Theor Appl Genet (1992) 83:495–499

Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis

Qifa Zhang [1,*], M. A. Saghai Maroof [2], T. Y. Lu [2], and B. Z. Shen [1]

[1] Biotechnology Center, Huazhong Agricultural University, Wuhan 430070, China
[2] Department of Crop and Soil Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Received March 25, 1991; Accepted May 29, 1991
Communicated by Hu Han

Summary. Genetic diversity and differentiation in *indica* and *japonica* groups of the cultivated rice (*Oryza sativa* L.) were studied by assaying DNA restriction fragment length polymorphisms of 12 *indica* and 14 *japonica* rice lines digested with three restriction endonucleases. A total of 49 probes were selected to represent the entire RFLP map at intervals of 20–30 cM. It was shown that 95 of the 145 possible probe/enzyme combinations, involving 43 probes and all three enzymes, detected restriction fragment length variation, and the degree of polymorphism varied greatly from one probe/enzyme combination to another. These results demonstrate that *indica* rice is genetically more diverse than *japonica* type. Significant differentiation between the two rice groups was detected by 33 probes representing 11 of the 12 rice chromosomes. It was deduced that the processes leading to differentiation involved a combination of molecular events that include base substitutions and insertion/deletions.

Key words: *Oryza sativa* – Phenotypic diversity – Differentiation – Randomization test – RFLP

Introduction

The genetic basis of diversity and differentiation in the *indica* and *japonica* groups of the cultivated rice (*Oryza sativa* L.) is pertinent to a wide range of research interests among scientists in a number of disciplines. Geneticists are interested both in the genes that are involved in such differentiation and in the genes that are capable of overcoming the partial reproductive isolation mechanisms between these two groups (Ikehashi and Araki 1985). Rice breeders are concerned with utilizing the very high degree of heterosis that has been observed in many intersubspecific crosses (e.g., Zeng et al. 1980), whereas evolutionary biologists are interested in the dynamics that have operated in the processes of *indica-japonica* differentiation (Oka 1964; Morishima and Oka 1982).

Experimental approaches have been undertaken to demonstrate the processes of differentiation in an attempt to understand the underlying genetic basis. Oka (1964) followed the associations of a number of morphological characters through the F_8 generation in progeny of several *indica-japonica* crosses propagated in bulk or in pedigree without conscious selection. He found that parental character combinations (*indica* or *japonica*) occurred much more frequently than would be expected on the basis of random association. These results indicate that the *indica-japonica* differentiation can perpetuate, at least in part, through the processes of hybridization and subsequent recombination and segregation, presumably as a result of joint effects of linkage and strong selection. This suggests that the parental *indica* and *japonica* gene arrays behave as cohesive units and that members within each array interact favorably to give rise to higher fitness than recombinants composed of portions from the two parental genomes. This in turn suggests that genetic differentiation between these two rice groups is extensive.

Attempts have also been made to estimate the extent of the genetic diversity and differentiation of the two rice groups. Large differences in allele frequency have been observed between *indica* and *japonica* types at a number of isozyme loci (e.g., Second 1982). The analysis of Morishima and Oka (1981), based on a number of morphological characters and isozyme loci, showed that about 50% of the total variation can be ascribed to genetic differentiation between these two groups. In addition, the results of these studies demonstrated substantial

* To whom correspondence should be addressed

variation within each of *indica* and *japonica* rice groups. However, a possible problem associated with the estimates in the above-mentioned studies is that only a minor portion of the genome was represented by those markers, and the markers themselves did not comprise a random sample of the genome. In fact, some of the morphological characters used were actually criteria for classification between *indica-japonica* types (Oka 1964), and hence of necessity, would give maximum observable morphological differentiation. Thus, estimates from these studies may be biased in one way or another.

In this paper, we report our study on genetic diversity and differentiation in *indica* and *japonica* rice groups in which we used a sample of 26 lines assayed with a total of 49 probes. These probes were selected to represent the rice RFLP map (McCouch et al. 1988) at intervals of 20–30 cM, and hence can be regarded as a random sample of the genome. Results of this study suggest that there is extensive genetic variation within the cultivated rice species that is readily detectable with RFLP analysis. Our analyses also clearly demonstrate that *indica* rice is genetically more diverse than *japonica* type and that a major portion of the genome, as represented by the RFLP markers, has been involved in genetic differentiation between these two rice groups.

Materials and methods

The genetic materials used in this study were 26 lines of cultivated rice (*Oryza sativa* L.) including 12 of *indica* type and 14 of *japonica* type. Most of these materials were obtained from the stocks maintained at the Hybrid Rice Research Unit of Huazhong Agricultural University, Wuhan, China.

Leaves were harvested from a single plant of each line grown in the field under normal agricultural conditions in Wuhan, China, and ground to a fine powder under liquid nitrogen. Total cellular DNA was extracted following, with minor modification, the CTAB method (Saghai Maroof et al. 1984).

These DNA samples were assayed for restriction fragment length polymorphisms with 49 cloned fragments. One of these probes was a clone that contains the entire wheat ribosomal RNA gene (rDNA) repeat unit (Gerlach and Bedbrook 1979); the other 48 probes were from the rice RFLP map (McCouch et al. 1988; T. Fulton personal communication) and were kindly provided by Dr. S.D. Tanksley. These probes were selected to represent the entire map at intervals of 20–30 cM. The majority of these clones represent single copy sequences.

For assaying restriction fragment variation with the 48 clones from the RFLP map, about 5 μg total cellular DNA was digested individually with each of three restriction enzymes EcoRI, HindIII, and XbaI, and resolved in 0.8% agarose gel. The DNA was then blotted onto nylon membranes following the manufacturer's instructions. Probes were labelled with α-^{32}P-dCTP to high specific activity using the random hexamer priming method (Feinberg and Vogelstein 1983). Prehybridization and hybridization were done following closely the procedures described by Saghai Maroof et al. (1984). Restriction fragment length variation of ribosomal DNA was assayed using about 1 μg DNA, which was digested with BamHI and resolved in an 1.0% agarose gel. Blotting and hybridization were done in the same manner as for the other probes with the exception that the probe was prepared by nick translation.

Results

Polymorphisms detected by various probe/enzyme combinations

Six probes, RG28, RG29, RG329, RG450, RG457, and RG463, did not detect any variation among the 26 rice lines assayed. On the contrary, rDNA appeared to be highly polymorphic, and the polymorphism has been demonstrated to be a result of length variation in the intergenic spacer region (Cordesse et al. 1990). Among the remaining 42 probes that revealed some degree of variation, 10 probes appeared to be polymorphic with only one restriction enzyme, 12 detected variation with two restriction enzymes, and the remaining 20 probes showed polymorphisms with all three enzymes (Table 1). Thus, in all, 95 probe/enzyme combinations detected polymorphisms among these 26 rice lines.

The banding patterns in the autoradiograph resolved by the majority of these 95 probe/enzyme combinations demonstrated typical single locus variation: almost all the lines in each case displayed one fragment length variant or another, and double-banded types appeared only in rare cases. Since the RFLPs of all probes have been subjected to Mendelian analysis (McCouch et al. 1988), we deduce that the single banded types represent homozygotes of one allele or another marked by the restriction fragment length variants, whereas the rare double-banded types represent heterozygotes of two different alleles that are marked by the two restriction fragment variants.

Banding patterns of a few probes cannot be explained by single locus variation; e.g., individuals are simultaneously variable for two or more bands, suggesting that more than one molecular event has been involved in such variation. However, these probes do not necessarily correspond to the clones that McCouch et al. (1988) determined as multiple copy or repetitive sequences in their analysis.

Some probes demonstrated obvious concurrence of RFLP patterns among different enzymes. To examine the relatedness of restriction fragment variation resolved by different enzymes for given probes, correlation analyses were performed between restriction patterns of paired enzymes within probes. There is one correlation within each of 12 probes that detect polymorphisms with two restriction enzymes, and three correlations within each of 20 probes that were variable with all three enzymes. Thus, in all, there are 72 possible correlations. The analyses (data not shown) demonstrated that the correlations were perfect for all three enzyme pairs in 2 probes (RG358 and RG528), and nearly perfect in 5 probes

Table 1. The level of genetic diversity and differentiation of *indica* and *japonica* rice averaged over three restriction enzymes for 43 polymorphic probes

Probe	Number of polymorphic enzymes	Diversity			Differentiation as % of total
		indica	*japonica*	Total	
RG2	2	0.362	0.137	0.271	10.95
RG13**	3	0.864	0.257	0.879	38.85[b]
RG25	1	0.150	0.137	0.143	0.12
RG64**	3	0.703	0.000	0.573	43.37[b]
RG69**	1	0.000	0.000	0.230	100.00[b]
RG101*	2	0.592	0.000	0.391	30.06[b]
RG109*	1	0.096	0.230	0.198	15.01[a]
RG122**	1	0.275	0.086	0.264	34.43[b]
RG134**	2	0.391	0.304	0.573	39.89[b]
RG144	3	0.434	0.176	0.365	19.03
RG152**	2	0.809	0.086	0.703	42.27[b]
RG167**	3	0.700	0.000	0.494	34.56[b]
RG171**	3	1.251	0.410	0.965	17.28[a]
RG182*	3	0.705	0.228	0.534	16.03[b]
RG213**	3	0.454	0.425	0.748	41.35[b]
RG236	1	0.102	0.000	0.056	16.27
RG252	2	0.300	0.000	0.181	23.32
RG256**	3	0.847	0.000	0.548	28.67[b]
RG257**	2	0.370	0.000	0.419	59.17[b]
RG303**	1	0.306	0.000	0.231	38.83[b]
RG324**	1	0.306	0.000	0.231	38.83[b]
RG341**	3	0.462	0.257	0.540	34.87[b]
RG346*	3	0.940	0.257	0.686	16.57[a]
RG348	2	0.246	0.173	0.262	21.20
RG351**	2	0.549	0.273	0.652	38.52[b]
RG358**	3	0.637	0.000	0.617	52.40[b]
RG365	1	0.096	0.000	0.054	18.79
RG396**	2	0.680	0.000	0.483	34.96[b]
RG400**	2	0.481	0.355	0.577	28.34[b]
RG403*	3	1.065	0.256	0.783	19.62[a]
RG424**	2	0.550	0.273	0.587	31.68[b]
RG470**	3	1.025	0.086	0.799	34.88[b]
RG528	3	0.451	0.000	0.271	23.32
RG553**	3	0.847	0.259	0.995	46.68[b]
RG570**	2	0.423	0.219	0.445	29.65[b]
RG598**	3	0.480	0.257	0.754	52.23[b]
RG620*	3	1.376	0.137	1.051	32.55[b]
RG650*	3	0.275	0.172	0.281	21.91[a]
RG778	1	0.000	0.086	0.054	15.00
RG811**	3	0.827	0.090	0.711	39.45[b]
RG869**	3	0.671	0.000	0.625	50.49[b]
RG885	1	0.289	0.173	0.248	8.52
rDNA**		1.907	1.195	2.128	28.37[b]
Average		0.476	0.128	0.438	34.06[b]

*,** Significant at probability level 0.05 or 0.01 as determined by the approximate t-test of the phenotypic frequencies resolved by the enzyme that detected largest amount of variation for each probe

[a], [b] Significant at probability level 0.05 or 0.01 as determined by the randomization test

(RG13, RG171, RG470, RG528 and RG598). In all, 56 of the 72 correlations were significant, 49 at the 0.01 probability level and 7 at probability level 0.05.

Diversity of indica and japonica rice lines

As mentioned in the previous section, a few probes did not show the typical single locus variation pattern, hence allelism could not be deduced on the basis of banding patterns in the autoradiographs. We henceforth refer to each distinct banding pattern as a phenotype. To evaluate the amounts of variation for *indica* and *japonica* rice lines, we adopted Shannon's information statistic (Bowman et al. 1971) to quantify the level of phenotypic diversity among these rice lines:

$$h_s = -\Sigma\, p_i \ln p_i$$

where p_i is the frequency of the i^{th} phenotype. Among the above-mentioned 95 polymorphic probe/enzyme combinations, 6 did not detect variation within the *indica* group, while 47 did not reveal polymorphism within the *japonica* group. A comparison of diversity values between the two rice groups showed that 86 of the 95 polymorphic probe/enzyme combinations revealed a higher level of diversity in the *indica* group than in the *japonica* group, 1 demonstrated equal amounts of diversity between the two groups, and 8 detected more variation in *japonica* rice than in *indica* rice. When the diversity statistic for each probe was averaged over the three enzymes (Table 1), 2 probes appeared to be monomorphic within the *indica* group and 15 within the *japonica* group. *Japonica* rice lines were more variable for 2 probes, but the reverse was the case for all the remaining 41 probes.

On average, the diversity value of *indica* rice lines was 3 to 4 times higher than that of *japonica* lines (Table 1).

Differences in phenotypic frequency between indica and japonica groups

The frequencies of phenotypic classes resolved by the enzyme that detected the maximum amount of diversity for each probe were compared between *indica* and *japonica* rice groups using an approximate t-test. The results of these tests showed that there were significant differences in phenotype frequencies for 33 of the 43 polymorphic probes. The differences were significant at the 0.01 probability level for 26 probes, and 0.05 probability level for 7 probes (Table 1).

It is interesting to note that the banding patterns resolved by 2 probe/enzyme combinations, RG69/XbaI and RG257/HindIII, appeared to be completely different in these two rice groups. In the first case (RG69/XbaI), all the *indica* lines were monotypic for the class 2 (arbitrarily numbered) phenotype, but all *japonica* lines were monotypic for the class 1 phenotype. In the second case

(RG257/HindIII), however, all the *japonica* lines were of class 1 phenotype but all the *indica* lines were of non-1 type.

It is also interesting to note that one probe (RG358), which exhibited a presence versus absence type of variation, also showed substantial differentiation between *indica* and *japonica* rice lines. Most *indica* lines hybridized to this probe, but none of the *japonica* lines contained sequences homologous to this probe.

Partitioning phenotypic diversity

The phenotypic diversity statistic for all 26 lines can be partitioned into two components: one corresponding to differentiation between the *indica* and *japonica* groups and the other, average diversity within the two groups. The results of such partitioning are given in Table 1, from which it can be seen that there is a wide distribution range in the relative magnitude of between-group differentiation among the 43 probes. They ranged from a low of 0.12% (RG25) to a high of 100% (RG69), with more than 90% of the probes falling between 10% and 60%.

It is difficult to characterize the statistical properties of the two components resulting from such partitioning. Therefore, a randomization test (Kempthorne 1955) was performed to evaluate the statistical significance of the differentiation component. In conducting this test, 12 of the 26 lines were selected at random and placed in one group, while the remaining 14 lines comprised the second group. The diversity index was calculated from this randomized data set and partitioned between the randomly permutated groups in the same manner as was done to the observed data. This procedure was repeated 150 times. Then the observed differentiation component was compared to those calculated from random permutations. If fewer than 5% of the random values were larger than the observed differentiation component, the differentiation was considered to be statistically significant at the 0.05 probability level. This procedure is believed to provide a nearly exact test for the null hypothesis. The results indicated that the differentiation components for 28 of the 43 polymorphic probes were significant at the 0.01 probability level (Table 1). It is also clear from Table 1 that the randomization test and t-test were very consistent in determining the statistical significance of the *indica-japonica* differentiation.

Discussion

We have analyzed genetic polymorphism resolved by 49 probes that cover the entire rice RFLP map at intervals of 20–30 cM among 26 rice lines encompassing a wide range of genetic variation in the species *O. sativa*. Among all the possible 145 probe/enzyme combinations, 95, involving 43 probes and all three enzymes, detected polymorphism. The analysis of phenotypic diversity showed that the average diversity indices ranged between 0 and 0.3 for 15 probes, between 0.35 and 0.80 for 23 probes, and above 0.85 for 5 probes. Thus, it is clear that there is extensive genetic variation in the cultivated rice species that is readily detectable with RFLP techniques.

Indica and *japonica*, the two major types of cultivated rice, appear to have distinct ranges of eco-geographical distribution. *Indica* rice is predominant in tropical and subtropical regions that comprise the major portion of rice-growing areas of the world, while *japonica* rice is more adapted to temperate regions. It is thus expected that *indica* rice would be genetically more diverse than *japonica* rice. Although the inference latitude in the present study is limited because of the small sample sizes, it is nevertheless clear that the amount of genetic variation in the *indica* group is much larger than that in the *japonica* group.

The *indica-japonica* differentiation has been recognized ever since ancient times in China, when it was referred to as *hsien* (*indica*) and *keng* (*japonica*). However, only recently have vigorous scientific studies been undertaken to characterize genetic differentiation between these two rice groups. Cheng (1985) considered six morphological characters to be adequate for classifying *indica* and *japonica* rices. Glaszmann (1987) found that 15 polymorphic loci coding for eight enzymes were sufficient for resolving most rice lines into respective groups. The results of Morishima and Oka (1981) indicated that the component representing *indica-japonica* differentiation accounted for about 50% of the total genetic diversity in their data set. In the present study, 33 out of 49 probes demonstrated significant differentiation, and, in all, about 34% of the total variation could be explained by this component. Of these 33 probes, 1 is located on chromosome 4, 2 are located on each of chromosomes 10, 11 and 12, 3 on each of chromosomes 1, 2, 6 and 7, 4 on chromosomes 3 and 9, respectively, and 5 on chromosome 5. rDNA has been mapped to two different chromosomes (9 and 10). Thus, the processes of *indica-japonica* differentiation have proceeded throughout the range of the entire genome and involved DNA segments in at least 11 of the 12 rice chromosomes.

The molecular nature of the *indica-japonica* differentiation may also be inferred by examining the concurrence of the restriction fragment patterns between enzymes within the 32 probes (rDNA was digested with only one enzyme) that demonstrated significant differentiation. Five probes detected polymorphism with one enzyme. Among the remaining 27 probes, 9 revealed polymorphism with two enzymes and 18 demonstrated variation with all three enzymes. There is only one possible correlation within each of the probes that were polymorphic with two enzymes: the banding patterns resolved by the two enzymes were significantly correlated

for seven of the nine probes that detected significant differentiation, and the correlation coefficients were very high (≥ 0.80) in five of the seven cases. Among the remaining 18 probes that detected variation with all three enzymes, all three correlations were very high (≥ 0.80) for 5 probes, only one correlation coefficient was ≥ 0.80 for 10 probes, and none of the correlation coefficients was ≥ 0.80 for the remaining 3 probes.

These results suggest that approximately half of the *indica-japonica* differentiation resolved in this RFLP analysis is attributable to insertion/deletion type of molecular events that occurred in one rice group but not in the other. In addition, about an equal amount of differentiation can be ascribed to base pair substitutions.

Acknowledgement. This project was supported in part by a grant from the Rockefeller Foundation and a grant from the Agricultural Ministry of China. MASM ackowledges contributions from the College of Agricultural and Life Sciences, VPI and SU. The authors thank Professor Wan Jinmeng and Mr. Jin Demin for providing the rice materials.

References

Bowman KD, Hutcheson K, Odum EP, Shenton LR (1971) Comments on the distribution of indices of diversity. Stat Ecol 3:315–359

Cheng KS (1985) A statistical evaluation of the classification of rice cultivars into *hsien* and *keng* subspecies. Rice Genet Newsl 2:46–48

Cordesse F, Second G, Delseny M (1990) Ribosomal gene spacer length variability in cultivated and wild rice species. Theor Appl Genet 79:81–88

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction fragment length polymorphisms to high specific activity. Anal Biochem 132:6–13

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885

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

Ikehashi H, Araki H (1985) Genetics of F_1 sterility in remote crosses of rice. In: IRRI (ed) Rice genetics. International Rice Genetics Symposium I Manila, Philippines, pp 119–130

Kempthorne O (1955) The randomization theory of experimental inference. J Am Stat Assoc 50:946–967

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 78:815–829

Morishima H, Oka HI (1981) Phylogenetic differentiation of cultivated rice. 21 Numerical evaluations of the *indica-japonica* differentiation. Jpn J Breed 31:402–413

Oka HI (1964) Considerations on the genetic basis of intervarietal sterility in *Oryza sativa*. In: IRRI (ed) Rice genetics and cytogenetics. Symposium on Rice Genetics and Cytogenetics Manila, Elsevier, Amsterdam, pp 159–174

Saghai Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza spp*): study of the polymorphism scored at 40 isozyme loci. Jpn J Genet 57:25–57

Zeng S, Yang X, Lu Z (1980) The heterosis of F_1 hybrids between *Oryza sativa* subsp *hsien* and *O. sativa* subsp *keng*. Acta Agron Sin 6:193–202