Theor Appl Genet (1999) 98: 1320–1328

R. K. Aggarwal · D. S. Brar · S. Nandi
N. Huang · G. S. Khush

Phylogenetic relationships among *Oryza* species revealed by AFLP markers

Received: 1 July 1998 / Accepted: 2 November 1998

Abstract The genus *Oryza* to which cultivated rice belongs has 22 wild species. Seventy-seven accessions of 23 *Oryza* species, five related genera, and three outgroup taxa were fingerprinted using amplified fragment length polymorphism (AFLP). A total of 1191 polymorphic markers were obtained using five AFLP primer combinations. AFLP data were analyzed to study species relationships using different clustering algorithms, and the resulting phenograms were tested for stability and robustness. The findings suggest a common ancestry to the genus *Oryza*. Moreover, the results demonstrate that: (1) evolution in *Oryza* has followed a polyphyletic path wherein multiple lineages underwent independent divergence after separation early in the evolution from a common ancestor/pool of related taxa; (2) newly assigned genomes, GG for *O. meyeriana* and HHJJ for *O. ridleyi* complexes, are among the most diverged in the genus; (3) CCDD tetraploids have a relatively ancient origin among the Officinalis complex; (4) *O. malampuzhaensis*, *O. indandamanica*, *O. alta*, and *O. grandiglumis* are diverged enough to deserve species status; (5) *O. officinalis* and *O. eichingeri* (CC) are putative progenitors of *O. minuta*/*O. malampuzhaensis* and tetraploid *O. punctata*, respectively, (6) *O. brachyantha* is most diverged species in the genus. AFLP is reliable molecular technique and provides one of the most informative approaches to ascertain genetic relationships in *Oryza*, which may also be true for other related species/organisms.

Key words *Oryza* species · AFLP · Polymorphism · DNA fingerprinting · Phenogram · Phylogenetic relationship

Communicated by B. S. Gill

R. K. Aggarwal[1] · D. S. Brar · S. Nandi
N. Huang · G. S. Khush (✉)
Plant Breeding, Genetics and Biochemistry Division,
International Rice Research Institute, P.O. Box 3127,
1271 Makati city, Philippines
Fax: + 63 2 891 1292, 816 8470
E-mail: gkhush@irri.cgnet.com

Present address:
[1] Centre for Cellular & Molecular Biology,
Hyderabad 500 007, India

Introduction

Variation in morphological traits, geographical distribution, cytogenetic relationships, breeding systems, cross-compatibility, and biochemical markers have been used extensively to elucidate the relationships among species. Although these classical methods are central to present-day taxonomy, they are restricted in their resolving power mainly because of the small number of variables available. In contrast, molecular approaches such as restriction fragment length polymorphism (RFLPs) (Wang et al. 1992) and DNA fingerprinting using hypervariable minisatellite probes (Aggarwal et al. 1994) provide genetically interpretable variability with extensive genomic coverage and have thus become immensely important in studies on population biology and systematics. While RFLPs have proven to be the most informative and reliable compared to all other approaches, they suffer serious limitations on account of being labor-, cost-, and time-intensive and because of their need for a relatively large sample size and, most importantly, for a library of cloned and mapped DNA probes. Amplified fragment length polymorphism (AFLP), a new approach for plant DNA fingerprinting (Vos et al. 1995), overcomes most of the limitations of RFLPs. It is emerging as an important technique for genome mapping (Becker et al. 1995; Maheswaran et al. 1997), gene tagging (Maksem et al. 1995) and also for diversity (Paul et al. 1997) and phylogenetic analysis of closely related plant species (Sharma et al. 1996; Hill et al. 1996). We demonstrate here the potential of AFLP for inferring phylogenetic

relationships among highly diverged species (comprising nine genomes) of the genus *Oryza*.

The genus *Oryza* (tribe Oryzeae; family Poaceae) comprises 22 wild species and two rice cultigens (*O. sativa* and *O. glaberrima*) and provides food for more than one-third of the world's population. The wild relatives of cultivated rice are useful sources of genetic variability for many agronomic traits and thus have great potential for its improvement (Jena and Khush 1990). However, to make effective use of this desirable pool of genetic diversity, we need to understand their genomic relationships and to assess inter- and intraspecific genetic diversity. Based on classical, isozyme, and RFLP studies, all species in the genus *Oryza* except *O. brachyantha* (FF genome) have been grouped into four main species complexes: (1) Sativa, (2) Officinalis, (3) Ridleyi, and (4) Meyeriana (Tateoka 1962; Second 1991; Wang et al. 1992; Vaughan 1989, 1994). The former two complexes have 12 diploid and 5 tetraploid species representing AA, BB, CC, BBCC, CCDD, EE genomes, whereas the Meyeriana and Ridleyi complexes have only 5 species with the recently assigned new genomes, GG and HHJJ (Aggarwal et al. 1997). So far, phylogenetic relationships in *Oryza* have been based on morphology, hybridization, isozyme, and RFLP analyses. We used AFLP to analyze 77 accessions of 23 *Oryza* species, five related genera and three outgroup taxa in order to investigate the phylogenetic relationships among *Oryza* species.

Materials and methods

Plant material

The material comprised 77 accessions of 23 *Oryza* species, five related genera, and three outgroup taxa–maize, sugarcane, and soybean (for details see, Figs. 1 and 2). Leaf samples for DNA extraction were obtained from the Genetic Resources Center and the Plant Breeding, Genetics, and Biochemistry Division, IRRI.

DNA extraction and template preparation for AFLP

Total genomic DNA was isolated from green leaves following the procedures of Dellaporta et al. 1983 with minor modifications (Aggarwal et al. 1997). AFLP analysis was done as described earlier (Vos et al. 1995; Maheswaran et al. 1997) with minor modifications. About 0.5 μg DNA digested with restriction enzymes *Pst*I and *Mse*I was ligated with corresponding adapters (Zabeau and Vos 1993) in a total volume of 50 μl. A 2- to 5-μl aliquot of the adapter-ligated DNA fragments was then pre-amplified using the primers 5′-CTC GTA GAC TGC GTACAT GCA-3′ (*Pst*I) and 5′-GAC GAT GAG TCC TGAGTA A-3′ (*Mse*I) in a volume of 25 μl for 30 cycles of: 94°C/30 s; 60°C/30 s; 72°C/60 s; this was followed by a 72°C/5-min extension. The preamplified samples were diluted approximately tenfold with water for use as a template for selective amplification.

Selective amplification and PAGE analysis

Selective amplification was conducted using five primer pair combinations (P1/M1, P1/M2, P1/M3, P1/M4, and P2/M5). Each primer contained three selective nucleotides at the 3′ end (written in 5′ to 3′ direction): CCA and GTT for *Pst*I-specific primers P1 and P2; and CAC, CAA, CAT, CCT, and CAG for *Mse*I-specific primers M1, M2, M3, M4 and M5, respectively. The core sequences of selective primers were from Zabeau and Vos (1993). Selective amplification was carried out in 25-μl reaction volumes using 5 μl of template, 20 ng of γ-[^{33}P]-end-labeled *Pst*I primer, and 50 ng of unlabeled *Mse*I primer (Vos et al. 1995). The polymerase chain reaction (PCR) amplification was done for 35 cycles, starting with the profile of 94°C/30 s, 65°C/30 s, 72°C/60 s and followed by 5 cycles with a stepwise reduction of annealing temperature by 1°C to 60°C that was subsequently maintained for the next 30 cycles. All amplification reactions were carried out in micro-titer plates in a Techne thermocycler. PCR products were resolved on a 6% denaturing polyacrylamide gel. Gels were dried and exposed to X-ray film for about 14–28 h depending on signal intensity.

Scoring of AFLPs and cluster analysis

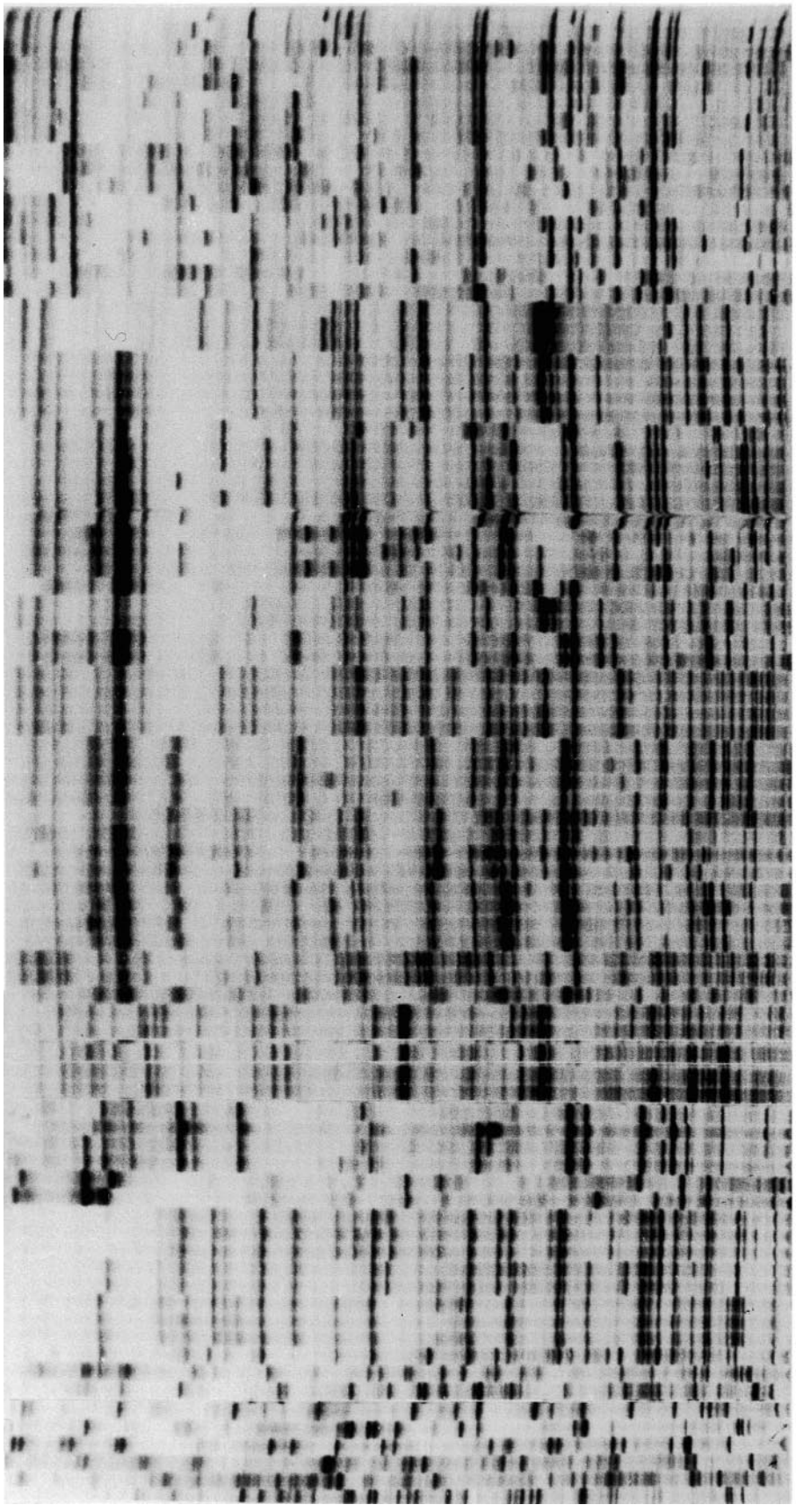
Each AFLP fragment/marker was treated as a unit character and scored as binary codes (1/0 = +/−). Only distinct, reproducible, well-resolved fragments were scored. The 1/0 matrix was used to calculate (dis)similarity coefficients following Nei and Li (1979) for each primer pair separately and also for the data pooled over all the primer combinations. The resulting distance matrices were used to construct an unweighted pair-group method with arithmetic means (UPGMA) (Sokal and Michener 1958) phenogram using software packages NTSYS-PC 1.8 (F.J. Rohlf, State University of New York, Stony Brook, USA) and 'PHYLIP version 3.57C' (J. Felsenstein, University of Washington, Seattle, USA) to infer phylogenetic relationships. The dendrograms were also constructed using other clustering methods based on distance matrices – single-link and complete-link available in NTSYS; Neighbor Joining [N-J method (Saitou and Nei 1987)], Kitsch, and Fitch (Fitch and Margoliash 1967) programs of the PHYLIP version 3.57C. The reliability, goodness of fit, and robustness of the phyletic trees were tested by comparing dendrograms from different methods, by deriving cophenetic correlations between cophenetic values for the tree obtained and the original data matrix (Sneath and Sokal 1973), by bootstrapping (Felsenstein 1985), and by using the Jackknife approach. Bootstrapping was done using the software package 'WINBOOT' developed at IRRI (Yap and Nelson 1996). Jackknife analysis was done by dropping half of the original data points selected at random.

Results

Polymorphism

AFLP analysis revealed a very large number of distinct, scorable fragments per primer pair (Fig. 1). In total, 1191 polymorphic markers were obtained. The level of polymorphism was much lower within species (approx. 2% in *O. minuta* to 21% in *O. officinalis*) and also between species carrying similar genome(s) (approx. 20% for the HHJJ genome to 35% for the BB/BBCC

Fig. 1 A portion of AFLP fingerprint patterns in the genus *Oryza*, related genera, and outgroup taxa using the P1/M3 primer combination. The species and accessions used are indicated at the *top* of the panel. (Entries marked with (*a*) are those whose inferred genome(s) differ from the expected; see Results)



	Species	Accession	Genome
1	*O. sativa*	IR-31917-45-3	*AA*
2	*O. sativa*	IR-56	*AA*
3	*O. sativa*	IR-64	*AA*
4	*O. rufipogon*	105908	*AA*
5	*O. rufipogon*	105909	*AA*
6	*O. rufipogon*	105910	*AA*
7	*O. rufipogon*	106412	*AA*
8	*O. rufipogon*	106423	*AA*
9	*O. longistaminata*	103886	*AA*
10	*O. longistaminata*	103890	*AA*
11	*O. longistaminata*	103902	*AA*
12	*O. barthii*	101937	*AA*
13	*O. nivara*	103407	*AA*
14	*O. nivara*	105721	*AA*
15	*O. nivara*	106185	*AA*
16	*O. glumaepatula*	100969	*AA*
17	*O. nivara*	104823	*AA*
18	*O punctata*	103896	*BB*
19	*O punctata*	104064	*BB*
20	*O punctata*	105690	*BB*
21	*O punctata*	105980	*BBCC*
22	*O punctata*	100884	*BBCC*
23	*O punctata*	101409	*BBCC*
24	*O punctata*	104975	*BBCC*
25	*O. officinalis*	100896	*CC*
26	*O. officinalis*	101116	*CC*
27	*O. officinalis*	101399	*CC*
28	*O. officinalis*	105100	*CC*
29	*O. officinalis*	105220	*CC*
30	*O. officinalis*	100176	*CC*
31	*O. rhizomatis*	103421	*CC*
32	*O. rhizomatis*	105448	*CC*
33	*O. rhizomatis*	105449	*CC*
34	*O. eichingeri*	101424	*CC*
35	*O. eichingeri* (a)	105181	*BBCC*
36	*O. eichingeri* (a)	105182	*BBCC*
37	*O. eichingeri*	105408	*CC*
38	*O. eichingeri*	105413	*CC*
39	*O. minuta*	101089	*BBCC*
40	*O. minuta*	101141	*BBCC*
41	*O. minuta*	103876	*BBCC*
42	*O. minuta*	105253	*BBCC*
43	*O. alta*	100888	*CCDD*
44	*O. alta*	100952	*CCDD*
45	*O. alta*	100967	*CCDD*
46	*O. alta*	105143	*CCDD*
47	*O. latifolia*	100168	*CCDD*
48	*O. latifolia*	100914	*CCDD*
49	*O. latifolia*	100955	*CCDD*
50	*O. latifolia*	103787	*CCDD*
51	*O. grandiglumis*	105155	*CCDD*
52	*O. grandiglumis*	105157	*CCDD*
53	*O. grandiglumis*	105560	*CCDD*
54	*O. grandiglumis*	105669	*CCDD*
55	*O. malampuzhaensis*	105223	*BBCC*
56	*O. malampuzhaensis*	105328	*BBCC*
57	*O. malampuzhaensis* (a)	105329	*CCDD*
58	*O. ridleyi*	100820	*HHJJ*
59	*O. ridleyi*	100821	*HHJJ*
60	*O. ridleyi*	105973	*HHJJ*
61	*O. longiglumis*	105147	*HHJJ*
62	*O. longiglumis*	105148	*HHJJ*
63	*O. australiensis*	100882	*EE*
64	*O. australiensis*	103318	*EE*
65	*O. australiensis*	105269	*EE*
66	*O. australiensis*	105272	*EE*
67	*O. brachyantha*	101232	*FF*
68	*O. brachyantha*	94-10482	*FF*
69	*O. granulata*	100879	*GG*
70	*O. granulata*	102118	*GG*
71	*O. granulata*	104503	*GG*
72	*O. granulata*	106449	*GG*
73	*O. granulata*	104986	*GG*
74	*O. meyeriana*	wsp-90-5	*GG*
75	*O. meyeriana*	106473	*GG*
76	*O. meyeriana*	106474	*GG*
77	*O. indandamanica*	105694	*GG*
78	*Porteresia coarctata*	104502	...
79	*Leersia perrieri*	105461	...
80	*Rhynchoryza subulata*	100913	...
81	*Hygroryza aristata*	105457	...
82	*Chikusichloa aquatica*	106186	...
83	Maize	Local cultivar	...
84	Sugarcane	Local cultivar	...
85	Soybean	cv clark	...

Table 1 Distribution of AFLP markers detected for different primer pairs across genomes of *Oryza*, related genera, and outgroup taxa

Primer pair used	Total polymorphic markers scored	Average number of markers/individual over:				Average number of markers per individual/genome in *Oryza*								
		Diploid *Oryza* species	Tetraploid *Oryza* species	Related genera	Outgroup taxa	AA	BB	CC	EE	FF	GG	BBCC	CCDD	HHJJ
P1/M1	234	28	40	33	37	22	22	31	32	25	36	39	40	41
P1/M2	260	39	52	43	49	36	37	38	37	39	45	50	51	56
P1/M3	218	32	48	39	39	33	36	34	32	18	40	53	48	43
P1/M4	255	39	52	47	44	37	37	36	40	40	44	52	47	57
P2/M5	224	28	39	34	40	29	22	26	33	26	34	38	42	37
Average	238	33	46	39	42	31	31	33	35	30	40	46	46	47
Total	1191	166	231	196	209	157	154	165	174	148	199	232	228	234

genome species). Table 1 shows the distribution of AFLP markers for different genomes of *Oryza*, related genera, and outgroup taxa. In general, more markers were obtained for allotetraploid *Oryza* species (average 46) than for their diploid relatives (average 33 per individual, Table 1).

Genetic variation within and between species

Nei's genetic distances (D-values) showed a linear increase from within species to between species and were most striking between different genomes. Within species, distances ranged from 0.024 (*O. meyeriana*) to 0.213 (*O. officinalis*), with the exception of 2 accessions of *O. brachyantha* which showed a much higher value (0.301). The average D-values 'between species carrying similar genome(s)' was approximately 2.8 times (0.293) than 'within species' (0.104); while 'between genomes' of *Oryza*, it was 0.737 (Table 2). The average genetic distances 'between related genera' and 'between outgroup taxa' were more than the 'between-genome' distances observed in the genus *Oryza*, thus confirming their status as outgroup genera.

Genetic affinities among Sativa complex species

The Asian species (*O. sativa*, *O. nivara*, and *O. rufipogon*) with the AA genome were more closely related to each other (average D = 0.173 within themselves) than to their relatives from Africa and America and they clustered together (Fig. 2). Their lower D-values also suggested a relatively recent differentiation of these forms in the Sativa complex. In comparison, African species *O. barthii* and *O. longistaminata* with the AA genome represented the two most distant groups in Sativa complex, having average distances of 0.266 and 0.441, respectively, from their Asian counterparts. The only Sativa complex species from America, *O. glum aepatula*, considered to be a subtype of *O. rufipogon*, showed a closer affinity with *O. barthii* rather than to *O. sativa*, *O. nivara* and *O. rufipogon*, the 3 Asian forms (average distance values being 0.289, 0.329, 0.307 and 0.332 respectively).

Genetic affinities among Officinalis complex species

Genetic distances revealed 5 well-distinguished groups corresponding to BB, BBCC, CC, CCDD and EE genomes which had an average intergenomic distance of 0.615 between them. Of the 3 diploid species in the Officinalis complex, EE (*O. australiensis*) was the most distant genome having an average distance of 0.776 from the other two genomes, BB and CC. Within the CC-genome species, *O. officinalis* showed the highest within-genome divergence. Unexpectedly, 2 of the *O. eichingeri* accessions (Acc. 105181 and 105182) showed distances comparable with that of tetraploid *O. punctata* and were closely aligned with the latter in the phenogram (Fig. 2). Subsequent cytological analysis revealed these to be tetraploid. Therefore, these accessions seem to be misclassified.

The 3 BBCC tetraploids showed closest affinity with the diploid *O. punctata* (average D = 0.365, the least in the Officinalis complex) and were accordingly clustered along with the latter (Fig. 2). Among themselves, *O. minuta* and *O. malampuzhaensis* were more closely related to each other (D = 0.219) than either to the *O. punctatas* (average D = 0.304), which in turn showed the minimum distance from its diploid forms. Similarly, with respect to, CC-genome, the former two were nearest to *O. officinalis*, while for *O. punctata* it was nearest to *O. eichingeri* (Table 2). One of the *O. malampuzhaensis* accessions (Acc. 105329) was found closely grouped with the CCDD tetraploid *O. alta* rather than to other accessions of *O. malampuzhaensis*; this was probably a case of mis-classification.

The genetic distance data revealed that the 3 CCDD allotetraploid species were more closely related to each other than to any of the others in the Officinalis complex. *O. alta* and *O. grandiglumis* were found to be closer to each other (D = 0.221) than either to the

Table 2 Distance matrix showing within- and between-species genetic distances for some of the diploid and tetraploid species in genus *Oryza*

	Within species[a]	Between species[b]	BBCC			CCDD			HHJJ	
			O. punctata	*O. minuta*	*O. malampuz haensis*	*O. alta*	*O. latifolia*	*O. grandiglumis*	*O. ridleyi*	*O. longiglumis*
O. sativa	0.0816		0.7162	0.7355	0.7054	0.7918	0.7767	0.7977	0.8369	0.8148
O. rufipogon	0.1177		0.7339	0.7507	0.7213	0.7767	0.7568	0.7985	0.7716	0.7575
O. longistaminata	0.1771		0.7154	0.6992	0.6873	0.7349	0.7330	0.7629	0.7813	0.7640
O. barthii	–		0.7445	0.7665	0.7282	0.7982	0.7730	0.7911	0.8015	0.7943
O. nivara	0.1380		0.7218	0.7429	0.7080	0.7808	0.7722	0.7908	0.8111	0.7880
O. glumaepatula	–	0.3024	0.7255	0.7419	0.7161	0.7596	0.7745	0.7839	0.7930	0.7694
O. punctata-BB	0.1047		0.2915	0.4286	0.4390	0.6624	0.6821	0.6824	0.7963	0.7914
O. officinalis	0.2134		0.5012	0.4315	0.4616	0.5000	0.5288	0.5239	0.7601	0.7309
O. rhizomatis	0.1194		0.4760	0.4883	0.4950	0.5267	0.5552	0.5264	0.7410	0.7330
O. eichingeri	0.1963	0.3689	0.4273	0.5158	0.5392	0.5127	0.5298	0.4940	0.7732	0.7698
O. punctata-BBCC	0.0664		0.0664	0.2934	0.3075	0.5092	0.5225	0.5173	0.7364	0.7316
O. minuta	0.0376		0.2934	0.0376	0.2193	0.5443	0.5524	0.5455	0.7578	0.7195
O. malampuzhaensis	0.0375	0.2704	0.3075	0.2193	0.0375	0.5540	0.5622	0.5805	0.7407	0.7028
O. alta	0.0611		0.5092	0.5443	0.5540	0.0611	0.3027	0.2206	0.7254	0.6993
O. latifolia	0.1653		0.5225	0.5524	0.5622	0.3027	0.1653	0.3530	0.7214	0.7006
O. grandiglumis	0.1018	0.2921	0.5173	0.5455	0.5805	0.2206	0.3530	0.1018	0.7469	0.7296
O. ridleyi	0.1795		0.7364	0.7578	0.7407	0.7254	0.7214	0.7469	0.1795	0.2213
O. longiglumis	0.0251	0.2213	0.7316	0.7195	0.7028	0.6993	0.7006	0.7296	0.2213	0.0251
O. australiensis	0.1218		0.7424	0.7540	0.7373	0.5909	0.5955	0.6095	0.7390	0.7613
O. brachyantha	0.3012		0.8077	0.8082	0.7864	0.8018	0.7994	0.8081	0.7595	0.7640
O. granulata	0.1021		0.8130	0.7667	0.7602	0.7902	0.7686	0.7899	0.7255	0.7312
O. meyeriana	0.0243		0.8034	0.7837	0.7707	0.8388	0.8233	0.8398	0.7284	0.7078
O. indandamanica	–	0.3008	0.8061	0.7790	0.7594	0.8275	0.8098	0.8045	0.7146	0.6957

[a] Distances are given only for species for which more than 1 accession was analyzed
[b] Average genetic distance over all the species carrying similar genome(s)

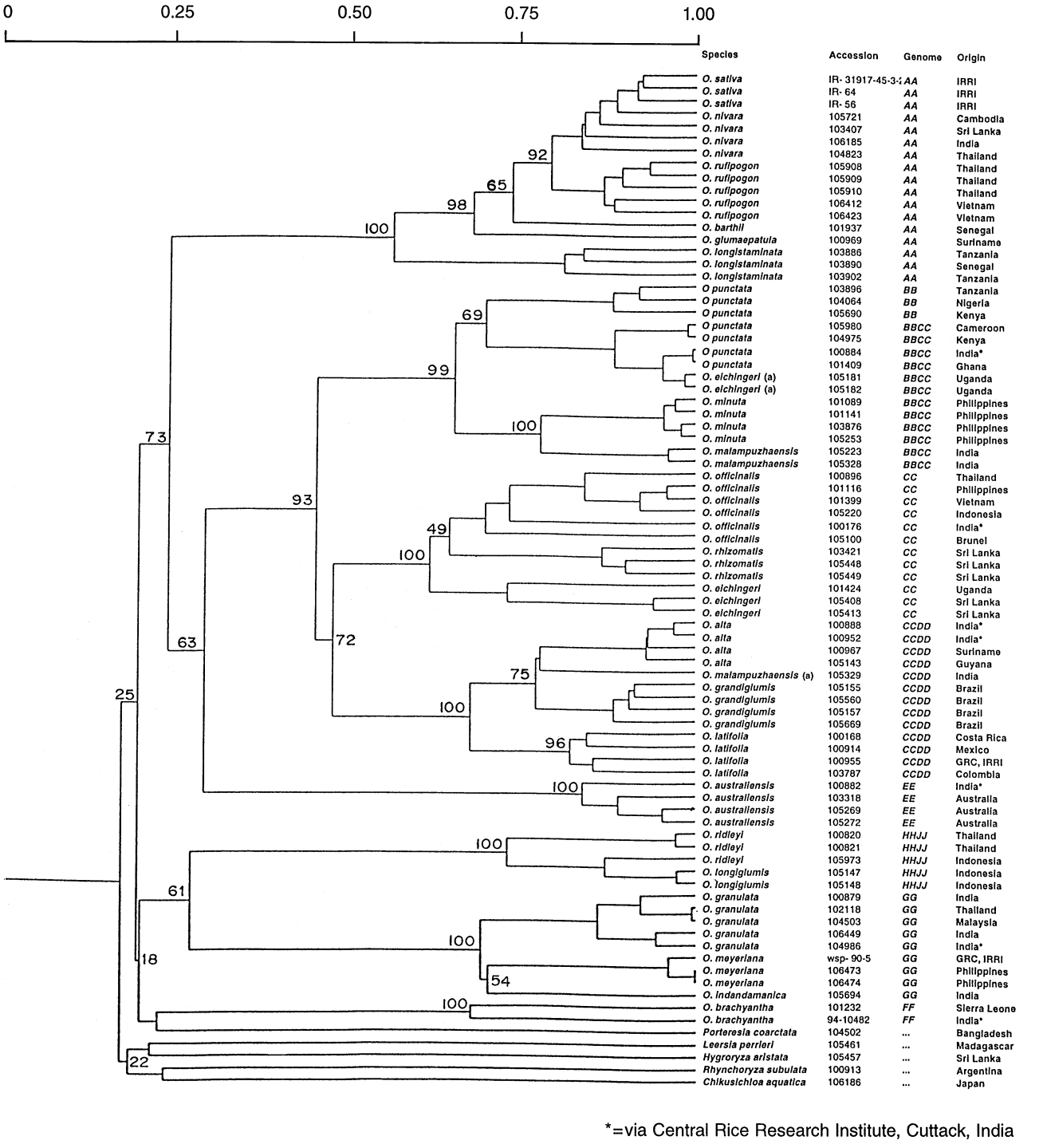


Fig. 2 An UPGMA phenogram based on AFLP markers obtained with five primer pairs. *Numbers* shown at different nodes represent percentage confidence limits obtained in the bootstrap analysis. The *scale* shown *above* is the measure of genetic similarity calculated according to Nei and Li (1979). The species/accessions used, their genome(s), and origin are indicated on the *right* side of the panel. (Entries marked with (*a*) are those whose inferred genome(s) differ from the expected; see Results)

O. latifolia accessions (average D = 0.328). In general, these species showed much larger genetic distances from other *Oryza* species than did their BBCC counterparts.

Genetic affinities among Meyeriana and Ridleyi complex species, and *O. brachyantha*

Species in the Meyeriana and Ridleyi complexes having the GG and HHJJ genomes, respectively, were found to

be closely related within themselves but showed very large distances (ranging from 0.708 to 0.851, and from 0.708 to 0.825, respectively) from all the other diploid or tetraploid species. On the whole, these two complexes were the nearest (D = 0.717) to each other than other complexes. The 3 diploid species – *O. granulata*, *O. meyeriana*, and *O. indandamanica* – of the Meyeriana complex appeared to be well-distinguished from each other with an approximately equal genetic distance of 0.301. Similarly, species of Ridleyi group, *O. ridleyi* and *O. longiglumis*, were well-differentiated from one another (average D = 0.221). Two accessions of *O. meyeriana* (Acc. 106473 and 106474) were found to be identical, and it is possible that these represent duplicates of the same entry collected at different times. *O. brachyantha* (FF genome) was found to be the most distant from other species in the genus. Interestingly, its 2 accessions (Acc. 101232 and 94-10482) exhibited the maximum divergence (0.301), which is comparable to the 'between-species' distance. It is probable that these accessions represent two different subspecies of the FF genome and need to be further investigated.

Phylogenetic analysis

Figure 2 shows an UPGMA clustering based on all the 1191 AFLP markers. This clustering revealed eight monophyletic groups of species in the genus *Oryza* (each corresponding to a different genomic constitution with confidence limits of 95–100%) at a 55% level of genomic similarity. At this level of similarity, BBCC tetraploids did not form a separate monophyletic group but were found to be closely aligned to their BB-genome relatives. The phenogram suggests that these might have evolved by relatively recent hybridization and polyploidization event(s) between the BB- and CC-genome species. The UPGMA tree further shows that six of the monophyletic groups were derived from only two lineages; one leading to the Officinalis complex comprising species with the BB, BBCC, CC, CCDD, EE genomes, and the other giving rise to two species complexes, Ridleyi (HHJJ genome) and Meyeriana (GG genome). Moreover, species of the Sativa and Officinalis complexes were closer to each other than to any other complex in the genus *Oryza*. Most importantly, the phylogenetic analysis suggested a common ancestry for all the species of genus *Oryza*, as all the lineages in this genus converge to a single point before separating from other related genera and species included as outgroup taxa. Although the common node for all the species under *Oryza* had a low confidence value, it was distinctly evident in all the clustering algorithms.

The various tests done to evaluate the 'goodness of fit' of the resulting phylogenetic trees revealed the reliability and stability of the inferred relationships vis-à-vis the input data. In general, very high cophenetic correlations values – 0.943, 0.902, and 0.915 – were obtained for UPGMA, complete-link, and single-link based clustering, respectively. Similarly, bootstrap analysis using 100, 500, and 2000 replications revealed high confidence values (>95–100%) for all the genomic groups of *Oryza*. Least square distance methods resulted in trees similar to those shown in Fig. 2, with relatively low values of percentage standard deviation, 7.65 (over 45 786 trees in Kitsch) and 6.32 (over 268 trees with global optimization option in Fitch). Also, the species relationships obtained from Kitsch, Fitch and N-J methods and from the Jackknife approach were essentially similar to those obtained using UPGMA (Fig. 2), except for minor differences in branch lengths between any two given taxa and some topographical rearrangements with respect to related genera and outgroup taxa.

Discussion

Molecular evaluation of genetic diversity is a means to study and partition the genetic variability of related species/genera, thus resolving their phylogenetic relationships. In recent years, this has been approached using isozyme markers, organelle and nuclear DNA typing, and RAPD and RFLP markers, with varying success. Most of such studies in *Oryza* have been restricted to the Sativa and Officinalis complexes (Tateoka 1962; Oka 1988; Dally and Second 1990; Wang et al. 1992). In the study presented here, we analyzed 23 species of genus *Oryza* in relation to other members of Oryzeae and outgroup taxa using the high-resolution approach of AFLP.

Technique

An advantage of AFLP-based DNA fingerprinting is its potential in exposing large genetic polymorphism giving near complete coverage of the whole genome. In the present study, 1191 polymorphic markers were generated using five primer pairs, a number which is approximately 1.5 times greater than that obtained with 46 RFLP probes for the same material (unpublished data). This demonstrates that the polymorphism revealed by AFLP is approximately 15-fold that of nuclear RFLPs. Considering that the recent AFLP map of rice has 208 markers spanning about 1500 cM (Maheswaran et al. 1997), the AFLP markers analyzed (Table 1) would be equal in diversity evaluation to more than 150 nuclear loci/diploid samples. In fact, the results of various tests done to check the robustness of the phenogram/estimates of phylogeny clearly establish that polymorphism revealed by AFLP is not only abundant but also stable and statistically reliable. Moreover, these results demonstrate that genetic resolution provided by AFLP is amenable to phylogenetic

analysis not only of closely related species as shown earlier (Sharma et al. 1996; Hill et al. 1996) but even of highly diverse species as represented by those in the genus *Oryza* comprising nine genomes.

Species relationships/taxonomic information

The genetic relationships revealed by AFLP are generally consistent with our understanding of *Oryza* taxonomy based on morphology, hybridization, isozyme, and RFLP analysis. Four species complexes were easily distinguished, and *O. brachyantha* does not belong to any of these complexes. As expected, the genetic variation (distances) increased from within species to between species and was most striking among different genomes. Variation was higher for *O. officinalis* (D = 0.213) and *O. latifolia* (D = 0.165), which have wider distributions and are probably of older origin, but low for BBCC tetraploids (average D = 0.047), suggesting these latter to be of relatively recent origin.

Eight monophyletic groups (each with 95–100% confidence intervals) were evident at a 55% level of similarity (measured in terms of Nei's distances) in the resulting phenogram (Fig. 2). Four groups representing five genomes – BB, BBCC, CC, CCDD, and EE – derived from a single lineage were evident in the Officinalis complex. This lineage was also nearest to the one leading to the AA genome species of the Sativa complex. This suggests that species of Officinalis and Sativa complexes have differentiated separately as distinct groups and have undergone parallel evolution. Despite their divergence as a distinct complex, AA-genome species showed relatively closer affinities to BBCC, BB, CC, and all CCDD species than to EE in terms of absolute genetic distances. A similar situation was also noted at the organelle DNA level (Dally and Second 1990).

Within the Sativa complex, the three geographic forms, African (*O. longistaminata* and *O. barthii*), American (*O. glumaepatula*), and Asian (*O. sativa*, *O. nivara*, *O. rufipogon*,) were distinct and evolved independently both in the phyletic representation and in terms of absolute genetic distances. This supports the conclusion of Morishima (1969) but is at variance with that of Second (1985) who, on the basis of isozyme data, observed that while African strains have diverged independently for millions of years, the American forms have evolved only in the last few centuries, probably following man-made introductions through introgressive hybridization between Asian rufipogons and African forms. The organelle polymorphism data observed by Dally and Second (1990), which were at variance with this conclusion, were thought to be due to large-scale nucleocytoplasmic substitutions (Dally and Second 1990; Second 1991). The present study clearly shows that the AA-genome American species is closer to the African species equally in terms of nuclear AFLPs as was seen earlier at the organelle level (Dally and Second 1990) and thus might have evolved either independently or might have descended from their African relatives over a long period of time.

The genetic distances observed between the diploid species with BB, CC, and EE genomes (average D = 0.752) were about twofold greater than those observed within the AA-genome species (average D = 0.302), suggesting that these species have evolved over a much longer period than did those of the Sativa complex. The 3 CCDD species were more closely related to each other than to any other species in the Officinalis complex. The closest relatives of these species were CC- and EE-genome species. Wang et al. (1992) also reported similar results on the relationships of CCDD species. It is highly improbable that the CCDD tetraploid species, which are the second most diverged in the Officinalis complex (after only *O. australiensis*), have evolved only in the last two to three centuries, as suggested by Second (1991). This is well-reflected in their genetic distances from their nearest diploid relatives having CC and EE genomes (D = 0.523 and 0.595, respectively). An ancient origin of CCDD species was also inferred using nuclear RFLPs (Wang et al. 1992).

O. officinalis was found to be the most diverged among the 3 CC-genome species. In contrast, *O. rhizomatis* showed the minimum between- and within-species variation and is thought to be of comparatively recent origin. A comparison of the genetic distances between putative parents of BBCC tetraploids suggested that these have evolved from at least two independent allotetraploidization events, as also suggested earlier (Dally and Second 1990; Wang et al. 1992). Our results show that the putative CC-genome donor of tetraploid *O. punctata* is *O. eichingeri* and of tetraploid *O. malamphuzaensis* and *O. minuta* is *O. officinalis.*

The species belonging to the *O. ridleyi* and *O. meyeriana* complexes and *O. brachyantha* accounted for three other distinct and highly diverged groups in the genus *Oryza.* The Ridleyi and Meyeriana complexes seemingly evolved from a single lineage (Fig. 2) and show a relatively closer affinity with each other. In comparison, *O. brachyantha* did not align with any other species complex of *Oryza* but rather overlapped with some of the related genera. These results support cytological findings on the FF genome of *O. brachyantha* (Li et al. 1961) but are in contrast to those of Wang et al. (1992) who, based on RFLP analysis, found some affinity of the *O. sativa* complex with *O. brachyantha.*

An evaluation of genetic distances within and between *Oryza* species supports a full species status for *O. malampuzhaensis*, *O. indandamanica*, *O. alta*, and *O. grandiglumis*, as each of these has genetic distances (from their closest relatives) comparable with those seen between related species but significantly more than intraspecific variation (data not shown). The species

status for CCDD species was also supported by total genomic DNA hybridization and RFLP analysis (Aggarwal et al. 1996).

Evolutionary significance

The present study provides evidence suggesting that the *Oryza* genus has originated from a common ancestral taxon or a pool of closely related taxa which separated early in evolution. This is apparent in the phyletic representation, wherein lineages leading to all the *Oryza* species originate from a single node (Fig. 2), separating them out from most other related genera. In addition, the results demonstrate that overall evolution in rice has followed a polyphyletic path, wherein multiple lineages (originating from common taxon or pool of taxa) underwent independent and parallel evolution leading to present-day distinct genomic groups. This supports the earlier speculation that all *Oryza* species have originated from the common ancestor prevalent in the Gondwanaland continent, which subsequently got dispersed to geographically isolated regions of the world following continental drift (Chang 1976).

Acknowledgments We are thankful to Dr. M.T. Jackson, Genetic Resources Center, IRRI, for providing some of the plant materials, and Dr. K. Majumdar, CCMB, India for his helpful suggestions on the manuscript. Financial support from the Rockefeller Foundation's Program on Rice Biotechnology is gratefully acknowledged.

References

Aggarwal RK, Majumdar KC, Lang JW, Singh L (1994) Genetic affinities among crocodilians as revealed by DNA fingerprinting with a Bkm-derived probe. Proc Natl Acad Sci USA 91: 10601–10605

Aggarwal RK, Brar DS, Huang N, Khush GS (1996) Differentiation within CCDD genome species in the genus *Oryza* as revealed by total genomic DNA hybridization and RFLP analysis. Rice Genet Newsl 13: 54–57

Aggarwal RK, Brar DS, Khush GS (1997) Two new genomes in the *Oryza* complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. Mol Gen Genet 254: 1–12

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

Chang TT (1976) The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica 25: 425–441

Dally AM, Second G (1990) Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, Section *Oryza*). Cladistic-mutation and genetic-distance analysis. Theor Appl Genet 80: 209–222

Dellaporta SL, Wood J, Hick JB (1983) A plant DNA mini preparation: version II. Plant Mol Biol Rep 1: 19–21

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783–791

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155: 279–284

Hill M, Witsenboer H, Zabeau M, Vos P, Kesseli R, Michelmore R (1996) PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. Theor Appl Genet 93: 1202–1210

Jena KK, Khush GS (1990) Introgression of genes from *Oryza officinalis* Well ex Watt to cultivated rice, *O. sativa* L. Theor Appl Genet 87: 737–745

Li HW, Weng TS, Chen CC, Wang WH (1961) Cytogenetical studies of *Oryza sativa* L. and its related species I. Hybrids of *O. paraguaiensis* Wedd × *O. brachyantha* A Chev. Et Roehr; *O. paraguaiensis* Wedd × *O. australiensis* Domin. and *O. australiensis* Domin. × *O. alta* Swallen. Bot Bull Acad Sin 2: 79–86

Maheswaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC, Huang N (1997) Polymorphism, distribution and segregation of AFLP markers in a doubled haploid rice population. Theor Appl Genet 94: 39–45

Maksem K, Leister D, Peleman J, Zabeau M, Salamini F, Gebhardt C (1995) A high-resolution map of the vicinity of the RI locus on chromosome V of potato based on RFLP and AFLP markers. Mol Gen Genet 249: 74–81

Morishima H (1969) Phenetic similarity and phylogenetic relationships among strains of *O. perennis*, estimated by methods of numerical taxonomy. Evolution 23: 429–443

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269–5273

Oka HI (1988) Origin of cultivated rice. Japan Scientific Society Press, Tokyo; Elsevier, Amsterdam

Paul S, Wachira FN, Powell W, Waugh R (1997) Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. kuntze) revealed by AFLP markers. Theor Appl Genet 94: 255–263

Saitou N, Nei M (1987) The neighbour joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425

Second G (1985) Evolutionary relationships in the *sativa* group of *Oryza* based on isozyme data. Genet Sel Evol 17: 89–114

Second G (1991) Molecular markers in rice systematics and the evaluation of genetic resources. In: Bajaj YPS (ed) Biotechnology in agriculture and forestry, vol. 14: rice. Springer, Berlin Heidelberg New York, pp 468–494

Sharma SK, Knox MR, Ellis THN (1996) AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. Theor Appl Genet 93: 751–758

Sneath PHA, Sokal RR (1973) (eds.) Numerical taxonomy: the principles and practice of numerical classification. Freeman & Co, San Francisco

Sokal RR, Michener CD (1958) A statistic method for evaluating systematic relationships. Univ Kans Sci Bull 28: 1409–1438

Tateoka T (1962) Taxonomic studies of *Oryza* II. Several species complex. Bot Mag Tokyo 75: 455–461

Vaughan DA (1989) The genus *Oryza sativa* L. current status of taxonomy. IRRI Res Pap Ser No. 138

Vaughan DA (1994) The wild relatives of rice. International Rice Research Institute, Manila, Philippines

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Freijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407–4414

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

Yap IV, Nelson RJ (1996) WINBOOT: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. IRRI Disc Pap Ser 14

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application, Publ no.: EP0534858, no. 92402629.7