Theor Appl Genet (1995) 91:978–986

S. Nair · U. Prasada Rao · J. Bennett · M. Mohan

Detection of a highly heterozygous locus in recombinant inbred lines of rice and its possible involvement in heterosis

Received: 24 April 1995 / Accepted: 12 May 1995

Abstract Forty-seven recombinant inbred (RI) lines derived from a cross between two indica rices, cv 'Phalguna' and the Assam land race ARC 6650, were subjected to restriction fragment length polymorphism (RFLP) analysis using cloned probes defining 150 single-copy loci uniformly dispersed on the 12 chromosomes of rice. Of the probes tested, 47 detected polymorphism between the parents. Heterozygosity was calculated for each line and for each of the polymorphic loci. Average heterozygosity per line was 9.6% but was excessive (> 20%) in the 5 lines that seemed to have undergone outcrossing immediately prior to harvest. Average heterozygosity detected by each probe across the 47 RI lines was 9.7%. The majority of probes revealed the low level of heterozygosity (< 8%) expected for F_5-F_6 lines in a species showing about 5% outbreeding. On the other hand, 7 probes exhibited heterozygosity in excess of 15%, while with a eighth probe (RG2 from chromosome 11) heterozygosity varied according to the restriction enzyme employed, ranging from 2% with *Sa*II to 72% with *Eco*RV. The presence of 34 recombination sites in a segment of the genome as short as 24 kb indicates a strong selection for recombination between two neighbouring loci, one required as homozygous for the 'Phalguna' allele, and the other heterozygous. Since selection was principally for yield advantage over that of the high-yielding parent, 'Phalguna', one or both of these loci may be important for heterosis in this cross. The results also indicate that heterozygosity as measured by RFLP can depend on the particular restriction endonuclease employed.

Key words Heterozygosity · *Oryza sativa* · Heterosis · RFLP · Recombinant inbred lines

Communicated by P. M. A. Tigerstedt

S. Nair · J. Bennett[1] · M. Mohan (✉)
International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi-110 067, India

U. Prasada Rao
Directorate of Rice Research, Rajendranagar, Hyderabad-500 030, India

Present address:
[1] International Rice Research Institute, P. O. Box 933, Manila, Philippines

Introduction

Restriction fragment length polymorphism (RFLP) analysis is used to map major genes and quantitative trait loci (QTL) and to follow gross changes in chromosome structure such as those that occur during introgression of genes from wide crosses (Snape et al. 1985; Burr et al. 1988; Burr and Burr 1991; Ellis et al. 1992; Reiter et al. 1992; Lister and Dean 1993; Causse et al. 1994). Populations suitable for mapping include F_2, doubled haploids (DH) from F_1, backcross 1 (BC_1) and recombinant inbred (RI) lines (Burr et al. 1988). In addition to their employment in the mapping of new markers, RI lines can also be used to map any phenotypic trait that is polymorphic between the two parents (Mohan et al. 1994). RI lines have been actively exploited to map disease resistance genes, for example resistance to the fungus *Peronospora parasitica* (Parker et al. 1993).

The ideal mapping population must satisfy a set of criteria, including the absence of biases, such as those consequent upon selection for desirable phenotypes. This means that lines produced by commercial breeding programmes are unlikely to be totally satisfactory for general mapping. Nevertheless, RFLP analysis can be applied to breeding lines. For this we need to set criteria for assessing whether such sets of lines are suitable: (1) do we have enough lines? (2) are the parents sufficiently different from each other? (3) do the parents show RFLPs? (4) are they consistently reproducing the same RFLP pattern?

One of the advantages of RI lines over F_2 or BC_1 populations in RFLP mapping is that in principle the former represent a genetically stable record of past

segregation events, thereby permitting their indefinite use for the mapping of many loci. The level of residual segregation can be made as low as desired by continuous inbreeding. However, in practice, heterozygosity is not eliminated entirely from RI lines because of the time required for exhaustive inbreeding and the difficulty of ensuring a total absence of selection for balanced polymorphisms. Indeed, it is possible to exploit RI lines prepared under intensive selection to focus attention on loci that are heterozygous for interesting reasons. The drive towards halving of heterozygosity at each locus with each generation is so strong that loci which remain heterozygous for many generations in many lines must do so as a result of strong selection for a locus for their retention. Such a locus may be identified by RFLP analysis if (1) the heterozygous region contains a suitable RFLP marker; and (2) it is very tightly linked to a RFLP marker. The ideal situation would be to also know the function of the locus in question, especially in the case of loci that might be of interest because of their contribution to heterosis or yield advantage attributable to hybrid vigour. However, in most cases the precise function of a heterozygous locus may be difficult to uncover if it contributes to a quantitative trait. Such loci are interesting for what they can teach us about heterosis, balanced polymorphisms, hybrid viability and the mechanism of homologous recombination in plants.

Patterns of restriction fragment length polymorphism have been proposed as estimators of genetic diversity among breeding lines and as predictors of heterosis and genetic variability. In general, results have suggested that heterosis for forage yield, fertility and tuber yield increase progressively with the theoretical levels of heterozygosity (Bonierbale et al. 1993). A common finding from many studies (Smith et al. 1990; Dudley et al. 1991; Stuber et al. 1992) has been that heterosis is significantly correlated with the heterozygosity of marker loci but that the level of correlation varies from one data set to another. Recently, molecular genetic markers have been used to measure multilocus heterozygosity in maize. Melchinger et al. (1990) found a poor association between their RFLP-based estimate of genetic distance and heterosis in crosses of unrelated maize lines. However, marker-based genetic distance estimates and heterosis are both positively correlated with genetic divergence within a set of related parents.

We report herein the identification of a highly heterozygous locus that is possibly correlated to heterosis in high-yielding RI lines. It is located within 24 kb of the marker RG2 on chromosome 11 of rice. Detection of this locus was dependent upon our decision to digest the DNAs of the parents and RI lines with various 6-base recognition restriction endonucleases and to carry out co-segregation analysis of the RG2 locus in the RI lines with all of the enzymes that revealed polymorphisms between the parents for this locus. The results also indicate that heterozygosity as measured by RFLP analysis can depend on the particular restriction endonuclease employed.

Materials and methods

Recombinant inbred lines

Forty-seven RI lines were derived by the method of single-seed descent (SSD) from a cross between 'Phalguna', a high-yielding parent, and ARC6650, a land race, in 1978 and have since been subject to periodic selfing where some outcrossing between different inbred lines might have occurred. All of the RI lines used in this study were of the F_5-F_6 generations and have been primarily subjected to selection for high yield in every generation.

Probes

RFLP probes

A total of 150 single-copy DNA probes distributed over the 12 chromosomes of rice were selected for analysis. These clones (the RG series) were originally selected from a *Pst*I genomic library of rice (McCouch et al. 1988; McCouch and Tanksley 1991) and were kindly provided by Dr. S. D. Tanksley of Cornell University, USA.

Random Amplified Polymorphic DNAs (RAPD) and other Probes

The RAPD markers (F8, F10, F19 and D7) employed were generated from a study on the tagging of an insect resistance gene that segregates in the RI lines. RAPD amplifications were carried out according to the published method (Williams et al. 1990). Details on the generation of the probes are described in Mohan et al. 1994. The *Salt T* gene (Claes et al. 1990) was also used as a RFLP marker. This gene was amplified directly form the genomic DNA of rice variety 'TN1' using *Salt T*-specific oligonucleotides.

DNA extraction and Southern hybridization

Rice seeds were germinated in the dark, and 14-day-old seedlings were harvested and kept frozen at − 80 °C until needed for DNA extraction by the method of Walbot (1988). Rice DNA (5 or 10 µg) was digested with restriction endonucleases and electrophoresed overnight at 25 V on 0.8% agarose gels in 1 × TBE buffer (Sambrook et al. 1989). The subsequent Southern transfer and hybridization were as described before (Williams et al. 1991). In the initial survey for RFLPs between the two parental DNAs, the following restriction enzymes were used: *Alu*I, *Bam*HI, *Bg*1II, *Eco*RI, *Eco*RV, *Hin*dIII, *Not*I, *Xba*I. Probes which did not reveal an RFLP with these enzymes were then hybridized to filters with parental DNAs digested with the following enzymes: *Aat*II, *Ava*I, *Bcl*I, *Cla*I, *Hin*cII, *Kas*I, *Kpn*I, *Msc*I, *Nco*I, *Nde*I, *Nla*IV, *Sal*I and *Xho*I.

Co-segregation analysis

The Mapmaker Macintosh Version 1.0 programme (E.I. duPont de Nemours and Company, Copyright 1990), based on the Mapmaker programme of Lander et al. (1987), was used for the analysis of segregation data. As recommended by Burr and Burr (1991), genetic distances were calculated from recombination percentages using the Haldane and Waddington (1931) formula to take into account the additional meioses involved in the production of RI lines compared with F_2 individuals.

Results

General RFLP survey of RI lines

Forty-seven RI lines derived from the 'Phalguna' × ARC6650 cross were examined for heterozygosity at 47 polymorphic loci out of 150 tested (covering all 12 chromosomes of rice) as defined by DNA markers mapped by McCouch et al. (1988), Causse et al. (1994) and Mohan et al. (1994). The yield of most of the lines was higher than that of the high-yielding parent, 'Phalguna', while in some RI lines the yield was equivalent to that of the latter. The upper limit of the range of RI lines for panicle length, grain weight, and yield per hectare more or less reached the heterotic level and often exceeded the value (in yield per hectare) of the higher yielding parent by 10–30% (data not included).

Table 1 lists the heterozygosity and parental bias of all 47 loci. With respect to the latter, most loci appeared to be reasonably unbiased with 30–70% of lines designated as either A (homozygous for ARC6650 allele) or P

Table 1 Heterozygosity and parental bias as detected by 47 polymorphic RFLP markers (*A* homozygous for ARC6650 allele, *P* homozygous for 'Phalguna' allele, *H* heterozygotes)

Locus	A	P	H	Missing data	%Heterozygosity
D7	18	19	6	4	12.7
F8	26	18	0	3	0.0
F10	28	19	0	0	0.0
F19	12	30	4	1	8.50
RG2 (*Sal*I)	10	36	1	0	2.120
RG2 (*Hin*dIII)	11	27	9	0	19.1
RG2 (*Nde*I)	11	19	14	3	29.7
RG2 (*Eco*RI)	12	5	30	0	63.8
RG2 (*Eco*RV)	11	1	34	1	72.3
RG13	6	39	2	0	4.20
RG16	12	25	10	0	21.20
RG103	14	31	2	0	4.20
RG118	19	20	2	6	4.20
RG120	5	22	13	2	27.6
RG123	30	12	2	3	4.20
RG152	35	9	2	1	2.12
RG167	11	29	3	4	6.38
RG171	6	26	11	4	23.4
RG182	16	26	5	0	10.6
RG190	19	19	6	3	12.70
RG207	29	17	1	0	2.12
RG213	27	14	5	1	10.6
RG214	21	18	8	0	17.02
RG235	13	27	9	0	14.90
RG241	12	30	5	0	10.60
RG257	14	26	3	4	6.38
RG329	21	14	12	0	25.5
RG333	8	39	0	0	0.0
RG341	13	30	3	1	6.38
RG365	37	10	0	0	0.0
RG381	5	41	1	0	2.12
RG396	19	18	7	3	14.90
RG424	28	13	2	4	4.20
RG435	39	7	0	1	0.0
RG456	29	14	1	3	2.12
RG463	16	27	4	0	8.50
RG472	15	26	3	3	6.38
RG476	21	19	7	0	14.90
RG520	10	30	0	7	0.0
RG528	12	33	2	0	4.20
RG532	17	26	4	0	8.50
RG543	16	25	6	0	12.70
RG553	7	34	6	0	12.70
RG570	5	29	13	0	27.60
RG573	8	36	3	0	6.38
RG620	24	14	7	2	14.90
RG648	30	13	2	2	4.20
RG667	7	25	15	0	31.90
RG776	23	17	7	0	14.90
RG869	13	30	3	1	6.38
SaltT	25	15	7	0	14.90

(homozygous for 'Phalguna' allele). However, RG152, RG365 and RG435 loci were heavily biased towards ARC6650 (> 70% A), while loci RG2 (RG2/*SalI*), RG13, RG333, RG381, RG528, RG553 and RG573 were heavily biased towards 'Phalguna' (< 30% A). These loci may be linked to genes defining important parental traits selected during the breeding programme. Likewise, for heterozygosity, the majority of loci approached the expected values of 3–6% for the F_5–F_6 generations. However, some loci, RG16, RG120, RG171, RG329, RG570 and RG667, greatly exceeded the expected value with a heterozygosity of more than 20%, while locus RG2 (RG2/*Eco*RV) was truly exceptional in showing 72% heterozygosity.

Bias and heterozygosity were also assessed for each RI line (Table 2). Most lines showed low bias and low heterozygosity and should be considered as a reasonably acceptable population for RFLP analysis. However, lines 3 and 44 were heavily biased towards ARC6650 (> 70% A); lines 9, 14, 15, 19, 24, 37, 38, 41, 43, 45 and 47 towards 'Phalguna' (< 30% A) and lines 1, 2, 10, 11, 12, 16, 17, 31, 32, 34 and 42 were markedly heterozygous (> 15%). These lines would appear to have undergone intercrossing with other RI lines while they were grown together, thereby resulting in outcrossing during the inbreeding process.

Detailed analysis of locus RG2

A DNA probe can reveal high heterozygosity if it hybridizes to a heterozygous locus or if it is tightly linked

Table 2 Bias and heterozygosity in RI lines as detected by 47 polymorphic RFLP probes (*A* homozygous for ARC6650 allele, *P* homozygous for 'Phalguna' allele, *H* heterozygotes)

RI Lines	A	P	H	Missing data	%Heterozygosity
1	23	14	10	0	21.2
2	18	13	16	0	34.0
3	33	14	0	0	0.0
4	19	15	2	11	4.2
5	22	24	1	0	2.1
6	21	20	6	0	12.7
7	15	25	6	0	12.7
8	29	18	0	0	0.0
9	7	40	0	0	0.0
10	12	23	10	2	21.2
11	17	10	20	0	42.5
12	12	16	19	0	40.7
13	28	14	5	0	10.6
14	1	42	4	0	8.5
15	1	44	2	0	4.2
16	14	21	12	0	25.5
17	17	22	8	0	17.0
18	27	14	3	3	6.3
19	1	46	0	0	0.0
20	29	11	7	0	14.8
21	25	21	1	0	2.1
22	26	19	2	0	4.2
23	28	11	2	6	4.2
24	6	39	2	0	4.2
25	24	22	0	1	0.0
26	23	16	1	7	2.1
27	15	18	4	10	8.5
28	26	17	3	1	6.3
29	28	17	2	0	4.2
30	27	18	2	0	4.2
31	20	18	9	0	19.1
32	18	18	11	0	23.4
33	29	13	4	1	8.5
34	27	10	9	1	19.1
35	23	22	2	0	4.2
36	19	23	5	0	10.6
37	1	46	0	0	0.0
38	1	45	1	0	2.1
39	16	20	4	7	8.5
40	18	23	5	1	10.6
41	5	40	2	0	4.2
42	18	21	8	0	17.0
43	1	45	0	1	0.0
44	33	14	0	0	0.0
45	1	45	1	0	2.1
46	23	23	1	0	2.1
47	4	32	1	10	2.1

Table 3A Heterozygosity at locus RG2 varies with restriction enzyme used to restrict the genomic DNAs

Locus	Scores[a]	Heterozygosity (%)
RG2 (*Eco*RV)	HHAHHHHHHHAHAHHHHAHAAHAHA-HHHHHHAAHHHHHHHHHAHHP	72.3
RG2 (*Eco*RI)	HHAHHPHHHPAHAHHHHAHAAHAPAHHHHHHHAAHHHHHPHHAAHHP	63.8
RG2 (*Nde*I)	HHA-HPPHHPAHAHHHPA-AAPAPAHPPHPPHAAHPPPPPPP-APHP	29.7
RG2 (*Hin*dIII)	HHAHHPPPPPAHAPPPPAPAAPAPAHPPHPPPAAHPPPPPPPPAPHP	19.1
RG2 (*Sal*I)	PPAHPPPPPPAPAPPPPAPAAPAPPPPPPPPPAAPPPPPPPPPAPPP	2.1

to such a locus. In favourable cases these two possibilities may be distinguished by performing RFLP analysis with several different restriction enzymes together with the specific DNA probe in question. If the probe hybridizes to a heterozygous locus, it should reveal the same degree of heterozygosity with each restriction enzyme, but if it hybridizes to a basically homozygous locus a short distance away from a heterozygous locus, the degree of heterozygosity is likely to vary with the restriction enzyme. In the case of RG2, heterozygosity was observed to vary from 72% with *Eco*RV to 2% with *Sal*I (Table 3A), other restriction enzymes revealing an RFLP between 'Phalguna' and ARC6650 with RG2 gave intermediate values for heterozygosity. Thus, while locus RG2 is basically homozygous, it is tightly linked to a heterozygous locus.

The distance between locus RG2 and the heterozygous locus would be expected to be different for each RI line and be determined by the exact site of homologous recombination between the two loci in each line. The factors involved are illustrated in Fig. 1. In Figure 2 we show how two RI lines (X and Y) might differ near locus RG2. For each of the 3 restriction enzymes shown, *Eco*RV, *Eco*RI and *Hin*dIII, we have arbitrarily assumed that an insertion in 'Phalguna' or ARC6650 is responsible for the RFLP that distinguishes the ARC6650 allele from the 'Phalguna' allele for that

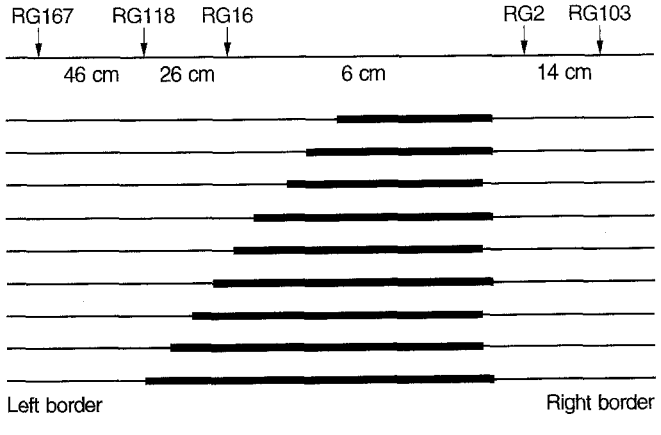

Fig. 1 Diagrammatic representation of the heterozygous region with reference to locus RG2 and neighbouring RFLP loci on chromosome 11. Each *horizontal line* is a graphical representation of a RI line: *bold lines* represent the area of heterozygosity; *thin lines* flanking the bold lines are homozygous regions. Distances shown [in centiMorgans (cM)] are not to scale

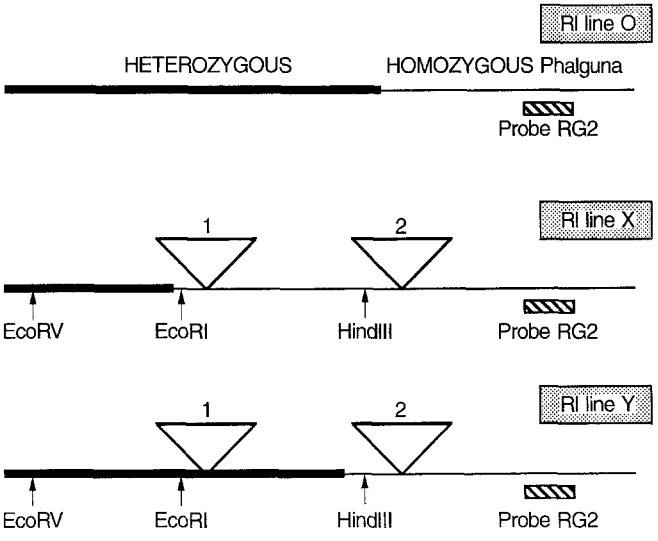

Fig. 2 RFLP probe RG2 reveals different levels of heterozygosity depending on the restriction endonuclease employed to restrict the genomic DNAs. Each *horizontal line* is a graphical representation of a RI line: *bold lines* represent the area of heterozygosity; *thin lines* flanking the bold lines are homozygous regions. *RI line O* represents a line showing the maximum heterozygosity possible. *X* and *Y* represent two RI lines with different levels of heterozygosity. *Triangles 1* and *2* depict insertions leading to RFLPs between the two parents ARC6650 and 'Phalguna'. *Vertical arrows* show restriction sites. Both lines X and Y reveal heterozygosity when restriction enzyme *Eco*RV and probe RG2 are used. With *Eco*RI and probe RG2 line X is homozygous for the 'Phalguna' allele, while Y is heterozygous since it cuts in the heterozygous region; with *Hin*dIII and probe RG2 both lines are homozygous for the 'Phalguna' allele

enzyme. The argument is not affected in essence if other types of mutations are responsible for the RFLPs. What is not arbitrary is the depiction of different restriction enzyme sites and the placement of all three sites within the heterozygous region. Insertion allows restriction enzyme *Eco*RV to detect the heterozygous region of lines X and Y. However, only the heterozygous region of line Y will be detected by restriction enzyme *Eco*RI, and restriction enzyme *Hin*dIII will detect neither heterozygous region, even though it restricts within the heterozygous region.

This phenomenon may be seen diagrammatically for locus RG2 by looking closely at the pattern of heterozygous lines revealed by each restriction enzyme (Table 3A). Notice that the 30 heterozygous lines as revealed by *Eco*RI are a subset of the 35 RI lines revealed by *Eco*RV. Similarly, the heterozygous RI lines revealed by the other 3 restriction enzymes are progressively smaller

nested sets of the 30 heterozygous lines detected by *Eco*RI. Thus, it is evident that RG2 identifies a junction between a heterozygous region and a region homozygous for the 'Phalguna' allele.

With respect to the heterozygosity detected by other loci neighbouring RG2, RG16 detected about 21%, which is also high for F_5–F_6 lines; this decreased to 4% with RG118 and 6% with RG167. It is evident that the region between RG2 and RG16 is highly heterozygous and is about 6 cM apart. The marker RG103, which is on the other side of RG2, also detected a low level of heterozygosity (Table 3B) that was within the expected range.

Table 4 summarizes the data on the survey for RFLPs between ARC6650 and 'Phalguna' using different restriction enzymes and on subsequent hybridization to probe RG2. Though the enzyme *Eco*RI produced the largest fragment (24 kb), it was *Eco*RV that revealed the maximum number of heterozygotes as *Eco*RV restricts deeper into the heterozygous region (Fig. 2). This figure is further supported by double digests of parental DNAs with *Hin*dIII in combination with *Eco*RI, *Eco*RV, *Nde*I and *Sal*I (data not shown).

Discussion

RFLP analysis of the 47 RI lines derived from a cross between 'Phalguna' and ARC6650 indicates that most lines and most loci are relatively unbiased and show low heterozygosity. The lines would thus appear to constitute a suitable population for the mapping of major genes governing interesting phenotypes. Indeed, we have used this population to map *Gm2*, a gene conferring resistance to biotype 1 of gall midge (*Orseolia oryzae*), the major dipteran pest of rice (Mohan et al. 1994). However, the small number of RI lines available precludes their use for mapping quantitative trait loci.

Given the close-to-ideal behaviour of most of the lines and loci examined here, the loci displaying high heterozygosity cannot be dismissed as being indicative of insufficient inbreeding or adventitious outcrossing in the lines on a whole. Rather, these loci stand out as candidates for maintenance of the heterozygosity through selection. Certainly, it is not credible that the heterozygosity seen at locus RG2 is ascribable to chance. The probability that locus RG2 is heterozygous in 1 F_5 line is 1/16, while the probability that it is heterozygous in 34 out of 47 RI lines is $(1/16)^{34} \times (15/16)^{13}$ or about 4.9×10^{-42}. For F_6 lines the probability would be smaller still. We conclude that RG2 is subject to a very high selection in order to remain in the heterozygous state in this cross.

Heterozygosity declines rapidly in early generations, as in the case of complete selfing, and declines to an equilibrium limit that is relatively insensitive to the degree of linkage and almost entirely determined by the frequency of crossing. The overall frequency of recombinant homozygotes is reduced by crossing, but the frequency itself increases in successive generations to an equilibrium limit that is independent of the level of linkage and that depends entirely on the frequency of the crossing (Louw 1994). Reductions in the frequency of recombinant homozygotes due to crossing are highest for independent loci and loose linkage. However, increase in the frequency of crossing from 10% to 50% leads to a marked reduction in the frequency of recombinant homozygotes. The initial heterozygosity of polymorphic loci is 1/2 and declines by 1/2 in each successive generation, assuming a large sample size and no selection. The decline in heterozygosity during propagation by SSD is reduced by cross-fertilization, which may occur as a natural characteristic of the species or through uncontrolled insect activity during flowering. Again, from the point of view of practical plant breeding using propagation by SSD where some uncontrolled crossing might occur, the main interest is in expected changes in heterozygosity and the frequency of recombinant homozygotes during the normal time span involved in the development of new lines in such experiments. The investigation by Louw (1994) on the effect of

Table 3B Heterozygosity of 4 other loci in the vicinity of the RG2 locus

Locus	Scores	Heterozygosity (%)
RG103	PHAAAPPPPPAPAPPPPPAPAAPHPAAPPPPPPAAAPPPPPPPPPPAP	4.2
RG2 (*Eco*RV)	HHAHHHHHHHAHAHHHHAHAAHAHA–HHHHHHAAHHHHHHHHHAHHP	72.3
RG16	HHAHHPPAPPAHAPPPPAPAAPAPAHPPHPPHAAHPPPPPPPPAPHP	21.2
RG118	HHA–PP–PAPAPAAAPP–AAPPAPP–APPPPAAAPPAA–PAPAAAP–	4.2
RG167	HPA–PPPPPPAHAPPPPAPAAP–PAAPPPPPPAAHPPP–PPPPAPP–	6.3

[a] For definitions of A, H and P, see legends of Tables 1 and 2, – missing data

Table 4 Molecular weights of polymorphic DNA fragments as detected by RFLP probe RG2 in the two parental genomic DNAs digested with various restriction endonucleases

Restriction enzyme	Size of fragment (kb)	
	ARC6650	Phalguna
*Eco*RV	8.0	21.0
*Eco*RI	22.0	24.0
*Nde*I	6.0	8.5
*Hin*dIII	5.5	3.0
*Sal*I	24.0	22.0

uncontrolled random crossing during the development of lines by the SSD method concentrated on two effects having major consequences with respect to practical plant breeding – the persistence of heterozygosity in advanced generations and the reduction in the frequency of recombinant homozygotes. The former has obvious implications for the genetic purity of lines at stages when selections may be made for commercial seed production, while uncontrolled crossing reduces the response to selection by reducing the frequency of homozygous recombinant lines.

The above results notwithstanding, some loci showed much more heterozygosity than expected for F_5–F_6 inbred lines. In particular, the RG2 locus is highly heterozygous when judged by its hybridization to genomic DNAs digested with *Eco*RV or *Eco*RI. However, the fact that the *Sal*I digest shows only 2% heterozygosity when probed with RG2 indicates that the RG2 locus is essentially homozygous but is tightly linked to a neighbouring heterozygous locus. The distance between the 2 loci varies among RI lines in accordance with the site of recombination that occurred during development of the lines. A low resolution map indicates that in some cases the crossing-over site is within 21 kb of the RG2 locus while in others it may be within 24 kb.

Given the low probability that F_5 or F_6 lines will exhibit heterozygosity at a given locus, it is clear that for 34 out of 47 RI lines to be heterozygous at a locus near locus RG2, some form of selection must be involved. During the breeding programme there was no indication of large-scale seed mortality, so it is unlikely that heterozygosity is high because of the death of individuals homozygous at this locus. It is more likely that retention of the heterozygotes was favoured during the breeding process. Since the major selection factor was for yield advantage over the high-yielding parent, 'Phalguna', it is possible that the heterozygous locus contributes to heterosis. The yield improvement of promising RI lines is encouraging, with some having a yield advantage of up to 20% over the better parent. This improvement is brought about through the combination of the desired traits in RI lines. It is worth mentioning here that hybrid rice outyields conventional varieties by 15–20% (Yuan and Virmani 1988). In this regard, highly significant correlations were observed between RFLP-based genetic distances, F_1 yield and heterosis in a group of 37 inbred maize lines representing a wide range of pedigree relationships (Smith et al. 1990). These researchers concluded that RFLP-based distance estimates, combined with pedigree and quantitative trait locus information, could accurately predict the best parental combinations for the production of high-yielding hybrids. However, even if marker loci are selected for their close proximity to a QTL, the relationship between heterozygosity at the molecular-marker loci (i.e. RFLP-based genetic distance) and heterosis or genetic variance will also depend on the extent of linkage disequilibrium between the marker loci and QTL in the germ plasm under consideration (Charcosset et al. 1991). Moser and Lee (1994) are of the view that dispersed markers are unlikely to predict heterosis or population genetic variance in oats. In their opinion, to accurately predict heterosis or genetic variance from molecular data, each allele at a QTL must be linked to a unique marker allele. The validity of this assumption will vary with the reference population or the source of germ plasm of the parents under consideration. Heterozygosity was found to be significantly correlated with several attributes of performance and heterosis (Zhang et al. 1994). The average grain yield of a maize plant was found to increase with increasing number of heterozygous marker loci in the genome (Edwards et al. 1987). While the association of heterosis with heterozygosity is clear it has long been debated why heterozygosity results in heterosis. It could be explained by the dominance theory of heterosis or overdominance theory. According to Xiao et al. (1995), dominance is the major genetic basis of heterosis in rice.

Alternatively, we could be seeing compensatory selection for two neighbouring loci, one preferable as the 'Phalguna' allele and the other as the ARC6650 allele. In the absence of crossing-over between the loci, the heterozygotes would be favoured. Of the 47 RI lines 12 showed no heterozygosity with probe RG2 for any enzyme tested. These lines could have experienced such a cross-over event, and the 2 loci could have become homozygous for the appropriate parental allele, in which case the predominance of the ARC6650 allele among these 12 lines would indicate that the locus favoured as the 'Phalguna' allele is proximal to the RG2 locus, while the locus favoured as the ARC6650 allele is distal. Alternatively, all 47 lines might be heterozygous near locus RG2, but the recombination event for these lines might be very distant from RG2, and we have not yet found a restriction enzyme which detects an appropriate RFLP sufficiently far from RG2 (equivalent to not having discovered enzyme *Eco*RV in Fig. 2).

A fundamental question concerns the length of the heterozygous region near locus RG2. We have approached the region from both sides of this locus. We know that the nearest RFLP marker on one side of the region (RG16) is basically homozygous, but this marker is known to be about 6 cM away from RG2 (McCouch et al. 1988). However, RG16 also detects more heterozygosity (21.2%) than expected. Other neighbouring loci (RG118 and RG167) and locus RG103, which is on the other side of RG2, exhibit heterozygosity within the expected values (6%) for F_5–F_6 RI lines. Thus, it is evident that the heterozygous region is somewhere between loci RG2 and RG16 and can be detected by these probes using specific restriction enzymes. If the heterozygous locus corresponds to a single gene, then it might be expected to be a few kilobases long plus an additional 0–24 kb at each end to allow for variation in the site of recombination. Thus, the region might be 5–55 kb in length. If two

genes were involved, it seems unlikely that they could be far apart because the drive to seperate the genes would lead to recombination over the same 0–24 kb distance. Thus, we estimate that at most the heterozygous region is about 80 kb in length. The whole region might be accommodated on a single clone of a yeast artificial chromosome library. Identification of such a clone would permit detailed genetic analysis and provide probes for characterizing the tissues and developmental stages when the genes are expressed.

If each of the 34 heterozygous lines is assumed to represent a distinct lineage, then the 21-kb region between RG2 and the *Eco*RV restriction site has accumulated 34 recombination events during inbreeding. Since the recombination frequency of exhaustively inbred lines is about twice the gametic frequency, the gametic recombination frequency can be estimated to be approximately 9 per 21 kb, or approximately 430 per megabase. This is in comparison with the overall recombination frequency in rice of 1% per megabase.

From a practical point of view these results have two important consequences. First, they establish that the heterozygosity of a locus is not necessarily a fixed quantity and its detection by RFLP varies markedly with the restriction enzyme used to digest the DNA. Second, the results throw light on the significance of maintenance of heterozygous loci in RI lines and their probable role in heterosis.

Acknowledgements We thank Prof. K.K. Tewari, Director, ICGEB, for his help and constant encouragement; Dr. E. Siddiq, Deputy Director General, Indian Council of Agricultural Research, New Delhi-110001, for providing rice varieties and RI lines. We are grateful to Dr. B. Burr, Brookhaven National Laboratory, for critical reading of the manuscript and valuable suggestions. Thanks are also due to Mr. K. Singh and Mr. C. Singh for assistance in the laboratory. This work was supported in part by a grant from the Rockefeller Foundation.

References

Bonierbale MW, Plaisted RL, Tanksley SD (1993) A test of the maximum heterozygosity hypothesis using molecular markers in tetraploid potatoes. Theor Appl Genet 86:481–491

Burr B, Burr FA (1991) Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. Trends Genet 7:55–60

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

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

Charcosset A, Lefort-Buson M, Gallais A (1991) Relationship between heterosis and heterozygosity at marker loci: a theoretical computation. Theor Appl Genet 81:571–575

Claes B, Dekeyser R, Villarroel R, Van den Bulcke M, Bauw G, Van Mantagu M, Caplan A (1990) Characterization of a rice gene showing organ-specific expression in response to salt stress and drought. Plant Cell 2:19–27

Dudley JW, Saghai Maroof MA, Rufener GK (1991) Molecular markers and grouping of parents in maize breeding programs. Crop Sci 31:718–723

Edwards MD, Stuber CW, Wendel JF (1987) Molecular-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution, and types of gene action. Genetics 116:113–125

Ellis THN, Turner L, Hellens RP, Lee D, Harker CL, Enard C, Domoney C, Davies DR (1992) Linkage maps in pea. Genetics 130:649–663

Haldane JBS, Waddington CH (1931) Inbreeding and linkage. Genetics 16:357–374

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4: 745–750

Louw JH (1994) The effects of low frequency random crossing on heterozygosity and frequency of recombinant homozygotes in lines developed by single seed descent. Euphytica 78: 91–96

McCouch SR, Tanksley SD (1991) Development and use of restriction fragment length polymorphism in rice breeding and genetics. In: Khush GS, Toenniessen GH (eds) Rice biotechnology. CAB Int, Wallingford, UK, pp 109–133

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

Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990) Genetic diversity for restriction fragment length polymorphisms: relations to estimated genetic effects in maize inbreds. Crop Sci 30:1033–1040

Mohan M, Nair S, Bentur JS, Prasada Rao U, Bennett J (1994) RFLP and RAPD mapping of the rice *Gm2* gene that confers resistance to biotype 1 of gall midge (*Orseolia oryzae*). Theor Appl Genet 87:782–788

Moser H, Lee M (1994) RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats. Theor Appl Genet 87:947–956

Parker J, Szabo V, Staskawicz B, Lister C, Dean C, Daniels M, Jones J (1993) Phenotypic characterization and molecular mapping of the *Arabidopsis thaliana* locus RPP5 determining disease resistance to *Peronospora parasitica*. Plant J 4:821–831

Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci USA 89:1477–1481

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Smith OS, Smith JSC, Bowan SL, Tenborg RA, Wall SJ (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, grain yield heterosis and RFLPs. Theor Appl Genet 80:833–840

Snape JW, Flavell RB, O'Dell M, Hughes WG, Payne PP (1985) Intrachromosomal mapping of the nucleolar organiser region relative to three marker loci on chromosome 1B of wheat (*Triticum aestivum*). Theor Appl Genet 69:263–270

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from elite maize inbred lines using molecular markers. Genetics 132:823–839

Walbot V (1988) Preparation of DNA from single rice seedlings. Rice Genet Newsl 5:149–151

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Williams MNV, Pande N, Nair S, Mohan M, Bennett J (1991) Restriction fragment length polymorphism analysis of poly-

merase chain reaction products amplified from mapped loci of rice (*Oryza sativa* L.) genomic DNA. Theor Appl Genet 82: 489–498

Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. In: Plant Genome III. San Diego, p 89

Yuan LP, Vermani SS (1988) Status of hybrid rice research and development. In: Hybrid rice. Int Rice Res Inst, Manila, The Philippines, pp 7–24

Zhang Q, Gao YJ, Yang SH, Ragab RA, Saghai Maroof MA, Li ZB (1994) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor Appl Genet 89:185–192