belong to diverse groupings based on data from both molecular (63 markers) and all 10 quantitative traits scored (data not shown).

The procedure can also be used as an initial screening method for the identification of QTLs. The established method for this is the selection of two parents that differ markedly in a particular quantitative character, and then the determination of associations between markers and that character in F_2 or backcross progeny. The apparent advantages of using diverse germplasm instead are (i) that this could allow the detection of QTLs that vary across a wide spectrum of biodiversity rather than just between two parental lines, and (ii) that QTLs for any quantitative trait can be studied in the same investigation.

Regardless of the underlying causes of the associations which we have detected, the use of molecular markers such as RAPD, which are more or less randomly distributed across the genome (Kurata *et al.*, 1994), coupled with multiple regression analysis could substantially change and improve the way in which crop biodiversity is used in the future. The combination of techniques should allow us to predict what a plant will look like in terms of quantitative agronomic traits prior to elaborate field trials. If a diverse test array of rice germplasm is scored for important traits requiring specialized assessment conditions (such as stress tolerances) then marker data can provide an efficient means of predicting the value of additional germplasm for such characteristics, and even identifying suitable material amongst germplasm *in situ*.

Our results therefore demonstrate the value of *ex situ* plant germplasm collections not just as repositories of useful genes, but also as sources of information about phenotypic characters. One of the major criticisms which has regularly been levelled at genetic resource conservationists over the last 40 years is that they have frequently been unable to provide appropriate material for crop improvement programmes. However, with appropriate organization of conserved material and the application of current DNA-based marker technology, genebanks can more easily counter these criticisms and become much more valuable interfaces between the activities of conservationists on the one hand and those wishing to exploit germplasm for the benefit of humankind on the other.

Acknowledgements

We are grateful to an anonymous reviewer for suggestions to improve the statistical analysis and Alan Paysan and Hipolito Elec for their assistance with the field measurements. This work was funded by the Plant Sciences Programme of the Overseas Development Administration, U.K. (grant number R5059).

References

DRAPER, N. AND SMITH, H. 1981. *Applied Regression Analysis*, 2nd edn. John Wiley and Sons, 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.

IBPGR-IRRI RICE ADVISORY COMMITTEE 1980. *Descriptors for Rice* (*Oryza sativa* L.). International Rice Research Institute, Manila, Phillipines.

JACKSON, M. 1994. Preservation of rice strains. *Nature*, **371**, 470.

JACKSON, M. AND HUGGAN, R. 1993. Sharing the diversity of rice to feed the world. *Diversity*, **9**, 22–25.

KURATA, N., NAGAMURA, Y., YAMAMOTO, K. *et al.* 1994. A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genetics*, **8**, 365–372.

MATHER, K., AND JINKS, J. L. 1982. *Biometrical Genetics*, 3rd edn. Chapman and Hall, London.

ROHLF, F. J. 1992. NTSYS-PC: numerical taxonomy and multivariate analysis system. Exeter Software, New York.

SAS 1990. *SAS/STAT User's Guide.* Version 6, 4th edn, vol. 1. SAS Institute Inc., Cary, NC.

STOMBERG, L. D., DUDLEY, J. W. AND RUFENER, G. K. 1994. Comparing conventional early generation selection with molecular marker assisted selection in Maize. *Crop Sci.*, **34**, 1221–1225.

VIRK, P. S., FORD-LLOYD, B. V., JACKSON, M. T. AND NEWBURY, H. J. 1995a. Use of RAPD for the study of diversity within plant germplasm collections. *Heredity*, **74**, 170–179.

VIRK, P. S., NEWBURY, H. J., JACKSON, M. T. AND FORD-LLOYD, B. V. 1995b. The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. *Theor. Appl. Genet.*, **90**, 1049–1055.

WELSH, J. AND MCCLELLAND, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.*, **18**, 7213–7218.

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**, 6531–35.