Theor Appl Genet (2004) 109: 965–977
DOI 10.1007/s00122-004-1700-2

ORIGINAL PAPER

Sunita Jain · Rajinder K. Jain · Susan R. McCouch

Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers

Received: 28 October 2003 / Accepted: 2 April 2004 / Published online: 7 August 2004
© Springer-Verlag 2004

Abstract Genetic relationships among Indian aromatic and quality rice (*Oryza sativa*) germplasm were assessed using 30 fluorescently labeled rice microsatellite markers. The 69 rice genotypes used in this study included 52 Basmati and other scented/quality rice varieties from different parts of India and 17 *indica* and *japonica* varieties that served as controls. A total of 235 alleles were detected at the 30 simple sequence repeat (SSR) loci, 62 (26.4%) of which were present only in Basmati and other scented/quality rice germplasm accessions. The number of alleles per locus ranged from 3 to 22, with an average of 7.8, polymorphism information content (PIC) values ranged from 0.2 to 0.9, with an average of 0.6, and the size range between the smallest and the largest allele for a given microsatellite locus varied between 3 bp and 68 bp. Of the 30 SSR markers, 20 could distinguish traditional Basmati rice varieties, and a single panel of eight markers could be used to differentiate the premium traditional Basmati, cross-bred Basmati, and non-Basmati rice varieties having different commercial value in the marketplace. When estimates of inferred ancestry or similarity coefficients were used to cluster varieties, the high-quality Indian aromatic and quality rice genotypes could be distinguished from both *indica* and *japonica* cultivars, and crossbred varieties could be distinguished from traditional Basmati rices. The results indicate that Indian aromatic and quality germplasm is genetically distinct from other groups within *O. sativa* and is the product of a long, independent pattern of evolution. The data also suggest that there is scope for exploiting the genetic diversity of aromatic/quality rice germplasm available in India for national Basmati rice breeding programs.

Communicated by P. Langridge

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-004-1700-2.

S. R. McCouch (✉)
Department of Plant Breeding, Cornell University,
240 Emerson Hall,
Ithaca, NY 14853-1901, USA
e-mail: SRM4@cornell.edu
Tel.: +1-607-2550420

S. Jain · R. K. Jain
College of Basic Sciences and Humanities, CCS Haryana
Agricultural University,
Hisar, 125004, Haryana, India

Introduction

Aromatic rices constitute a special group of rice accessions well known for their aroma and/or superfine grain quality (Ahuja et al. 1995; Khush and de la Cruz 1998; Kumar et al. 1996; Nene 1998; Singh et al. 2000a, b). The center of diversity of Indian aromatic rices is the foothills of the Himalayas in the Indian states of Uttar Pradesh and Bihar and the Tarai region of Nepal. From there, aromatic rices spread northwestwards and northeastward to other Indian states and neighboring countries, and today there are many locally adapted genotypes of aromatic and quality rices (Khush 2000). The Indian aromatic and quality rice germplasm is comprised of small-, medium-, and long-grain types with mild to strong aroma (Singh et al. 2000a, b). Of these, aromatic rice types with long, slender grains (having a kernel length of 6 mm and more), a length to breadth (L/B) ratio of three and above, and high kernel elongation after cooking have been categorized as Basmati. Based on conventional taxonomy, Indian aromatic rices have been classified as *indicas* (Khush 2000), but the plants have a morphologically distinctive phenotype, and both isozyme and microsatellite data have shown that most of the aromatic rices of the Indian sub-continent, including Basmati types, are identifiable as a genetically distinct cluster (Aggarwal et al. 2002; Garris et al. 2003; Glaszmann 1987; Nagaraju et al. 2002).

The popularity of Basmati rice is increasing internationally. The export of Basmati rice from India, which was 67,100 tons in 1978, was 638,380 tons in 2000, a ninefold increase, and is expected to continue to rise. Basmati rices command premium prices, typically two to three times higher than lower quality rices in both domestic and

international markets (Bhasin 2000). For example, in India the 2000 market value of *Taraori Basmati* paddy rice was twice that of *Pusa Basmati 1*. As a result, adulteration of superior Basmati rice supplies with less expensive and/or lower quality *indica* or Basmati rice is quite common. DNA fingerprinting has been suggested as a means of differentiating high-quality Basmati rice from other, cheaper long-grain rice cultivars and of documenting instances and levels of seed mixing in the marketplace.

Microsatellites [also known as simple sequence repeats (SSRs)] are simple, tandemly repeated, di- to tetra-nucleotide sequence motifs flanked by unique sequences (Hamada et al. 1982). They are valuable as genetic markers because they detect high levels of allelic diversity, are co-dominant, are easily and economically assayed by PCR (Weber and May 1989), and are easily automated (Reed et al. 1994; Smith 1998; Westman and Kresovich 1997). Thousands of microsatellites have been identified in rice, and over 2,500 have been developed as molecular markers (for review, see McCouch et al. 2002; Temnykh et al. 2001). Blair et al. (2002) and Coburn et al. (2002) recently assembled a sub-set of these markers into kits for semi-automated fluorescent detection.

In the investigation reported here, we used 30 SSRs combined into four multiplex panels to fingerprint 69 diverse *O. sativa* accessions. The varieties included Basmati and other aromatic/quality rice germplasm collections from different parts of the Indian sub-continent as well as non-Basmati (*indica* and *japonica*) cultivars. The SSR data was used to evaluate the level of genetic diversity within Basmati rices, to assess genetic relationships among the accessions, and to determine whether existing SSR markers provide adequate power of resolution to discriminate between high- and low-quality aromatic varieties for use in commercial evaluation.

Materials and methods

Plant material

A total of 69 rice genotypes were evaluated in this study, including 52 varieties/accessions from Basmati and quality rice germplasm available in India (Table 1). Five varieties were selected to represent the *japonica* sub-species: cv. *Nipponbare*, a temperate *japonica* from Japan, cv. *Azucena*, an aromatic tropical *japonica* from the Philippines, and three additional tropical *japonicas* classified as "new plant type" lines developed at the IRRI, Philippines. Twelve genotypes were selected to represent the *indica* subspecies: cvs. IR36 and IR64 (from IRRI), cvs. *Gobind*, HKR120, and HKR126 (high-yielding varieties developed at the Rice Research Station Kaul, Haryana, India), cvs. CSR-10, -11, -13 and -18 (salt-tolerant lines developed at the CSSRI, Karnal, India), cv. BS125 (from the Ivory Coast), cv. *Kasalath* (an Indian landrace), and cv. *Pokkali*, a highly salt-tolerant landrace (from Kerala, India). Varieties IR36, IR64, *Azucena*, *Kasalath*, BS125 and *Nipponbare* also served as controls for determining allele molecular weight in this study because they had previously been assayed using the same SSR markers on the automated system of Coburn et al. (2002) as well as on silver-stained gels (Cho et al. 2000).

Plants were raised in pots in a net house and, at 3 weeks of age, about 2 cm of leaf from each of five plants was harvested and bulked for each genotype. Total genomic DNA was extracted from the bulked leaf samples by the potassium acetate method (Dellaporta et al. 1983).

Panel design

Thirty SSRs were selected from those originally mapped by Chen et al. (1997) and Temnykh et al. (2001). The SSRs were divided into four panels (A, B, C and D) containing seven to eight fluorescently labeled markers each (Table 2). Two of these panels (C and D) are the same as those described in Blair et al. (2002), and the other two were designed independently. The markers were selected from all 12 chromosomes and combined into multiplex panels based on the range of allele molecular weights reported by Cho et al. (2000) and Coburn et al. (2002). Markers were combined so that the allele size ranges of those labeled with the same fluorophore dye would differ by at least 20 bp, while markers that overlapped in allele size were distinguished by being labeled with different colored dyes. Panels A and D included markers from the same chromosome arms, thereby providing an opportunity to evaluate genetic diversity at linked markers, while panels B and C contained markers from different chromosomes or chromosome arms. Forward primers were labeled with either hexachloro-6-carboxyfluorescein (HEX), tetrachloro-6-carboxyfluorescein (TET) or 6-carboxyfluorescein (FAM) dye phosporamidites and synthesized on an Applied Biosystem 392 apparatus (Foster City, Calif.) by the Cornell Bioresource Facility.

Fluorescently-labeled microsatellite marker analysis

Individual PCR amplifications for each microsatellite were performed using the PTC100 96 V thermocycler (MJ Research, Watertown, Mass.). The PCR protocols were a slightly modified version of those described by Chen et al. (1997) and involved a total volume of 15 μl containing 20 ng genomic DNA, 1 pmol of each primer, 1.5 μl of 10× buffer (0.1 *M* Tris pH 8.3, 0.5 *M* KCl; 7.5 m*M* $MgCl_2$, 0.1% gelatin), 100 m*M* each of dNTPs and 1.0 U *Taq*. The PCR profile was: 5 min at 94°C; 35 cycles of a 1-min denaturation at 94°C, a 1-min annealing at 55°C, and a 2-min extension at 72°C; 5 min at 72°C for final product extension. The PCR products for each set of microsatellites were mixed together in a ratio of 1:2:4 for FAM:TET: HEX. The fluorophore dye HEX gives a weaker signal, and hence the PCR products labeled with HEX were used in higher amounts compared to products labeled with the other dyes. A 0.5-μl aliquot of the mixed microsatellite samples was combined with 1 μl of a loading buffer (98%

Table 1 List of rice varieties/germplasm accessions used for SSR analysis

Sample no.	Genotype	Abbreviations	Type	Source[a]	Accession no.
1	Acc277805	ACC	Aromatic/quality	DRR	1010/129(A)
2	AKP-1	AKP-1	Aromatic/quality	DRR	1019/292
3	*Ambemohar*	Ambe	Aromatic/quality	DRR	1001/54
4	*Ambemohar 157*	Ambe157	Aromatic/quality	DRR	1003/881
5	*Ayepyung*	A-pyung	Aromatic/quality	DRR	1017/181(B)
6	*Azucena*	Azucena	*Japonica*	Cornell	RA199
7	*Barhail*	Barhail	Aromatic/quality	DRR	1292/705
8	*Basmati 150*	B150	Basmati	DRR	1165/111
9	*Basmati 213*	B213	Basmati	DRR	1100/246
10	*Basmati 217*	B217	Basmati	DRR	1061/16
11	*Basmati 334*	B334	Basmati	DRR	1246/669
12	*Basmati 370*	B370	Premium Basmati	CCSHAU	na[c]
13	*Basmati 370B*	B370B	Premium Basmati	DRR	1162/109
14	*Basmati 370* (Bangladesh)	B370BD	Premium Basmati	DRR	1187/195
15	*Basmati 370* (Pakistan)	B370P	Premium Basmati	DRR	1072/508
16	*Basmati 372A*	B372A	Basmati	DRR	1250/674
17	*Basmati 397*	B397	Basmati	DRR	1083/262
18	*Basmati 405*	B405	Basmati	DRR	1139/102B
19	*Basmati 5836*	B5836	Basmati	DRR	1162/109
20	*Basmati 6113*	B6113	Basmati	DRR	1199/204
21	*Basmati 6131*	B6131	Basmati	DRR	1203/206
22	*Basmati 6313*	B6313	Basmati	DRR	1211/211a
23	*Basmati A3-3*	BA3-3	Basmati	DRR	1065/53
24	*Basmati Bahar*	Bahar	Basmati	DRR	1060/484
25	*Basmati Kamon*	Kamon	Basmati	DRR	1104/107
26	*Basmati Mehtrah*	Mehtrah	Basmati	DRR	1046/14
27	*Basmati Norat 439*	Norat	Basmati	DRR	1077/220
28	*Basmati Sathi*	Sathi	Basmati	DRR	1063/496
29	*Basmati Sufaid 100*	Sufaid	Basmati	DRR	1135/100A
30	*Basmati T3* (Haiti)	BT3H	Basmati	DRR	1288/576
31	*Basmati Tall*	BTall	Basmati	DRR	1072/508
32	*Begami T-1*	BegT-1	Aromatic/quality	DRR	1042/1
33	*Balugyun*	Balu	Aromatic/quality	DRR	1186/182
34	*Blomberg*	Blom	Aromatic/quality	DRR	1294/709
35	*Bogi Joha*	Bogi	Aromatic/quality	DRR	1277/364
36	*Boga Joha*	Boga	Aromatic/quality	DRR	1274/361
37	*Bokul Joha*	Bokul	Aromatic/quality	DRR	1273/360
38	BPT1235	BPT	Aromatic/quality	DRR	1264/297
39	BS125	BS125	*Indica*	Cornell	1974
40	*C. Basmati*	C-Bas	Basmati	DRR	1308/400
41	*Cemposelak*	Cempo	Aromatic/quality	DRR	1313/553
42	*Chini Guri*	C.Guri	Aromatic/quality	DRR	1312/431
43	*Chini Sakkar*	C.Sak	Aromatic/quality	DRR	1311/430
44	*Chok-jye-bi-chal 263*	Chok1	Aromatic/quality	DRR	1299/146A
45	*Chok-jye-bi-chal 264*	Chok3	Aromatic/quality	DRR	1301/565
46	*Chu xiang xion*	C-Xion	Aromatic/quality	DRR	1314/733
47	CSR-10	CSR10	*Indica*	CSSRI	CSR10
48	CSR-11	CSR11	*Indica*	CSSRI	CSR11
49	CSR13	CSR13	*Indica*	CSSRI	CSR13
50	CSR18	CSR18	*Indica*	CSSRI	CSR18
51	D66	D66	Aromatic/quality	DRR	1323/301A
52	DM24	DM24	Aromatic/quality	DRR	1327/766B
53	*Dulhaniya*	Dulha	Aromatic/quality	DRR	1318/56
54	*Gobind*	Gobind	*Indica*	CCSHAU	na[c]

Table 1 (continued)

Sample no.	Genotype	Abbreviations	Type	Source[a]	Accession no.
55	HBC-19[b]	HBC19	Premium Basmati	CCSHAU	HBC19
56	HKR-120	HKR120	*Indica*	CCSHAU	HKR120
57	HKR93-401	HKR93	Cross-bred Basmati	CCSHAU	HKR93(401)
58	HKR126	HKR126	*Indica*	CCSHAU	HKR126
59	*Haryana Basmati 1* (HKR228)	HKR228	Cross-bred Basmati	CCSHAU	HKR228
60	IR36	IR36	*Indica*	Cornell	1594
61	IR64	IR64	*Indica*	Cornell	1754
62	*Kasalath*	Kasal	*Indica*	Cornell	1730
63	*Nipponbare*	Nippon	*Japonica*	Cornell	1411
64	New plant type-1	NPT-1	*Japonica*	IRRI	IR65600-38-1-2-1
65	New plant type-2	NPT-2	*Japonica*	IRRI	IR68552-100-1-2-2
66	New plant type-3	NPT-3	*Japonica*	IRRI	IR68011-15-1-1-2-3
67	*Pokkali*	Pokkali	*Indica*	CSSRI	na[c]
68	*Pusa Basmati 1*	P.Bas	Cross-bred Basmati	CCSHAU	na[c]
69	*Taraori Basmati*	Taraori	Premium Basmati	CCSHAU	na[c]

[a]Source abbreviations: DRR, Directorate of Rice Research, Rajendranagar, Hyderabad 500 030, India; CCSHAU, CCS Haryana Agricultural University Rice Research Station, Kaul 132021, India; CSSRI, Central Soil Salinity Research Institute, Karnal 132001, India; Cornell, Department of Plant Breeding, Cornell University, Ithaca, NY 14853-1901, USA

[b]A pure line selection from Taraori Basmati

[c]Certified breeder's seed from CCSHAU, Rice Research Station, Kaul 132021, India

formamide, 10 m*M* EDTA, blue dextran), and 0.1 µl of an internal-lane size standard, TAMRA-labeled Genescan-350 (Applied Biosystems). The samples were denatured at 95°C for 2 min and run for 3 h on 96-well, 5% denaturing Longranger polyacrylamide gels (8.0 *M* urea) in 1x TBE buffer with the recommended run module (constant 30 W) and with filter set B using an ABI-377A automated DNA sequencer (Perkin Elmer Analytical Instruments, Norwalk, Conn.). The experiments were repeated to confirm cases of null alleles or multiple alleles.

Data analysis

Molecular weights for products amplified from microsatellite markers were estimated with GENESCAN 672 software (Applied Biosystems) by the local Southern method. This approach compares the migration of a sample fragment (in base pairs) to the migration of within-lane size standards. Allele calling was done using the GENOTYPER software, and "allele binning" was conducted as described in Blair et al. (2002). The polymorphism information content (PIC) for each microsatellite marker was calculated according to Anderson et al. (1993). Allele molecular weight data were scored in binary format, with the presence of an allele given a score of '1' and its absence scored as '0' to generate similarity matrices using the SIMQUAL subprogram of software NTSYS-PC (Rohlf 1993). The similarity coefficients were used for cluster analysis of the rice genotypes performed using the SAHN subprogram of NTSYS-PC and to build dendrograms by unweighted pair-group method for arithmetic average (UPGMA). Allele molecular weight data were also used to determine the genetic distance for phylogeny construction based on the neighbor-joining method (Saitou and Nei 1987) implemented in POWER MARKER, a genetic analysis software distributed by Kejun Liu (Kliu2@unity.nesu.edu). The inferred ancestry of the rice genotypes was calculated using J. Pritchard's 'structure' analysis (STRUCTURE 2.00) (Pritchard et al. 2000). The posterior probabilities of k (i.e., the likelihood of k as a proportion of the sum of the likelihoods for different values of k) were estimated assuming prior values of k between one and eight. The posterior probabilities were estimated using the Markov chain Monte Carlo method (MCMC). The results presented in this study were based on burnin period length = 10,000; MCMC iterations after burnin = 100,000, where alpha was constant.

Results

Number of alleles per locus

Using 30 microsatellite markers, a total of 235 alleles (electromorphs) were detected among the 69 rice accessions (Table 2 and Supplemental Table 1 at http://ricelab.plbr.cornell.edu/publications/2004/jain/.

The average number of electromorphs per locus was 7.8, with a range of three (RM133, RM323) to as many as 22 (RM252). In agreement with Cho et al. (2000), loci with perfect di-nucleotide repeat motifs tended to detect a greater number of alleles (average: 8.8; n=20) than those with tri-nucleotide repeats (average: 4.5; n=6) (P<0.05, t-test). RM171 (chromosome 10) had a poly(GATG)$_n$ motif and detected an unusually high number of alleles (12) for a tetra-nucleotide SSR.

Table 2 Data on number of alleles, number of rare alleles, number of genotypes with null or multiple alleles, allele size range, highest frequency allele and polymorphism information content (PIC) found among 69 rice genotypes for 30 fluorescently labeled microsatellites from four multiplex panels (A, B, C and D)

Panel/ fluorescent label[a]	Marker[b,c]	Genebank or clone no.	Chromosome no.	Repeat type	No. of alleles	No. of rare alleles	Genotypes with:		Size range	Difference (bp)	Reference allele (bp)[c]			High-frequency alleles		PIC value
							Null alleles	Multiple alleles			Clone	ABI	Difference	Size (bp)	Frequency (%)	
Panel A																
FAM	RM135[b,c]	Osm35	3	$(CGG)_{10}$	4.0	0.0	0.0	6.0	119–131	12.0	125	125	0.0	122	43.5	0.613
	RM280[b,c]	CT780	4	$(GA)_{16}$	9.0	4.0	1.0	6.0	151–181	30.0	156	153	3.0	171	58.0	0.625
TET	RM170[b,c]	Osm68	6	$(CCT)_7$	7.0	3.0	0.0	5.0	99–119	20.0	117	116	1.0	105	58.0	0.611
	RM153	Osm53	5	$(GAA)_9$	4.0	1.0	0.0	2.0	190–205	15.0	195	195	0.0	205	59.4	0.521
	RM182[c]	M1	7	$(AT)_{16}$	4.0	2.0	3.0	1.0	328–349	21.0	330	330	0.0	330	73.9	0.412
HEX	RM286[b,c]	CT806	11	$(GA)_{16}$	9.0	7.0	0.0	4.0	98–126	28.0	112	112	0.0	112	63.7	0.514
	RM110[c]	Osm10	2	$(GA)_{15}$	8.0	4.0	0.0	0.0	138–159	21.0	139	138	1.0	154	55.1	0.646
	RM174[b]	Osm74	2	$(ACG)_7$ $(GA)_{10}$	6.0	3.0	0.0	3.0	207–223	16.0	207	207	0.0	223	49.3	0.636
Mean					6.4	3.0	0.5	3.4		20.4			0.6		57.6	0.572
Panel B																
FAM	RM312	GT165	1	$(ATTT)_4$ $(GT)_9$	6.0	2.0	0.0	7.0	96–106	10.0	96	96	0.0	100	58.0	0.589
	RM105[c]	Osm5	9	$(CCT)_6$	5.0	1.0	0.0	6.0	126–138	12	131	129	2.0	132	60.9	0.545
	RM171[b]	Osm71	10	$(GATG)_5$	12.0	6.0	0.0	9.0	320–350	30	334	327	7.0	336	34.8	0.793
TET	RM133[b,c]	Osm33	6	$(CT)_8$	3.0	1.0	0.0	4.0	230–233	3.0	226	229	3.0	230	85.5	0.247
	RM103[b,c]	Osm3	6	$(GAA)_5$	4.0	0.0	0.0	3.0	327–335	8.0	328	334	6.0	332	55.1	0.602
HEX	RM282[b,c]	CT787	3	$(GA)_{15}$	5.0	2.0	0.0	0.0	125–137	12.0	135	137	2.0	129	56.5	0.583
	RM337[b,c]	CTT64	8	$(CTT)_{4\text{-}19}$- $(CTT)_8$	8.0	5.0	0.0	6.0	157–192	10.0	182	192	10.0	157	71.0	0.454
Mean					6.1	2.4	0.0	5.0		12.1			4.3		60.3	0.545
Panel C																

Table 2 (continued)

Panel/ fluorescent label[a]	Marker[b,c]	Genebank or clone no.	Chromosome no.	Repeat type	No. of alleles	No. of rare alleles	Genotypes with:		Size range	Difference (bp)	Reference allele (bp)[c]			High-frequency alleles		PIC value
							Null alleles	Multiple alleles			Clone	ABI	Difference	Size (bp)	Frequency (%)	
FAM	RM1[b,c]	GA12	1	$(GA)_{26}$	9.0	4.0	0.0	4.0	78–116	39.0	110	110	0.0	78	50.7	0.686
	RM122[b,c]	–	5	$(GA)_{11}$	6.0	3.0	0.0	5.0	228–247	19.0	227	232	5.0	228	79.7	0.339
TET	RM5[b,c]	GA273	1	$(GA)_{15}$	8.0	3.0	0.0	3.0	106–129	23.0	113	113	0.0	116	49.3	0.683
	RM55[b]	GA587	3	$(GA)_{17}$	7.0	3.0	0.0	8.0	219–240	22.0	226	229	3.0	234	55.1	0.594
HEX	RM248[b,c]	CT469	7	$(GA)_{25}$	9.0	4.0	0.0	6.0	72–104	32.0	102	104	2.0	72	58.0	0.624
	RM231[b,c]	CT234	3	$(GA)_{16}$	10.0	7.0	0.0	5.0	168–194	26.0	182	186	4.0	186	65.2	0.473
	RM38	GA344	8	$(GA)_{16}$	10.0	6.0	6.0	1.0	231–263	31.0	250	252	2.0	250	36.2	0.787
Mean					8.4	4.3	0.9	4.6		27.4			2.3		56.3	0.598
Panel D																
FAM	RM253[b]	CT452	6	$(GA)_{25}$	12.0	7.0	2.0	3.0	88–115	27.0	109	111	2.0	106	29.0	0.816
	RM224	CT199	11	$(GA)_{13}$	9.0	4.0	2.0	4.0	121–159	38.0	157	157	0.0	133	50.7	0.679
	RM252[b]	CT206	4	$(GA)_{19}$	22.0	17.0	2.0	9.0	194–262	68.0	216	221	5.0	209, 235	11.6	0.921
TET	RM229[b,c]	CT224	11	$(GA)_{11}$	8.0	5.0	1.0	1.0	104–136	32.0	116	116	0.0	120	53.6	0.627
	RM17[b]	GA56	12	$(GA)_{21}$	7.0	3.0	0.0	7.0	158–187	29.0	180	188	8.0	160	42.0	0.709
	RM222[b]	CT193	10	$(GA)_{18}$	13.0	6.0	0.0	5.0	193–225	32.0	213	214	1.0	206	42.0	0.775
HEX	RM44[b]	GA408	8	$(GA)_{16}$	8.0	4.0	4.0	9.0	103–131	29.0	99	103	4.0	103	39.1	0.688
	RM323[b,c]	CAT69	1	$(CAT)_{5}$	3.0	0.0	0.0	4.0	242–245	3.0	244	245	1.0	242	59.4	0.497
Mean					10.2	5.8	1.4	5.3		32.3			2.6		40.9	0.714
Total Mean					7.8	3.8	0.7	4.5		23.3			2.4		53.5	0.607

Table 2 (continued)

Panel/ fluorescent label[a]	Marker[b, c]	Genebank or clone no.	Chromosome no.	Repeat type	No. of alleles	No. of rare alleles	Genotypes with:		Size range	Difference (bp)	Reference allele (bp)[c]			High-frequency alleles		PIC value
							Null alleles	Multiple alleles			Clone	ABI	Difference	Size (bp)	Frequency (%)	
SSR																

[a]Hex, Hexachloro-6-carboxyfluorescein; Tet, tetrachloro-6-carboxyfluorescein; FAM, 6-carboxyfluorescein
[b]Markers that can differentiate between premium traditional Basmati and one or more of cross-bred Basmati and/or non-Basmati rice varieties
[c]Markers that can detect a particular allele common in more than 90% of the 31 traditional Basmati and aromatic rice varieties. Reference alleles for var. IR36 clone: length based on sequence data; ABI: length based on comparison to size standard

Allele size range

The size variation between the smallest and the largest allele at a given SSR locus was correlated with the number of alleles per locus. Thus, RM133 and RM323 presented the smallest allele size range (3 bp) and had only three alleles per locus, while RM252 had the largest allele size range (68 bp) and a total of 22 alleles (Table 2). The size range of alleles affects the degree of multiplexing that can be accommodated in a single lane on a gel and is therefore an important statistic when selecting markers for use in automated SSR analysis. Markers in panel D had the largest number of alleles (82) and the largest average allele size range (32.3 bp), followed by those in panel C (59 alleles; range of 27.4 bp), panel A (51 alleles; range of 20.4 bp), and panel B (43 alleles; range of 12.1 bp). It is of interest that all of the markers in panels D and C contained only perfect poly$(GA)_n$ motifs, with the exception of one marker in panel D [RM323; a poly$(CAT)_5$], while panels A and B contained a mixture of perfect and compound di-, tri-, and tetra-nucleotide SSR motifs.

PIC values

The PIC values ranged from a low of 0.24 (RM133) to a high of 0.92 (RM252) and averaged 0.61 (Table 2). PIC values also showed a significant, positive linear correlation with the number of alleles (r=0.72) and allele size range (r=0.68) for microsatellites evaluated in this study.

Null alleles

A genotype was assigned a null allele for an SSR locus whenever an amplification product(s) was not detected for the particular genotype × marker combination. Experiments detecting null alleles were all repeated at least once to ensure that the absence of an amplified product was not due to experimental error. Of the 30 SSR loci eight contained null alleles in one to six of the 69 genotypes. The loci harboring the highest frequency of null alleles were RM38 (nulls detected in six genotypes) and RM44 and RM182 (nulls detected in four and three genotypes respectively). The genotypes having the largest proportion of null alleles were NPT-3 (null alleles at four loci) and *Dulhaniya* (null alleles at two loci).

Rare alleles

An allele that was observed in fewer than three of the 69 genotypes (less than 5%) was considered to be a rare allele. Rare alleles were observed at all of the SSR loci in one or more of the 69 accessions, except at RM103, RM135, and RM323 (Table 2). In general, markers detecting a greater number of alleles per locus detected more rare alleles. Marker RM252 detected the greatest number of alleles, and 77% (17/22) of the alleles were

observed in fewer than three genotypes each. Most of the markers derived from genomic libraries [poly(GA)$_n$] detected a higher number of rare alleles than markers derived from expressed sequence tags (ESTs). The exceptions were RM282 [poly(GA)$_{15}$], a marker from a genomic library that detected only two rare alleles, and RM171 [poly(GATG)$_5$], a marker derived from an EST that detected six rare alleles. Rare alleles are highly informative when fingerprinting varieties.

High-frequency alleles

On average, 53.6% of the 69 genotypes shared a common allele at any given locus. Within the group of 12 *indica* varieties, the average frequency of shared alleles was 55%, and within the five *japonica* varieties, the shared allele frequency was 65%. Within the group of 52 Indian aromatic and quality rices, the average frequency of shared alleles was 63%. The number of alleles per locus and the frequency of the most common allele at that locus were negatively correlated (r=0.71) (Table 2). Some alleles were consistently more common in one cluster of germplasm than another. The most common allele at any given locus among the 52 Indian aromatic/quality rices differed from the highest frequency alleles in the *japonica* and the *indica* groups at 24/30 loci, while the most common allele among the five *japonica* accessions differed from the highest frequency allele in the 12 *indica* accessions at 28/30 loci. Based on allele frequency only, the three groups of germplasm could be distinguished from each other at 53% (16/30) of the loci where the most common allele differed for each group.

Genetic similarity among accessions

Genetic similarity values between the various rice accessions were used to produce a dendrogram (cluster tree analysis, Fig. 1) and a two-dimensional scale diagram (PCA analysis, Fig. 2), which clearly explain the relationships among the 69 rice genotypes. Using 80% similarity as the threshold for UPGMA clustering, we observed four major groups (Fig. 1). Group I contained most of the Basmati and other aromatic and quality rice accessions used in this study; group II represented the *indica* accessions, along with some of the cross-bred aromatic rices; group III was populated by a single salt-tolerant variety, *Pokkali*; group IV contained the *japonica* varieties.

Group I consisted of 41 accessions and contained all of the Indian Basmati/aromatic germplasm accessions as well as the landrace *indica* variety, *Kasalath*. The similarity coefficients between any two rice genotypes in this cluster ranged from 79% to 99.6%. The average pair-wise similarity coefficients among the Basmati rices in the group was 0.89, compared to an average coefficient of 0.81 for *Kasalath*. The cluster can be further sub-divided into several sub-groups having varying degrees of

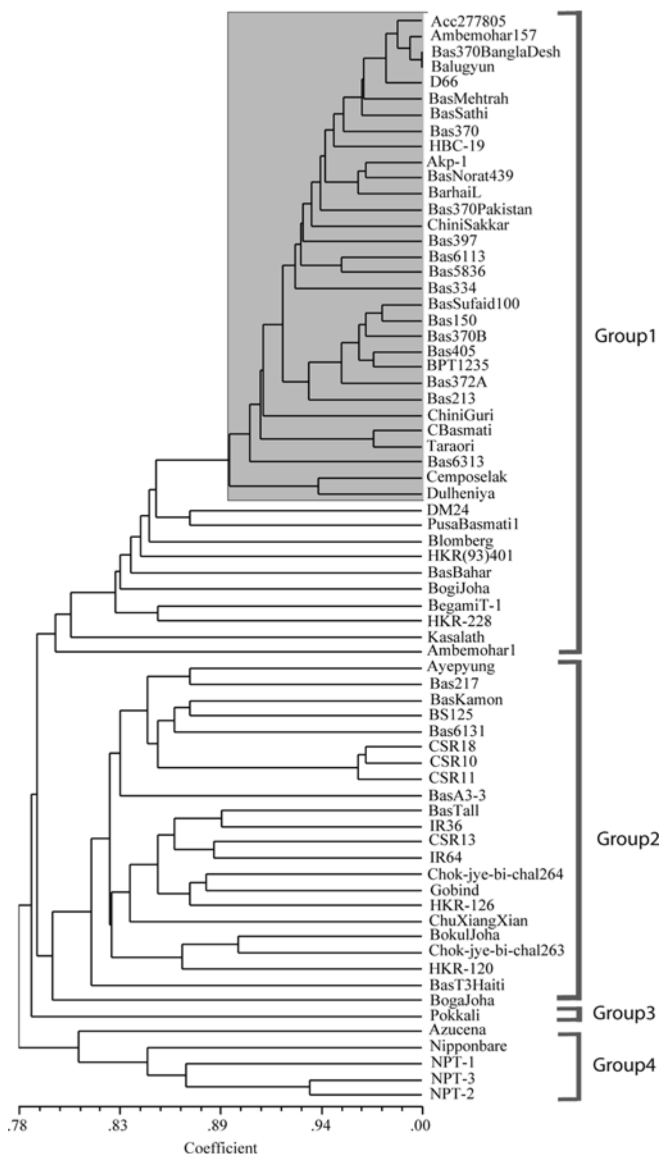

Fig. 1 Dendrogram showing four major groups from (NTSYS-PC) of 69 rice genotypes using genetic diversity data for 235 alleles at 30 SSR loci. The *grey area* indicates a subgroup of 31 Basmati and aromatic rice accessions with a similarity coefficient of 90%

similarity. A major sub-group at about a 90% similarity coefficient is comprised of 31 Basmati rices, including the six commercially important traditional Basmati rice varieties (noted as premium *Basmati* in Table 1). Varieties in this group had an average similarity coefficient of 0.92 and are of interest because some of them command the highest price of any Basmati rices on the international market. The remaining ten rice varieties in group I were more diverse, with an average similarity coefficient of 0.81. These subgroups were situated between the traditional Basmati and the *indica* groups in Figs. 1 and 2 and contained the three commercially popular cross-bred Basmati rice varieties, *Pusa Basmati 1*, HKR228 and HKR93-401 as well as a number of other aromatic and quality rices. This intermediate positioning makes sense based on the mixed parentage represented in the pedigrees

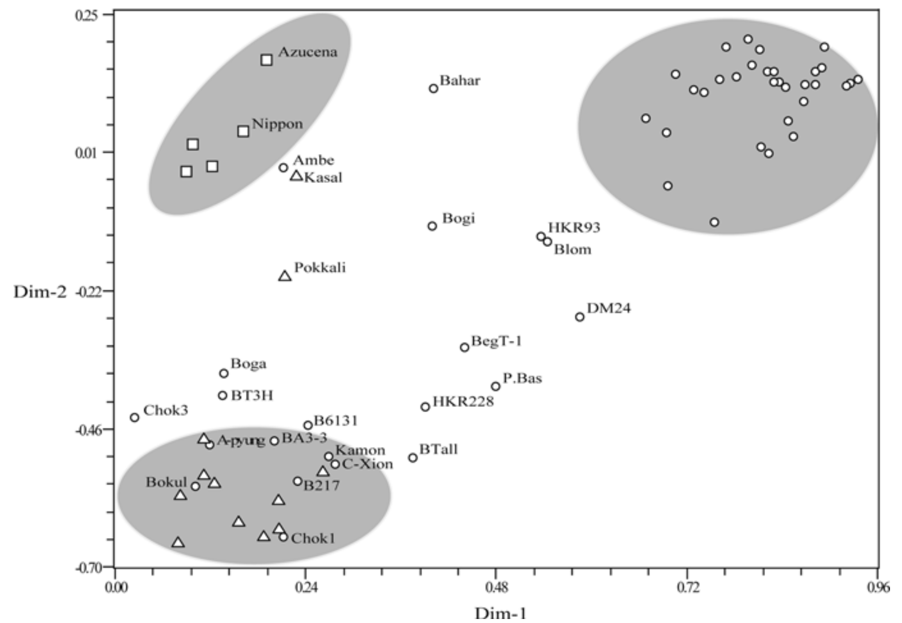

Fig. 2 The two-dimensional scaling resulting from PCA analysis (NTSYS-PC) of 69 genotypes using genetic diversity data for 235 alleles at 30 SSR loci. *Grey areas* represent the three major clusters of germplasm: *open circle* Basmati and aromatic rice, *triangleindica*, *squarejaponica*

of the cross-bred varieties. In each case, there was one *indica* parent (*Pusa 167*, *Sona*, and IR50) and one traditional Basmati parent: *Pusa Basmati 1* was derived from the cross *Pusa167* × *Karnal local Basmati*, *Haryana Basmati 1* (also known as HKR228) was obtained from *Sona* × *Basmati 370*, and HKR93-401 was a selection from the cross IR50 × *Taraori Basmati*.

Group II was more diverse than either groups I or IV and was comprised of both *indica* and aromatic/quality rices. This group included the high-yielding *indica* varieties developed in India and at the IRRI, the salt-tolerant lines released by CSSRI, India, five aromatic/quality rices (*Ayepyung*, *Boga Joha*, *Bokul Joha*, *Chu-xiang-xion*, and *Chok-jye-bi-chal*), six Basmati rice accessions (*Basmati A3-3*, *Basmati T3 Haiti*, *Basmati Kamon*, *Basmati Tall*, *Basmati 217*, *Basmati 6131*) and three improved *indica* rice varieties, (HKR120 and HKR126, and *Gobind*).

Group III consisted only of *Pokkali*, which proved to be the single most unusual genotype in this collection. While generally classified as an *indica*, it demonstrated a highly unusual combination of alleles and did not cluster with genotypes in any of the other three groups in this study.

Group IV, the *japonica* cluster, consisted of only five accessions, and it was interesting to note that the three 'New Plant Type' lines from IRRI clustered more closely with *Nipponbare*, the temperate *japonica* variety from Japan, than with the aromatic, tropical *japonica* variety from the Philippines, *Azucena*.

When taken as a whole, the 69 rice varieties had an average similarity coefficient of 0.83. Rices in groups II and IV had similarity coefficients of 0.79 and 0.78, respectively, while the 41 rices in group I had an average similarity coefficient of 0.89. When rices in the major Basmati sub-group described above were compared with *indica* and *japonica* varieties that contain no known Basmati parents in their pedigrees, similarity coefficients of less than 0.79 were observed. This value is comparable to the similarity coefficient of 0.77 observed between the *indica* (group II) and *japonica* (group IV) rice varieties used in this study. This analysis suggests that Basmati rices have a long and independent history of evolution that distinguishes them from most of the *indica* and *japonica* rices.

The groupings identified by PCA (Fig. 2) were very similar to those identified by the UPGMA tree cluster analysis (Fig. 1). Of the 52 genotypes of Indian aromatic/quality rice genotypes, 31 formed a distinct cluster in the upper-right corner of the diagram, well-separated from the *indica* and *japonica* clusters. This cluster corresponded to the 31 varieties that appear as a major sub-group in Fig. 1. A second sub-group of nine aromatic/quality genotypes included the three cross-bred Basmati's and corresponded well with the group of "intermediate" varieties between Groups I and II in the dendrogram. The remaining 11 aromatic/quality genotypes were interspersed within the *indica* cluster. Two *indica* genotypes (*Kasalath* and *Pokkali*) and one aromatic/quality Basmati genotype (*Ambemohar 1*) were located close to the *japonica* cluster in the PCA, suggesting that gene flow among and between these gene pools has occurred.

The inferred ancestry from a population of 69 genotypes can also be divided into four major subpopulations using Pritchard's structure analysis (Pritchard et al. 2000) (Fig. 3).The natural log (ln) probability of the data, which is proportional to the posterior probability, is maximized at k=4 subpopulations (−2071.9). Genotypes belonging to the major groups shared at least 75% ancestry with other group members. Thirty-one genotypes comprised the largest group, which contained only aromatic and Basmati rice varieties and was identical to the Basmati subgroup identified by the UPGMA and neighbor-joining methods. Out of these 31 rice varieties, 27 showed more than 90% shared ancestry. The second largest group

consisted of 17 rice varieties that included nine *indica* rice varieties and some aromatic and Basmati accessions. Twelve of these, including the nine *indicas*, had more than 90% shared ancestry. The third group comprised seven genetically diverse varieties, including *Kasalath* and *Pokkali*. The varieties in this group harbored many rare alleles but six of the seven varieties had more than 90% shared ancestry. Group IV consisted of the five *japonica* varieties along with *Basmati Bahar*, a variety that clustered with the aromatics using distance-based methods. Ten varieties were identified as admixtures having only 52–67% shared ancestry with a major group (Supplemental Table 2 at http://ricelab.plbr.cornell.edu/publications/2004/jain/.

Markers for distinguishing Basmati rices

Of the 30 SSR markers, 22 (indicated with an "[a]" in the second column of Table 2) were useful in differentiating between premium traditional Basmati rice varieties and cross-bred Basmati or non-Basmati rice varieties, of which the most informative marker was RM252. The application of a hypervariable marker, such as RM252, in combination with other, less variable SSR markers offers a powerful approach to fingerprinting and variety identification and can be used for distinguishing genetically relevant sub-sets of germplasm. Eighteen SSR loci (indicated with a "[b]" in the second column of Table 2) had a particular allele that was common in more than 90% of the 31 traditional Basmati and aromatic rice types but was also sometimes detected in non-Basmati types. From this SSR fingerprint database, a subset of eight SSR markers (RM1, RM5, RM103, RM135, RM171, RM174, RM222, RM252) has been organized into a single panel of fluorescent markers and used to efficiently differentiate among the traditional Basmati, crossbred Basmati, and non-Basmati rice varieties. For example, these markers can distinguish either *Pusa Basmati 1*, *Haryana Basmati* (HKR228) and/or HKR93-401 from the six commercially important premium traditional Basmati rice varieties, including the three *Basmati 370* accessions, *Taraori Basmati* and HBC19.

Heterogeneity

Microsatellite markers in this study detected allele mixtures in 62% (32/52) of the Indian aromatic/Basmati rice accessions. Heterogeneity was detected at one or more loci even in some modern varieties, such as IR36 and *Nipponbare*. For any given marker/locus, heterogeneity was detected in an average of 4.5 genotypes, and hypervariable markers (that detected large numbers of alleles) were most likely to detect the allele mixtures. In most cases of within-accession polymorphism, one of the alleles was rare. For instance, in HKR228, three loci (RM122, RM5 and RM248) showed within-accession polymorphism and, in each case, one of the two alleles occurred in less than 10% of the accessions. Of the Indian

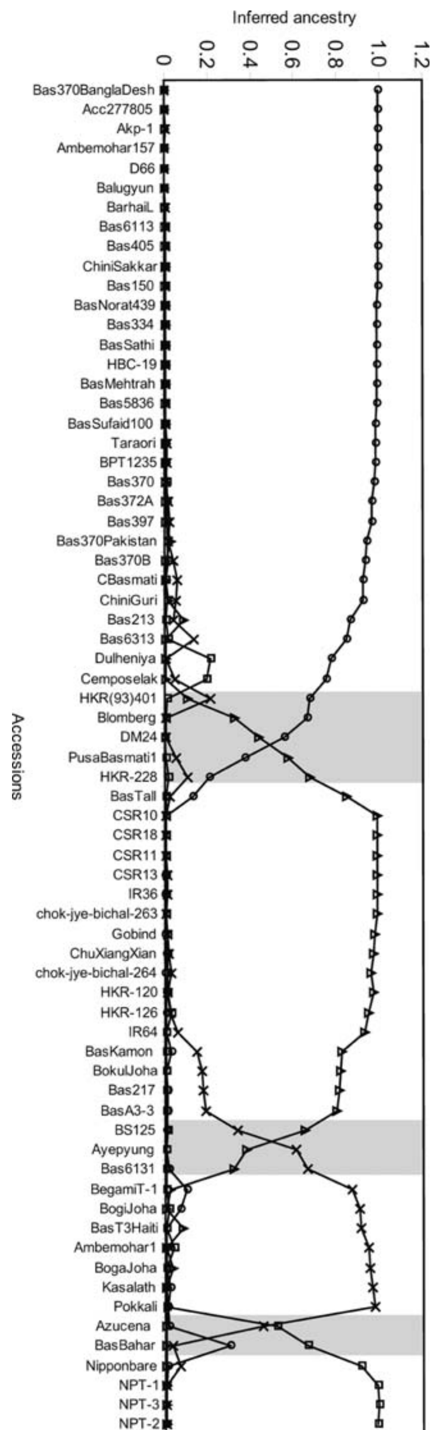

Fig. 3 Graph showing the proportion of shared ancestry among the 69 genotypes based on allele molecular weight data from 30 SSR loci using Pritchard's model-based method structure (Pritchard et al. 2000). *Open circle* Group 1 Basmati, *triangle* group 2 *indica*, *X* group 3 diverse *indica*, *square* group 4 *japonica*. *Grey areas* represent admixtures

aromatic/Basmati rice accessions showing within-accession polymorphism, the most heterogeneous accessions (i.e., those with the largest number of multi-allelic loci)

were HKR93-401, *Blomberg*, *Basmati 334*, DM24, *Boga Joha*, *Chu xiang xion*, *Dulhaniya*, and *Chini sakkar*.

Discussion

The availability of markers that distinguish commercially important traditional Basmati rice cultivars, such as *Taraori Basmati* and *Basmati 370*, from cross-bred varieties, such as *Pusa Basmati 1* and *Haryana Basmati 1*, is of interest to exporters, commercial suppliers, and consumers to ensure varietal purity in the face of large price differentials. Several of the microsatellites used in this study can readily distinguish one Basmati variety from another. The most powerful marker for discriminating among closely related Basmati varieties was RM252 where 13 distinct electromorph alleles were detected in Basmati and other aromatic/quality rices. The hypervariable nature of this marker makes it highly informative but suggests that it is highly mutagenic and potentially unstable. Thus, we suggest that a combination of markers be used together for fingerprinting purposes. In this study, we demonstrated that 22 microsatellite markers were particularly useful for differentiating among types of Basmati varieties (traditional vs. cross-bred) and distinguishing Basmati from non-Basmati aromatic and quality rice varieties. The most valuable markers for this purpose were those that detected Basmati-specific alleles, such as RM1, RM5, RM135, RM170, RM171, RM222, RM252, and RM253 (see Supplemental Table 1). If used in combination with a database containing SSR-based fingerprints of registered varieties, these DNA markers would provide Trading Standards Officers with a reliable method for authenticating labeling claims, determining the authenticity of imported shipments of rice, and assessing the level of mixing of premium Basmati rice with cheaper Basmati or long-grain *indica* rices in the marketplace. Were appropriate sampling techniques to be used, even heterogeneous varieties could be accurately identified and unknown material could be evaluated for genetic relatedness to known varietal standards. Thus, with minimal up-front investment, publicly available, low-cost, and highly efficient SSR fingerprinting technology could be used immediately to help safeguard the interests of both the consumers and producers of traditional Basmati rices.

The use of previously mapped SSR markers for varietal fingerprinting provides insights into which putative chromosomal segments have been introgressed from specific parental genomes, provided that the parental lines are included in the analysis. SSR data such as that generated in this study or that presented by Macaulay et al. (2001) for barley allows researchers to identify regions of the genome that have been retained from each of the specific parents in a cross and to identify regions where the allele constitution has been fixed in breeding material derived from different crosses. This information, in combination with information on *indica*-specific and *japonica*-specific alleles at these and other loci (Coburn et al. 2002; Ni et al. 2001), provides the basis for association mapping studies and offers insights into gene flow among different clusters of germplasm, such as the traditional Basmati, *indica* and *japonica* gene pools.

Genetic heterogeneity is a common feature of rice accessions, despite the highly inbred nature of the species. Two or three alleles per locus were detected in more than one-half of the Indian aromatic/Basmati rice genotypes in this study. There are several genetic hypotheses that can explain the observed polymorphism. One is remnant heterozygosity in some varieties, another is heterogeneity due to inadvertant seed mixtures, and a third is expected varietal heterogeneity where landrace varieties consist of mixtures of pure lines that contribute to their broad adaptation in traditional farming systems. In this study, heterozygosity cannot be clearly distinguished from heterogeneity because our DNA samples were extracted from bulked leaf samples. It is reasonable to assume that most of the Basmati germplasm and other traditional landrace varieties analyzed in this study represent mixtures of pure lines as a feature of the varieties themselves. Some residual heterozygosity can also be expected in such cases where low levels of outcrossing occur among adjacent plants. The level of varietal heterogeneity may vary among the accessions and loci, but it is a predictable feature of older varieties that were not bred as pure lines.

Questions on how to accurately estimate allele frequencies in these heterogeneous germplasm samples warrant further research. This is especially important for rare alleles that occur at a low frequency in heterogeneous accessions. While these rare alleles have a high value as genotype-specific markers, rigorous evaluation is often required to reliably detect them. This generally involves the testing of multiple individuals in an accession, and it is critical to know how many individuals per accession are required to obtain reliable molecular fingerprints. Quantitative information on allele frequencies would help guide efforts to effectively propagate and curate these seed collections as well as assist in performing fingerprinting analysis that accurately represented the allelic constitution of the varieties.

The automated detection of fluorescently labeled microsatellite markers proved to be a powerful and efficient technique for fingerprinting rice accessions and assessing genetic variation among closely related varieties. The use of multiple fluorophores increased the multiplexing potential compared to manual detection using silver staining so that we were able to evaluate an average of eight markers per lane in this diverse collection of germplasm. Automated allele detection using within lane size standards and specialized software also improved the accuracy and efficiency of the analysis by reducing the time required to estimate the molecular weight of each allele. In this study, only five 96-well polyacrylamide gel electrophoresis gels were required to assess polymorphism at 30 SSR loci for all 69 rice genotypes.

Multiplex panels of rice microsatellites providing whole genome coverage have recently been designed to facilitate diversity analysis of rice germplasm (Coburn et al. 2002). These panels have been evaluated to confirm their

reliability and efficiency using a wide range of diverse *O. sativa* accessions. While multiplex panels with up to 24 microsatellites per lane have been used for human genome analysis (Reed et al. 1994), factors such as type of repeat (di-, tri-, or tetra-nucleotide repeats), range of allele sizes per locus, and genetic diversity among genotypes affect the density of multiplex panels that can be designed (Ghosh et al. 1997). With the emergence of genomic sequence information for rice, new SSR markers can be easily designed in specific regions of interest and used to expand existing multiplex panels.

The average number of alleles per DNA locus in this study was 7.8, with a range of 3–22. This is significantly higher than the average of 2.0–5.5 alleles per locus for various classes of microsatellites reported by Cho et al. (2000) and is similar to the average of 7.4 (range 3–17) alleles per locus reported by Olufowote et al. (1997) using different rice germplasm. The presence of numerous "rare alleles" in this collection of Indian germplasm that were not observed in other cultivated *indica* and *japonica* varieties indicated that Indian aromatic and quality rices represent a unique source of genetic diversity among *O. sativa* cultivars.

Some of the landraces, particularly those coming from the presumed center of diversity for aromatic/quality rices, are phenotypically unusual and revealed higher levels of polymorphism and unique loci than others. To give a few examples, *Boga joha*, *Bokul joha*, and *Bogi joha* are all aromatic short-grain traditional varieties from Assam, the eastern state of India. All three landraces were easily distinguishable from each other, as evidenced by their similarity coefficient values (Fig. 1). Despite their short-grain characteristics, *Boga joha* and *Bokul joha* clustered with the *indicas*, while *Bogi joha* found its place in the Basmati group. *Ambemohar 1* is another example of a genetically distinct type. It is a landrace from Maharastra State (India) with aromatic, blackish, small grains. SSR data shows that it contains three unique alleles and 11 rare alleles. *Ambemohar 1* is genetically equidistant from the *japonica*, *indica* and aromatic-Basmati rice groups, with similarity coefficient values varying between 0.78 and 80 (Fig. 1). A total of six aromatic/quality rice varieties (*Ayepyung*, *Boga joha*, *Bokul joha*, *Chok-jye-bi-chal 263*, *Chok-jye-bi-chal 264*, *Chu xiang xion*) and five Basmati rice varieties (*Basmati 217*, *Basmati Kamon*, *Basmati T3 Haiti*, *Basmati A3-3*, *Basmati 6131*) are interspersed between *indicas* in the two-dimensional scaling obtained using PCA and most are also diagnosed as admixtures using Structure analysis (Fig. 3). Most of these aromatic/quality rice varieties do not resemble Basmati rice with respect to their grain or plant characteristics and could be classified as aromatic *indica* varieties. They may have arisen via hybridization between Basmati and *indica* rice varieties.

While molecular fingerprinting can be used to distinguish varieties from one another, it can also suggest the possibility that two varieties with different names may be genetically identical. Aromatic rices, including Basmati, have been cultivated in many countries, including India, Pakistan, Bangladesh, Vietnam, Thailand, Iran, and Japan for more than 1,000 years. This leads to the possibility that the same variety may be cultivated under a different name in different countries. *Balugyun* is an aromatic Basmati cultivar cultivated in Myanmar that shows 100% SSR similarity to *Basmati 370BD*, which is being grown in the neighboring country of Bangladesh (Fig. 1). This may be a case where the same variety is being cultivated under different names in two countries. Though microsatellite analysis at 30 loci cannot confirm that this is indeed the case, it is clear that *Balugyun* and *Basmati 370BD* are at least very recently derived from a common ancestral variety.

In summary, our SSR-based analysis of Indian aromatic and quality rices native to the foothills of the Himalayas suggests a long, independent, and complex pattern of evolution for this cluster of germplasm. Greater appreciation of the genetic diversity contained in the Basmati/aromatic rice gene pool will facilitate the classification, proper maintenance, conservation, and utilization of this unique and valuable resource. In addition, marker-based identification and differentiation of traditional Basmati rices may help serve to maintain the integrity of this high-quality product to the benefit of both farmers and consumers.

Acknowledgements We gratefully acknowledge financial support from the Rockefeller Foundation (RF 99001#736, RF99001#726 and RF2001FS048). We thank Sharon Mitchell for help with the genotyping, Steve Kresovich for his generosity in providing access to the ABI Sequencer at the Institute of Genomic Diversity, and Navinder Saini for help with DNA extraction. We also thank Marappa Maheswaran, Mande Semon, and Amanda Garris for assistance with the data analysis and Lois Swales for help with formatting.

References

Aggarwal RK, Shenoy VV, Ramadevi J, Rajkumar R, Singh L (2002) Molecular characterization of some Indian Basmati and other elite rice genotypes using fluorescent-AFLP. Theor Appl Genet 105:680–690

Ahuja SC, Panwar DVS, Ahuja U, Gupta KR (1995) Basmati rice—the scented pearl. Directorate of Publications, CCS Haryana Agricultural University, Hisar

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186

Bhasin VK (2000) India and the emerging global rice trade. In: Singh RK, Singh US, Khush GS (eds) Aromatic rices. Oxford & IBH, New Delhi, pp 252–276

Blair MW, Hedetale V, McCouch SR (2002) Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). Theor Appl Genet 105:449–457

Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor Appl Genet 95:553–567

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GeneBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Coburn JR, Temnykh SV, Paul EM, McCouch SR (2002) Design and application of microsatellite marker panels for semi-automated genotyping of rice (*Oryza sativa* L.). Crop Sci 42:2092–2099

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

Garris A, McCouch S, Kresovich S (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics 165:759–769

Ghosh S, Karanjawala ZE, Hauser ER, Ally D, Knapp JI, Rayman JB, Musick A, Tannenbaum J, Te C, Shapiro S, Elrridge W, Musick T, Martin C, Smith JR, Carpten, Brownstein MJ, Powell JI, Whiten R, Chines P, Nylund SJ, Magnuson VL, Boehnke M, Collins FS (1997) Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. Genome Res 7:165–178

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 42:233–240

Hamada H, Petrino MG, Kakunaga T (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc Natl Acad Sci USA 79:6465–6469

Khush GS (2000) Taxonomy and origin of rice. In: Singh RK, Singh US, Khush GS (eds) Aromatic rices. Oxford & IBH, New Delhi, pp 5–13

Khush GS, de la Cruz N (1998) Developing Basmati rices with high yield potential. Cahiers Options Mediterraneennes, vol 24. Rice quality. A pluridisciplinary approach. (CD-ROM computer file). CIHEAM, Paris

Kumar S, Shobha Rani N, Krishnaiah K (1996) Problems and prospects of food grain aromatic rice in India. In: IRRI (ed) Rep INGER Monitoring Visit Fine Grain Aromatic Rice India, Iran, Pakistan, and Thailand. IRRI, Manila, pp 21–29

Macaulay M, Ramsay L, Powell W, Waugh R (2001) A representative, highly informative genotyping set of barley SSRs. Theor Appl Genet 102:801–809

McCouch S, Teytelman L, Xu Y, Lobos K, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development of 2,240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Nagaraju J, Kathirvel M, Kumar RR, Siddiq EA, Hasnain SE (2002) Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. Proc Natl Acad Sci USA 99:5836–5841

Nene YL (1998) Basmati rice: a distinct variety (cultivar) of the Indian subcontinent. Asian Agric Hist 2:175–188

Ni J, Colowit PM, Oster JJ, Mackill DJ (2001) Evaluation of genetic diversity in rice sub-species using SSLP markers. Crop Sci 42:601–607

Olufowote JO, Xu Y, Chen X, Park WO, Beachell HM, Dilday RH, Goto M, McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. Genome 40:370–378

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure from multilocus genotype data. Genetics 155:945–959

Reed PW, Davies JL, Copeman JB, Bennett ST, Palmer SM, Pritchard LE, Gough SCL, Kawaguchi Y, Cordell HJ, Balfour KM, Jenkins SC, Powell EE, Vignal A, Todd JA (1994) Chromosome specific microsatellite sets for fluorescence-based, semi-automated genome mapping. Nat Genet 7:390–395

Rohlf FJ (1993) NTSYS-PC: numerical taxonomy and multivariate analysis system, version 1.8, Exeter Software, Setauket, N.Y.

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

Singh RK, Gautam PL, Saxena S, Singh S (2000a) Scented rice germplasm: conservation, evaluation and utilization. In: Singh RK, Singh US, Khush GS (eds) Aromatic rices. Oxford & IBH, New Delhi, pp 107–133

Singh RK, Singh US, Khush GS, Rohilla R, Singh JP, Singh G, Shekhar KS (2000b) Small and medium grained aromatic rices of India. In: Singh RK, Singh US, Khush GS (eds) Aromatic rices. Oxford & IBH, New Delhi, pp 155–177

Smith S (1998) Cultivar identification and varietal protection. In: Caetano-Anollés G, Gresshoff PM (eds) DNA markers: protocols, applications, and overviews. Wiley-VCH, New York, p 283

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice *(Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441–1452

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Westman AL, Kresovich S (1997) Use of molecular marker techniques for description f plant genetic variation. In: Callow JA, Ford-Lloyd BV, Newbury HJ (eds) Plant genetic resources. Conservation and use. CAB Int, Wallingford, pp 9–48