Theor Appl Genet (2002) 105:680–690
DOI 10.1007/s00122-002-0973-6

R. K. Aggarwal · V. V. Shenoy · J. Ramadevi
R. Rajkumar · L. Singh

Molecular characterization of some Indian *Basmati* and other elite rice genotypes using fluorescent-AFLP

Received: 30 May 2001 / Accepted: 28 December 2001 / Published online: 2 July 2002

Abstract Cultivated rice is a high-volume, low-value cereal crop providing staple food to more than 50% of the world populace. A small group of rice cultivars, traditionally produced on the Indo-Gangetic plains and popularly known as *Basmati*, have exquisite quality grain characteristics and are a prized commercial commodity. Efforts to improve the yield potential of *Basmati* have led to the development of several crossbred *Basmati*-like cultivars. In this study we have analysed the genetic diversity and interrelationships among 33 rice genotypes consisting of the traditional *Basmati*, improved *Basmati*-like genotypes developed in India and elsewhere, American long-grain rice and a few non-aromatic rice using a DNA marker-based approach – fluorescent-amplified fragment length polymorphism (f-AFLP). Using a set of nine primer-pairs we scored a total of 10,672 data points over all of the genotypes in the size range of 75–500 bp. The scored data points corresponded to a total of 501 AFLP markers (putative loci/genome landmarks) of which 327 markers (65%) were polymorphic. The f-AFLP marker data, which were analysed using different clustering algorithms and principal component analysis, indicate that: (1) considerable genetic variability exists in the analysed genotypes; (2) traditional *Basmati* cultivars could be distinctly separated from the crossbred *Basmati*-like genotypes as well as from the non-aromatic rice; (3) the crossbred *Basmati*-like cultivars from the subcontinent and elsewhere are genetically very distinct; (4) f-AFLP-based clustering, in general, conforms to the putative pedigree of the improved genotypes. Moreover, analysis to ascertain the scope of AFLP as a technique suggests that the polymorphism revealed by three selective primer-pair combinations is sufficient to obtain reliable estimates of genetic diversity for the type of material used in this study. However, its utility to identify group-specific DNA markers was discounted due to a low frequency of observed group-specific discrete markers.

Keywords *Oryza sativa* · *Basmati* · Aromatic rice · Genetic diversity · DNA polymorphism · Fluorescent-AFLP

Communicated by P. Langridge

R.K. Aggarwal · V.V. Shenoy · J. Ramadevi · R. Rajkumar
L. Singh (✉)
Centre for Cellular and Molecular Biology, Uppal Road, Tarnaka, Hyderabad 500 007, India
e-mail: lalji@gene.ccmbindia.org
Fax: +91-40-7160591

Introduction

Rice (*Oryza sativa* L.) feeds more than 50% of the tropical populace. Though a high-volume, low-value commodity, a class of aromatic, superfine grade premium rice has evolved its own market niches, making rice trade a commercial success internationally. A class of premium rice with specific grain characteristics, traditionally grown on either side of the Indus river, is popularly known as the *Basmati* rice. The superiority of the *Basmati* cultivars over the other premium rice is due to its superfine grains having a distinct aroma and excellent elongation ability and the soft, flaky texture of the cooked rice (Khush and Juliano 1985; Siddiq and Shobha Rani 1998).

The traditional *Basmati* rice cultivars that command a very high price in domestic and international markets due to their superior grain characteristics have some agronomically less desirable traits leading to a low yield potential. Consequently, there have been extensive breeding efforts to improve the yield potential of the *Basmati* by crossing them with high-yielding, short-stature plant types. These efforts have led to the development of many *Basmati*-like high-yielding cultivars, both in India and elsewhere. It is foreseen that more crossbred *Basmati*-like cultivars will follow (Khush and dela Cruz 1998). However, for such efforts to be successful, the genetic variability available to breeders should be assessed as accurately as possible.

In addition, implications of the trade-related intellectual property rights under the World Trade Organization

agreements demand an unambiguous identification and distinction of the crossbred *Basmati* types from the traditional *Basmati* genotypes, thereby necessitating newer means that can effectively differentiate between the two types. The latter is also desirable as a means to check the possibility of trade-related malpractices wherein crossbred cultivars can be traded as genuine traditional *Basmati* rice.

A large number of methodologies are available for the assessment of genetic variability, diversity and interrelationships in the germplasm, as well as for individualization through macromolecular fingerprinting. While the protein-based technologies are influenced by environment and more complex to analyze, the DNA-based ones have provided reliable tools, enabling not only the assessment of genetic variability but also a high-throughput individual DNA typing (Bligh 2000). Circumventing the limitations and drawbacks of several molecular profiling and fingerprinting techniques, such as restriction fragment length polymorphism, random amplified polymorphic DNA, sequence-tagged repeats, etc., Zabeau and Vos (1993) introduced the amplified fragment length polymorphism (AFLP) technique as a means of precise genotyping. The technique does not need prior sequence information of the genome being studied, is reproducible, gives the best genome-wide coverage and has the advantage of being amenable to semi-automated genotyping (Zhu et al. 1998; Aggarwal et al. 1999; Sharma et al. 2000). Therefore, AFLP has become a method of choice in plant research for the study of genetic diversity, varietal identification, mapping, quantitative trait loci analysis as well as gene isolation. In the investigation reported here, we used the fluorescence-based AFLP (f-AFLP) approach to compare 33 rice genotypes comprising the traditional *Basmati* and *Basmati*-like rice genotypes for their genetic diversity and interrelationships in order to assess the feasibility of obtaining individual genotype-specific profiles for documentation. An attempt was also made to define the minimum size of analysis needed to derive at reliable estimates of genetic diversity.

Materials and methods

Plant material

The 33 rice genotypes used in the study are listed in the Table 1. These genotypes were broadly classified into six groups: the traditional *Basmati* rice (TB); high-yielding, crossbred *Basmati*-like rice from Indian sub-continent (HYB); improved, patented, aromatic rice (IPR); American long-grain rice, both aromatic and non-aromatic (AmR); non-premium *japonica* rice (JapR) and *indica* rice (IndR). The seeds of the cultivars/genotypes were obtained from the Directorate of Rice Research, Hyderabad, India.

DNA extraction

Genomic DNA was isolated from green leaves of young seedlings. Each sample comprising leaves pooled from five to eight seedlings (approx. 5 g) was processed as per Dellaporta et al. (1983) until the crude-DNA precipitation step. The crude DNA mat was then spooled-out, washed twice with 70% ethanol, dissolved in 3–5 ml TES (10 m*M* Tris, 1 m*M* EDTA, 5 *M* NaCl, pH: 8.0) containing 20 µg/ml RNAse-A, incubated at 37 °C for 30 min and extracted with chloroform:iso-amyl alcohol (24:1, v/v); the DNA was then re-precipitated with two volumes of chilled absolute ethanol. The purified DNA was finally dissolved in TE (10 m*M* Tris, 0.1 m*M* EDTA, pH 8.0) buffer and checked for its quality and quantity.

AFLP analysis

The AFLP analysis was carried out using the fluorescent dye-based AFLP Plant Mapping Kit from Perkin Elmer Applied Biosystems (Foster City, Calif.) following the manufacturer's instructions with modifications. About 0.5 µg of genomic DNA was used for each sample. The sequences of adaptors, primers, selective primer-pair combinations and fluorescence tags used are detailed in Table 2. Each selective primer contained three unique nucleotides at the 3′ end. The selective amplification was carried out in 10 µl volumes using the PE9600 thermocycler, as per the cycling conditions described by the manufacturer. The amplified product was mixed with internal lane standard GeneScan-500 TAMRA or ROX (Perkin Elmer Applied Biosystems) and denatured at 95 °C for 5 min. Finally, 0.5–1.0 µl from each sample was analysed on the ABI Prism 377 automated DNA sequencer (Perkin Elmer Applied Biosystems).

Data analysis

The AFLP electrophoretograms in the range of 75–500 bp were analysed and compared using the software GENESCAN version 3.1 and GENOTYPER version 2.1 (Perkin Elmer Applied Biosystems), respectively. Each fragment size was treated as a unit character for analysis and converted to binary code (1/0 = +/–), and binary matrices were generated for all of the genotypes for each primer-combination separately. The latter were re-checked manually by comparing them with the corresponding AFLP gel images to take care of the doubtful data points that were apparently sized due to noise signal. For analysing genetic diversity and testing the robustness of the AFLP marker data, we used multiple data sets; (1) marker data sets for individual primer-pairs; (2) data sets created by pooling the marker data for different, randomly chosen combinations of 2, 3, 4, 5, 6, 7 and 8 primer-pairs; (3) whole data pooled over all primer-pairs. In each case, the 1/0 matrix was used to calculate the genetic (dis)similarities as Dice coefficients (Nei and Li 1979). The resulting similarity matrices from the pooled data set were utilized to ascertain the genetic inter-relationships by: (1) partitioning the variance of the data sets using principal component analysis (PCA) and plotting the first two principal components; (2) constructing phenetic trees using UPGMA (unweighted pair grouping with arithmetic mean averages). Cluster analysis to infer the phenetic relationships was also done using the maximum likelihood method MIX employing directly the pooled 1/0 matrix. The reliability, goodness of fit and robustness of the phenetic trees were tested by deriving the cophenetic correlations (Sneath and Sokal 1973) and bootstrapping (Felsenstein 1985). The similarity matrices generated from different sub-sets of AFLP data were subjected to Mantel test (Mantel 1967) to ascertain the size of analysis needed for deriving reliable diversity estimates. The analyses were done using different routines available in the software packages NTSYS-PC version 2.02i (Applied Biostatistics, New York), WINBOOT (Yap and Nelson 1996) and PHYLIP version 3.57c (http://evolution.genetics.washington.edu/phylip.html).

Results and discussion

The assessment of genetic diversity is important not only for crop improvement efforts but also for efficient manage-

Table 1 Details of rice genotypes used in the present study for fluorescent-AFLP-based genetic diversity analysis

Selection no.	Genotype	Origin	Putative pedigree	Main characteristics
	Traditional *Basmati* genotypes (TB)			
1	*Basmati*-370	Kapurthala (Punjab), India	Native genotype	Aromatic, long slender grain, tall plant stature
2	Karnal Local	Kaul (Haryana), India	Native genotype	Aromatic, long slender grain, tall plant stature
3	Taroari *Basmati*	Kaul (Haryana), India	Selection from local *Basmati*	Aromatic, long slender grain, tall plant stature
4	Ranbir *Basmati*	R S Pura (J & K), India	Selection from *Basmati*-370-90-95	Aromatic, long slender grain, tall plant stature
5	*Basmati*-397 (A)[a]	Kapurthala (Punjab); Pakisatan	Native genotype	Aromatic, long slender grain, tall plant stature
6	*Basmati*-397 (B)[a]	Kapurthala (Punjab); Pakisatan	Native genotype	Aromatic, long slender grain, tall plant stature
7	*Basmati*-433	Kapurthala (Punjab); Pakisatan	Native genotype	Aromatic, long slender grain, tall plant stature
8	*Basmati*-122[b]	Kapurthala (Punjab); Pakisatan	Native genotype	Aromatic, long slender grain, tall plant stature
	High yielding improved *Basmati* genotypes from India (HYB)			
9	IET 10363	IARI, New Delhi, India	Pusa167/Karnal Local	Aromatic, long slender grain
10	IET 11341	IARI, New Delhi, India	Pusa167/Karnal Local	Aromatic, long slender grain
11	IET 12021	HAU RS, Kaul (Haryana), India	IR579/*Basmati*-370// *Basmati*-370	Aromatic, long slender grain
12	IET 13162	DRR, Hyderabad, India	PR109/Pakistan *Basmati*	Aromatic, long slender grain
13	IET 13548	DRR, Hyderabad, India	Pusa *Basmati*-1/IET-8585	Aromatic, long slender grain
14	Pusa *Basmati*-1	IARI, New Delhi, India	Pusa 150/Karnal Local	Aromatic, long slender grain, semi-dwarf plant stature
15	Haryana *Basmati*	HAU RS, Kaul (Haryana), India	Sona/*Basmati*-370	Aromatic, long slender grain, semi-dwarf plant stature
16	Kasturi	DRR, Hyderabad, India	*Basmati*-370/CRR-88-17-1-5	Aromatic, long slender grain, semi-dwarf plant stature
17	*Basmati*-385	Kapurthala (Punjab), Pakisatan	T(N)1/*Basmati*-370	Aromatic, long slender grain, tall plant stature
	Improved aromatic genotypes developed and patented in USA (IPR)			
18	*Basmati* 867	Rice Tec, Texas, USA	Pakistani TRB (?)/American rice (?)	Aromatic, long slender grain, high yielding
19	RT 1117	Rice Tec, Texas, USA	Pakistani TRB (?)/American rice (?)	Aromatic, long slender grain, high yielding
20	RT 1121	Rice Tec, Texas, USA	Pakistani TRB (?)/American rice (?)	Aromatic, long slender grain, high yielding
	Long-grain rice genotypes developed and released from USA (AmR)			
21	Della	Crowley, Louisiana, USA	Rexoro/Delitus//Century/ Rexoro-Zenith	Aromatic, long grain, high yielding
22	Jasmine 85	Beaumont, Texas, USA	IR-262/Jasmine (Khao-Dwak-Mali-105)	Aromatic, long grain, low amylose aontent, high yielding
23	Lebonnet	Beaumont, Texas, USA	Bluebelle//Belle Patna/Dawn	Non-aromatic, traditional southern long-grain, high yielding
24	SkyBonnet	Beaumont, Texas, USA	Bluebelle//Belle Patna/Dawn selection	Non-aromatic, traditional southern long-grain, high yielding
25	Texmont	Beaumont, Texas, USA	[(CI 9881/PI331581//L-201) Selection]/Lemont	Non-aromatic, long-grain, high yielding, anther-culture derivative
	Non-aromatic rice genotypes, *japonica* type (JapR)			
26	Akihikari	Aomori, Japan	?	Short bold grain, low amylose content, dwarf
27	Norin-18	Japan	?	Short bold grain, low amylose content

Table 1 (continued)

Selection no.	Genotype	Origin	Putative pedigree	Main characteristics
	Non-aromatic rice genotypes, *indica* type (IndR)			
28	Dular	India	Native genotype	Long, bold grain
29	IR-24	IRRI, Los Banos, Philippines	IR-8/IR127-2-2	Long slender grain, semi-dwarf plant stature
30	IR-64	IRRI, Los Banos, Philippines	Complex derivative of several crosses	Non aromatic, semi-dwarf plant stature
31	Rasi	DRR, Hyderabad, India	T(N)1/Co29	Non aromatic, semi-dwarf plant stature
32	MTURH-2020	ANGRAU, Hyderabad, India	Heterotic hybrid of *indica*/*indica* type	Non aromatic, semi-dwarf plat stature, heterotic hybrids
33	MTURH-2048	ANGRAU, Hyderabad, India	Heterotic hybrid of *indica*/*indica* type	Non aromatic, semi-dwarf plant stature, heterotic hybrids

[a] Two different samples obtained
[b] Seemingly admixture; labelled as *Basmati*-122

Table 2 Details of the primers and flourophores used in the AFLP analysis

Pre-selective amplification primers				
EcoR-1 specific primer core sequence + A:	*Eco*-A 3′			
Mse-1 specific primer core sequence + C:	*Mse*-C 3′			
Selective amplification primer over-hangs (+3 b on 3′ end)				
*Eco*RI specific	*Eco*-AGG	*Eco*-ACA	*Eco*-ACC	*Eco*-ACT
*Mse*I specific	*Mse*-CTC	*Mse*-CAG	*Mse*-CTT	*Mse*-CAA
Nine primer combinations used				
Eco-AGG+*Mse*-CTC	*Eco*-ACA+*Mse*-CAC		*Eco*-ACA+*Mse*-CAG	
Eco-ACA+*Mse*-CTT	*Eco*-ACC+*Mse*-CAA		*Eco*-ACT+*Mse*-CTA	
Eco-ACT+*Mse*-CTC	*Eco*-ACT+*Mse*-CTC		*Eco*-AGG+*Mse*-CAA	
Fluorescent tags of selective primers				
Eco-AGG: JOE (green)		*Eco*-ACA: FAM (blue)		
Eco-ACT: FAM (blue)		*Eco*-ACC: TAMRA (yellow)		

ment and protection of the germplasm resources. For this purpose, the recent DNA fingerprinting approaches capable of revealing extensive variability have become methods of choice. In the present study, we analysed a number of premium *Basmati* cultivars using the powerful DNA fingerprinting approach of AFLP, employing fluorescent dye-labelling and semi-automated detection technology. The present investigation and our earlier experiences with radioactive AFLP (Aggarwal et al. 1999) clearly indicate that the fluorescence-based AFLP technique offers significant improvements over the original method by increasing the scoring and typing efficiency as well as reducing the risk of exposure to the hazardous radioactive isotopes and probably the related costs of the analysis, an opinion shared by Mitchell et al. (1997) and Sharma et al. (2000).

AFLP markers in the analysed rice materials

The summary of AFLP markers produced by nine primer-pairs and their distribution across the rice groups are given in the Tables 3 and 4. Preliminary experiments involving three independent attempts to amplify AFLP markers for five genotypes with three different selective primer-pairs revealed a high reproducibility of the data (results not shown). The nine primer-pairs generated a total of 501 differently sized fragments of which 327 (65%) were polymorphic over all the genotypes. In total, 10,672 marker data points could be scored with an average of 1,186 ± 116 markers per primer-pair, across the genotypes, thereby confirming the high multiplex ratio expected for the AFLPs (Powell et al. 1996). The frequency of individual polymorphic AFLP marker ranged from 0.033 (i.e. present only in one genotype) to 1.0 (present in all). The capability of different primer-pairs to generate AFLP markers varied significantly, ranging from 35 to 90 (average of 55.7±5.4) markers per primer pair over all genotypes, and 22 to 61 (average of 35.6±3.6) markers per primer pair per genotype. Nonetheless, the polymorphic information content (PIC) measured as percentage polymorphic markers for all the primer pairs was high and varied in a relatively narrow

Table 3 Summary of the AFLP markers scored for 33 elite genotypes of rice using nine selective primer-pairs each with 6-bp selection (*SD* standard deviation, *SEM* standard error of the mean)

Parameter	6-bp Selective primer-pair combinations									Across nine primer-pairs			
	AAG-CTC	ACA-CAC	ACA-CAG	ACA-CTT	ACC-CAA	ACT-CTA	ACT-CTC	AGG-CTT	AGG-CAA	Total markers	Average markers	SD	SEM
Total markers[a]	46	44	47	68	35	57	62	90	52	501	55.67	16.27	5.42
Average markers[b]	27.40	30.60	31.80	39.60	22.00	34.00	38.40	60.60	36.00	320.40	35.60	10.86	3.62
SD	1.95	2.51	1.48	1.82	1.00	2.24	1.67	4.34	4.00				
SEM	0.34	0.44	0.26	0.32	0.17	0.39	0.29	0.75	0.70				
Polymorphic markers[c]	30	32	27	48	21	37	40	59	33	327	36.33	11.49	3.83
Polymorphic markers (%)	65	73	57	71	60	65	65	66	63		65	5.00	2.00
Total data points[d]	921	1,003	1,030	1,354	764	1,215	1,302	1,975	1,108	10,672	1,185.78	349.5	116.5

[a] Total number of differently sized AFLP markers amplified across all 33 genotypes
[b] Average number of AFLP markers scored per genotype
[c] Total number of AFLP markers found to be polymorphic across the 33 genotypes
[d] Total number of AFLP markers scored for all genotypes

range of 57% (for *Eco*-ACA:*Mse*-CAG) to 73% (for *Eco*-ACA:*Mse*-CAC), with an average of 65±2%, suggesting that despite large variations in the marker index there is a near-uniform distribution of AFLPs (polymorphisms) across the genome, which is a desirable attribute for their utility in diversity analysis.

An advantage of AFLP fingerprinting is its potential to expose large genetic polymorphism, thereby providing broader genome coverage. In the present study, a very large number of markers were generated, with an average of 320±2.7 markers per genotype over all nine primer-pairs and 35.6±3.6 markers per genotype per primer pair. These multiplex marker indices are comparable to the ones reported earlier for diploid rice cultivars/species using the radioactive AFLP approach (Aggarwal et al. 1999). In comparison to a recent AFLP map of the rice with 208 markers spanning approximately 1,500 cM of genome (Maheshwaran et al. 1997), the AFLP markers scored in the present study would be equivalent to more than 320 nuclear loci per genotype with all nine primer-pairs, thus providing extensive coverage of the whole genome. Accordingly, the genetic groupings inferred from the present data were supported by the high bootstrap and cophenetic values (see below), indicating the robustness of the data for this purpose.

A perusal of the marker data for their group-wise distribution revealed that while the average number of markers scored per genotype in different rice groups over all the primers were almost similar (ranging between 34.8 and 36.7 markers), the level of diversity discernible from percentage polymorphic markers varied extensively from a low of 29% for IPR to a high of 61.4% for the IndR group (Table 4). These results suggest that the patented aromatic rice group is the least genetically diverse one, whereas the IndR group is the most diverse. Indeed, the IndR group has genotypes from diverse origins and derivations. In addition, heterotic hybrid rice entries MTURH-2020 and MTURH-2048 are also included, adding to the variation. On the other hand, the IPR genotypes studied herein are supposed to have been derived from a closely related gene pool (Sarreal et al. 1997), depicting the least divergence. The above demonstrates that the AFLP analysis can provide reliable indices that are proportional to the genetic base of the germplasm collections and thus has great potential in biodiversity studies.

Genotype-specific profiles and fingerprinting

The data revealed a considerable number of polymorphic markers within each rice group (Table 4). Group-wise, the number of such differentiating AFLP markers ranged from a minimum of 32 in the JapR group comprising only two entries to 199 in the most diverse IndR group comprising six entries. The occurrence of such a large number of polymorphic markers within each group suggests the potential of the AFLP approach to generate genotype-specific fingerprints useful for germplasm reg-

Table 4 Summary of polymorphic and unique AFLP markers scored in different rice groups comprising 32 genotypes

Rice groups	6-bp Selective primer pair combinations used									Total[a]	Average
	AAG-CTC	ACA-CAC	ACA-CAG	ACA-CTT	ACC-CCA	ACT-CTA	ACT-CTC	AGG-CTT	AGG-CAA		
Traditional *Basmati* genotypes (TB)[b]											
Average markers	29.1	30.1	31.3	41.3	23.3	38.1	41.6	60.6	33.3		36.5
Polymorphic markers	13	11	13	22	10	19	16	24	16	144	16.0
Polymorphism (%)	44.6	36.5	41.6	53.3	42.9	49.8	38.5	39.6	48.1		43.9
Unique markers[c]	2	0	1	2	1	1	1	1	0	9	
Cross-bred *Basmati*-like genotypes from India (HYB)											
Average markers	28.4	30.4	31.6	42.6	22.4	38.7	40.0	62.0	34.4		36.7
Polymorphic markers	18	14	13	33	9	26	21	33	22	189	21.0
Polymorphism (%)	63.3	46.0	41.2	77.5	40.1	67.2	52.5	53.2	63.9		56.1
Unique markers[c]	0	0	0	0	0	3	1	0	0	4	
Patented improved aromatic rice from USA (IPR)											
Average markers	26.3	30.3	30.3	39.0	23.0	36.3	37.3	58.3	33.7		35.0
Polymorphic markers	7	10	9	10	6	10	9	21	11	93	10.3
Polymorphism (%)	26.6	33.0	29.7	25.6	26.1	27.5	24.1	36.0	32.7		29.0
Unique markers[c]	0	1	1	1	0	0	0	1	0	4	
American long-grain rices (AmR)											
Average markers	27.4	30.2	30.0	40.8	24.6	35.8	39.2	58.8	31.8		35.4
Polymorphic markers	9	13	14	23	13	19	23	34	22	170	18.9
Polymorphism (%)	32.8	43.0	46.7	56.4	52.8	53.1	58.7	57.8	69.2		52.3
Unique markers[c]	1	0	1	0	1	0	1	0	1	5	
Non-aromatic *japonica* type (JapR)											
Average markers	28.5	30.0	29.5	42.5	24.5	36.0	37.5	55.5	29.0		34.8
Polymorphic markers	11	4	3	3	1	4	3	1	2	32	3.6
Polymorphism (%)	38.6	13.3	10.2	7.1	4.1	11.1	8.0	1.8	6.9		11.2
Unique markers[c]	0	0	1	3	0	1	0	2	1	8	
Non-Aromatic *indica* type (IndR)											
Average markers	28.2	30.2	31.8	38.7	22.3	34.0	39.0	59.3	35.0		35.4
Polymorphic markers	18	19	15	33	11	19	26	38	20	199	22.1
Polymorphism (%)	63.9	63.0	47.1	85.3	49.3	55.9	66.7	64.0	57.1		61.4
Unique markers[c]	1	1	2	2	2	0	5	2	2	17	
Total unique markers	4	2	6	8	4	5	8	6	4	47	

[a] Over all nine primer-pairs
[b] Data of *Basmati*-122, possibly an admixture, is not included
[c] Group-specific AFLP markers, but may not be in all members

Table 5 Average genetic similarities calculated as Dice coefficients for different groups of rice genotypes based on AFLP markers (*SD* standard deviation)

Group	Genetic similarity between the groups[a]						Within the group	
	TB	HYB	IPR	AmR	JapR	IndR	Mean±SEM	SD (*n*)
TB		0.035 (63)[b]	0.018 (21)	0.031 (35)	0.018 (14)	0.021 (42)	0.914±0.006	0.030 (21)
HYB	0.864±0.004		0.015 (27)	0.020 (45)	0.014 (18)	0.027 (54)	0.877±0.005	0.029 (36)
IPR	0.871±0.004	0.836±0.003		0.053 (15)	0.011 (6)	0.016 (18)	0.901±0.004	0.007 (3)
AmR	0.845±0.005	0.822±0.003	0.884±0.014		0.066 (10)	0.033 (30)	0.882±0.030	0.094 (10)
JapR	0.834±0.005	0.793±0.003	0.872±0.005	0.875±0.021		0.014 (12)	0.949	
IndR	0.805±0.003	0.853±0.004	0.800±0.004	0.796±0.019	0.758±0.004		0.862±0.009	0.036 (15)

[a] Data for *Basmati*-122 that is seemingly an admixture is not included in the summary statistics

[b] Figures in parenthesis are the number of pair-wise comparisons; values above diagonal are standard deviation of similarity coefficient; values below diagonal are average similarity coefficients with standard error of mean (SEM)

istration and protection. The perusal of the marker profiles of different genotypes reveals them to be genotype-specific with a difference of at least nine AFLP markers between the two closest entries – Pusa *Basmati*-1 and Taroari *Basmati*. Although our results validate the utility of AFLP-based fingerprinting for the identification and authentication of genotypes, it should be mentioned that caution should be exercised if the number of genotypes being compared are large or the entries are closely related. In such cases, the number of unique differentiating polymorphic markers may become a constraint, and one may need to resort to either a lower +3/+2 nucleotide selection strategy, as suggested by Zhu et al. (1998), in place of the +3/+3 strategy employed in the present work to increase the level of detectable AFLP polymorphism or utilize additional marker techniques, such as sequence-tagged repeats or the DNA microsatellite loci. Further, the possibility of identifying AFLP marker(s) that can be diagnostic of the group as a whole is discounted in the present study.

Genetic similarity and inter-relatedness

The genetic similarity estimates, calculated as the Dice coefficient using the entire dataset (excluding the Pakistan *Basmati*-122, the sample of which seems to be an admixture), ranged from 0.737 to 0.980, which is comparable to earlier reports on cultivated rice germplasm (Mackill et al. 1996). In general, within-group estimates of genetic variation corroborated well with the type and number of cultivars that comprised different groups (Table 5). Accordingly, the IndR group was revealed as being the most diverse having the least within-group average similarity (0.862±0.009) but with the broadest range (0.809–0.945). In contrast, the JapR group represented by only two cultivars showed the least variation with the highest within-group average similarity of 0.949. Similarly, the IPR group showed a high average similarity (0.901) within itself, ranging from 0.894 to 0.908. The low variation in the JapR and IPR groups can be due to the lesser number of genotypes; in the latter, a low variation was also expected due to the related and/or overlapping pedigrees. The remaining three groups – TB, HYB and AmR – exhibited similar levels of genetic variation (average genetic similarity of 0.88).

Furthermore, between-group genetic variation was found to be informative about the possible origin/pedigree of the analysed cultivars. The average between-group genetic similarity was observed to be highest (0.884) between the IPR and AmR groups, suggesting that some member(s) of the latter group may be involved in the derivation of the patented lines analysed. It was noteworthy that the patented aromatic rice *Basmati*-867 showed a maximum relatedness of 0.876 with the traditional Ranbir *Basmati*, which happens to be from the Jammu and Kashmir region of India, geographically adjoining Pakistan. This would suggest that the unidentified Pakistani traditional *Basmati* used in the development of *Basmati*-867 could either be Ranbir *Basmati* itself or one of its closely related cultivars. The patented RT1117 was the closest to Della (aromatic AmR), with a similarity of 0.918. All of these similarity estimates strongly suggest that the IPR genotypes analysed are derivatives from crosses involving the traditional *Basmati* and the American long-grain cultivars.

The aromatic American long-grain cultivar Jasmine-85 showed a higher similarity with the crossbred HYB and the IndR groups than with the AmR group. This was not unexpected as Jasmin-85 is a derivative of a cross between IR-262 (a non-aromatic *indica* type from IRRI, Philippines) and Khaw Dawk Mali-105, also known as Jasmine (an aromatic *indica* type from Thailand). Similarly, the higher genetic similarity of AmR to JapR (with an average of 0.884) than with the IndR (average similarity of 0.796) is substantiated by the fact that most of the American rice are *japonica* types (Rutger et al. 2000). The traditional *Basmati* as a group was equidistant from the HYB and IPR groups (approx. 0.86 similarity). This was expected as there is a definite involvement of some members of the traditional *Basmati* group in the development of both the crossbred *Basmati*-like HYB as well as the IPR rice cultivars (Table 1).

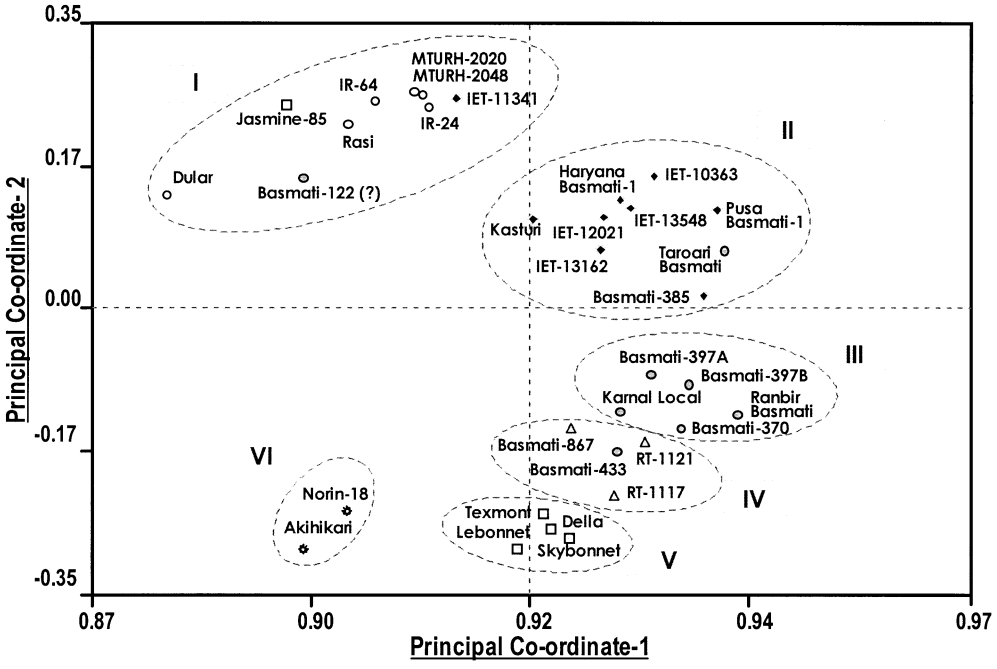

Fig. 1 Two-dimensional plot of genetic diversity among the 33 rice genotypes as revealed by the PCA based on AFLP markers obtained with nine primer-pairs. The first two co-ordinates explain more than 87% of the diversity

Principal component analysis

The PCA is one of the multi-variate approaches of grouping based on the similarity coefficients or variance-covariance values of the component traits of the entities. It is expected to be more informative about differentiation among major groups, while the cluster-analysis provides higher resolution among closely related populations (Liu et al. 2001). In our PCA analysis, more than 90% of the variation in the estimates of genetic similarity was explained by the first three components, indicating the suitability of the AFLP approach for genetic clustering. The first principal component explained a significant variation (approx. 82%) and could clearly separate all of the aromatic/*Basmati* rice genotypes from the non-aromatic rice types. These two broad groups could further be resolved into relatively distinct clusters corresponding to different rice types by the second principal component that explained approximately 4% variation of the data.

Overall, six distinct clusters were revealed by the first two principal components (Fig. 1). Cluster 1 included in addition to all of the members of the IndR group, three genotypes from the other groups, *Basmati*-122 (from Pakistan), IET-11341 and Jasmin-85. Among these exceptions, the inclusion of TB genotype *Basmati*-122 in the IndR group cluster is seemingly due to it being an admixture with some *indica* genotype, as was indicated in an independent study using rice-specific DNA microsatellite markers (data not shown). On the other hand, the grouping of Jasmine-85 in the IndR cluster was rather expected as it is derived from *indica* type parents (see above), whereas in the case of IET-11314 (a derivative of *indica*/TB genotypes) it could be due to a stronger selection bias for the *indica* traits/genetic background during its development.

The second cluster comprised HYB cultivars that are derivatives of traditional *Basmati* cultures crossed with modern high-yielding *indica* types. Accordingly, it was placed between the clusters I and III comprising IndR and TB genotypes, respectively. Interestingly, the traditional Pakistan *Basmati*-433 clustered along with the IPR genotypes, suggesting its close genetic affinity with the patented aromatic rice analysed. This is supported by the US Patent document (no. 5663484) on the patented aromatic rice, which shows it as one of the parents used (Sarreal et al. 1997). Cluster IV of the patented aromatic rice, similar to cluster II, was placed between cluster III of TB and cluster V of the AmR genotypes, in line with their parentage. Similarly, cluster V comprising AmR genotypes was found to be nearest to cluster VI of the JapR genotypes, suggesting their origin from the *japonica* types (Rutger et al. 2000). All these observations amply demonstrate the utility of AFLP data in deducing the genetic lineage(s) of the crossbred genotypes.

The UPGMA cluster analysis

The cluster analysis using UPGMA and MIX, based on similarity coefficients, was done to resolve the closer phenetic relationships between the analysed rice genotypes. Both approaches produced similar phenograms, with the exceptions of minor variations in the branch lengths and the positioning of some genotypes on the terminal branches (data not shown). In general, the phenetic clusters of genotypes were supported with high bootstrap values (Fig. 2), indicating the reliability and stability of the inferred relationships as well as the robustness of the AFLP data used for diversity analysis. The goodness of the AFLP data for cluster-analysis was also supported by

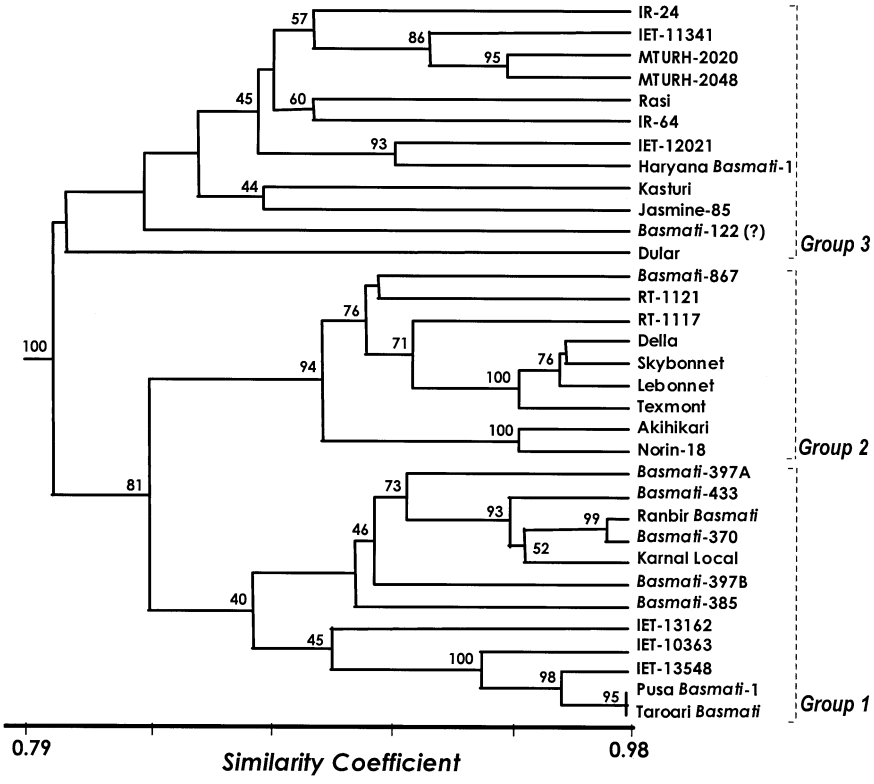

Fig. 2 UPGMA clustering of 33 rice genotypes based on AFLP markers obtained with nine selective primer-pairs. The *numbers* shown at different nodes represent the bootstrap values

high (>0.82) cophenetic correlation coefficients (Rohlf 1992).

The UPGMA cluster analysis provided a better resolution of the relationships among the rice genotypes, which were broadly clustered into three major groups (Fig. 2). It was interesting to note that all of the traditional and some of the crossbred *Basmati* types comprised Group 1 which was distinctly separated from Group 2, the latter including all of the genotypes of AmR, IPB and JapR except Jasmine-85. Both these groups were distinct from the relatively more diffused Group 3 that consisted of diverse IndR and a couple of crossbred *Basmati* genotypes. Within Group 1, the genotypes were distinctly separated into the HYB and the TB types, sharing a genetic similarity of around 87%. Similarly, within Group 2, three well-defined sub-clusters were observed corresponding to the AmR, IPR and JapR types, resolving from each other around at the 92%, 89.5% and 88% genetic similarity levels, respectively. While Group 3 was separated from the other two groups at around 80% genetic similarity, Group 1 shared slightly more genetic similarity – around 84% – with Group 2. The latter was expected as some of the Group 1 TB genotypes have contributed to the development of IPR genotypes of Group 2. These results, in general, are in agreement with what was observed in terms of similarity coefficients and the PCA analysis.

Among the crossbred *Basmati* genotypes from the subcontinent, there were two distinct sub-groups – one (IET-10363, 13548, Pusa *Basmati*-1) nearer to the traditional *Basmati* in Group 1 and the other (IET-12021, Haryana *Basmati*-1, Kasturi) to the non-aromatic *indica* genotypes in Group 3. The genotypes of the former subgroup invariably have the traditional *Basmati* Karnal Local as one of the parents, whereas those of the latter subgroup involved traditional *Basmati*-370 in their pedigree. This apparent genetic affinity towards a specific parent is probably the result of the selection bias for a particular plant type and/or differences in the general combining ability of the parents involved. Accordingly, it is a possibility that the traditional *Basmati* Karnal Local genotype may have a better combining ability than the traditional *Basmati*-370, with the elite *indica* genotypes involved in the *Basmati* improvement programmes.

Size of AFLP analysis for reproducible biodiversity estimates

In most diversity studies, one important point of consideration is how many markers one must analyse to rightly mirror the variations representative of the whole genome in order to derive reliable estimates of biodiversity. To date, there are no set rules except the realization that analysis should ideally provide wide genome coverage. In the present study, the average numbers of AFLP markers scored per genotype were more than the total number mapped on the AFLP linkage map of rice (Maheshwaran et al. 1997), suggesting that the number was probably in excess of what may optimally be needed for the diversity estimation. Furthermore, the earlier mapping studies suggest that the AFLP markers generat-

Table 6 Mantel statistic and cophenetic correlations for comparison of similarity matrices derived from AFLP marker data sub-sets of different sizes

Number of selective primer-pairs that comprised the data sub-set	Average correlation *r* between matrices derived from data sub-sets[a] of the same size	Number of total pair-wise matrices comparisons made for all the randomly generated data sub-sets[b]	Average correlation *r* between the matrices of individual data sub-sets and the entire data set (pooled over the nine primer-pairs)	Average cophenetic correlation coefficient
One	0.48 (0.48)	36 (36)	0.73	0.78
Two	0.69 (0.63)	36 (25)	0.84	0.79
Three	0.83 (0.77)	36 (4)	0.90	0.82
Four	0.89	36	0.94	0.83
Five	0.94	36	0.96	0.82
Six	0.96	36	0.97	0.82
Seven	0.97	36	0.98	0.83
Eight	0.99	36	0.99	0.84

[a] In each case only nine sub-data-sets were generated using data from random combinations of individual selective primer-pairs
[b] Values in parenthesis pertain to orthogonal comparisons out of the total 36 pair-wise combinations tested

ed by each primer combination are more or less randomly distributed throughout the genome and thus represent random subsets of the full dataset (Zhu et al. 1998). Therefore, we attempted to determine the minimum number of AFLP markers needed for reliable grouping of the test materials. For this purpose, the Mantel test and cluster analysis were carried out on 72 sub-datasets of AFLP markers generated by combining the basic dataset for individual primer-pairs in different combinations (see Materials and methods). The results revealed that, in general, sub-datasets comprising two or more basic sets produced genotype clusters similar to the one produced by the dataset pooled over all the nine primer pairs. Also, the clusters were supported by reasonably high cophenetic correlation values of greater than 0.8, indicating a good fit of the data for diversity analysis (Table 6). Similarly, the results of the Mantel test show highly significant correlations between the similarity matrices generated from different sub-datasets, suggesting further that the data generated by the use of two or three primer-pairs are rather sufficient to obtain the reliable estimates of genetic similarities among the genotypes tested in the study (Table 6). These results are significant, as our analysis involved the comparisons of a large number of orthogonal datasets, and they by and large substantiate the inferences of Zhu et al. (1998) that in rice the AFLP data generated from two or at the most three primer combinations appear to be sufficient for robust estimates and that there is little advantage in generating data from four or more primer combinations.

In conclusion, the present investigation demonstrates the potential of f-AFLP in exposing extensive genetic polymorphism useful for diversity analysis and genotype individualization. Most interestingly, the analyses not only distinguished the traditional *Basmati* genotypes from the crossbred *Basmati*-like genotypes, but they also revealed that the latter genotypes from the Indian subcontinent are genetically very different from the ones developed elsewhere. Furthermore, the analyses clearly indicated the involvement of the traditional *Basmati* genotype(s) in the derivation of the patented, aromatic genotypes. To a large extent, the observed clustering corresponds well with the available pedigree information, suggesting that AFLP is also a good genetic tool for identifying genetic lineages. The results of the Mantel tests suggest that the analysis with only two to three AFLP primer-pair combinations (with +6 selective bases) is more than sufficient for reliable estimates of genetic diversity in closely related germplasm like the rice material analysed herein.

Acknowledgements Authors are grateful to Drs. K. Krishaniah, former Project Director, N. Shobha Rani, Principal Scientist, Plant Breeding and E.A. Siddiq, National Professor (ICAR), Directorate of Rice Research, Hyderabad, India for the rice samples.

References

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS (1999) Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor Appl Genet 98:1320–1328

Bligh HFJ (2000) Detection of adulteration of *Basmati* rice with non-premium long grain rice. Int J Food Sci Technol 35:257–265

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

Khush GS, Juliano BO (1985) Breeding of high yielding rices for excellent cooking and eating qualities. In: IRRI (ed) Rice grain quality and marketing. International Rice Research Institute, Los Banos, Philippines, pp 61–69

Khush GS, dela Cruz N (1998) Developing *Basmati* rices with high yield potential. Cahiers Options Médeterraneennes 24: rice quality: a pluridisciplinary approach. CIHEAM, Paris, France (available on CD-ROM)

Liu F, Sun GL, Salomon B, Bothmer von R (2001) Distribution of allozymic alleles and genetic diversity in the American barley Core Collection. Theor Appl Genet 102:606–615

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

Maheshwaran 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

Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Mitchell SE, Kresovich S, Jester CA, Hernandez CJ, Szewc-McFadden AK (1997) Application of multiple PCR and fluorescence-based semi-automated allele-sizing technology for genotyping plant genetic resources. Crop Sci 37:617–624

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

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalaski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 3:225–238

Rohlf FJ (1992) NTSYS-PC. Numerical taxonomy and multivariate analysis system, version 1.8. Exeter Publ, New York

Rutger JN, Tai T, Eizenga GC, Bryant RJ (2000) Genetic improvement of rice for yield, quality pest resistance and stress tolerance. Status report of the USDA-ARS in house project no. 6225-21000-003-00D. ARS, Stuttgart, Ark.

Sarreal ES, Mann JA, Stroike JE, Robins AD (1997) *Basmati* rice lines and grains. US Patent No. 5663484. Date issued: September 2, 1997

Sharma A, Sharma R, Machii H (2000) Assessment of genetic diversity in a *Morus* germplasm collection using fluorescence-based AFLP markers. Theor Appl Genet 101:1049–1055

Siddiq EA, Shobha Rani N (1998) Protecting biodiversity and utilizing in rice products. In: Proc 4th IFCON. Mysore, India, pp 987–1000

Sneath PHA, Sokal RR (1973) Numerical taxonomy: The principles and practice of numerical classification. Freeman, San Fransisco

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 Discussion Paper Ser No. 14, International Rice Research Institute, Los Banos, Philippines

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application No. 92402629 (Publ No. 0534858A1)

Zhu J, Gale MD, Quarrie S, Jackson MT, Bryan GJ (1998) AFLP markers for study of rice biodiversity. Theor Appl Genet 96:602–611