Scientific Research and Essay Vol. 3 (12), pp. 577-598, December, 2008
Available online at http://www.academicjournals.org/SRE
ISSN 1992-2248 © 2008 Academic Journals

Full Length Research Paper

Genetic structure of rice samples from a germplasm bank

Allcochete, A.A.N.[1, 2, 3*], Rangel, P.H.N.[4] and Ferreira, M. E.[3,5]

[1]Dpto Biologia da Universidade Agostinho Neto, Luanda, Angola.
[2] Dpto Biologia Celular da Universidade de Brasília, Brasília, DF, Brasil
[3] EMBRAPA-Recursos Genéticos e Biotecnologia, Laboratório Genética Vegetal, Brasília-DF, Brasil.
[4]EMBRAPA-Arroz e Feijão, Santo Antônio de Goiás, GO, Brasil.
[5]Universidade Católica de Brasília-UCB, Brasília-DF, Brasil.

Accepted 5 December, 2008

The analysis of genetic relationships in crop species is an important component of crop improvement programs, as it serves to provide information about genetic diversity. Accurate assessment of levels and patterns of genetic diversity can be invaluable in crop breeding for diverse applications such as analysis of genetic variability in cultivars, identifying diverse parental combinations to create segregating progenies with maximum genetic variability for further selection, and introgressing desirable genes from diverse germplasm into the available genetic base. In this study, three multiplex panels of fluorescent microsatellite markers were used for automated genotyping of 298 rice accessions, part of it collected in different areas of Brazil and conserved in a genbank. Sixteen marker loci distributed throughout the rice genome were genotyped and the data used to estimate pairwise genetic distances between the accessions. A Neighbour-Joining based dendrogram was used as model to define clusters and infer possible genetic structuring of the collection. The analysis of the genetic relationships of these accessions suggested no significant correlation between clustering based on distance data and subpopulaition differentiation based on MCMC approach. The estimates of Wright's F-statistics revealed a high value of inbreeding coefficient (F_{IS}) and a relatively high overall fixation index (F_{IT}) but only moderate levels genetic differentiation (F_{ST}) of subpopulations defined according to the genetic distance clustering model. It is possible that the use of germplasm conserved in rice germplasm collections could be enhanced if molecular characterization and population genetics approaches could be more intensively applied to better define the sample of accessions which would be more appropriately suited to different genetic and breeding purposes.

Key words: Rice accessions, genetic diversity, microsatellite, multiplex genotyping.

INTRODUCTION

Accurate assessment of levels and patterns of genetic diversity can be invaluable in crop breeding for diverse applications including (i) analysis of genetic variability in cultivars (Smith, 1984; Cox et al., 1986), (ii) identification of diverse parental combinations to create segregating progenies with maximum genetic variability for further selection (Barrett and Kidwell, 1998), and (iii) introgression of desirable genes from diverse germplasm into the available genetic base (Brondani et al., 2002). An understanding of genetic relationships among inbred lines or pure lines can be particularly useful in planning crosses, in assigning lines to specific heterotic groups, in defining parental lines to be used in the development of new recombining populations or to enhance the use of germplasm in breeding programs (Ferreira and Grattapaglia, 1998).

The study of genetic structure of plant populations is fundamental to the understanding of their ecology and the evolutionary forces that affect them (Hartl, 1987). Because genetic structuring reflects the number of alleles exchanged between populations, it has major consequ-

*Corresponding author. E-mail: a_alcochete@yahoo.com, ferreira@cenargen.embrapa.br.

ences on the genetic composition of individuals themselves. Reliable estimates of population differentiation are crucial in conservation biology, where it is often necessary to understand whether populations are genetically isolated from each other. In the case of genetic resources, the knowledge of the genetic structure of accessions deposited in germplasm banks is essential to select those which are suited to different genetic and breeding purposes, thus enhancing their use.

Analyzing the genetic structure of populations or groups of accessions has been based in principles underlying Wright´s F-statistics (Wright, 1978; Weir and Cockerham, 1984). The hierarquichal F-statistics are used to measure the extent of genetic inbreeding within subpopulations (F_{IS}), the extent of genetic differentiation among subpopulations (F_{ST}) and the mean reduction in heterozigosity of an individual relative to the total population (F_{IT}). While this approach has been greatly used in the analysis of natural population, its application in the analysis of germplasm banks has been limited.

The study of genetic diversity of germplasm collections, coupled with genetic differentiation estimates, can facilitate the reliable classification of the accessions, the establishment of their pairwise and group genetic relationships, the selection of representative samples which capture the genetic diversity of the collection and the detection of patterns of differentiation in the whole collection as well as in samples drawn from it (Frankel and Brown, 1984).

Molecular genetic analysis, based on DNA polymerphism detected at marker loci, has been very useful in the analysis of genetic diversity of germplasm collections. Microsatellites or simple sequence repeats (SSRs) (Weber and May, 1989; Litt and Luty, 1989; Tautz, 1989) are very useful co-dominant molecular markers well known for their usually high information content (Ferreira and Grattapaglia, 1998). More than 6,000 microsatellite markers have been developed for rice research, offering great opportunity of application in different genetic studies and on breeding efforts. Micro-satellite marker have been used to explore the rice genome on issues such as genetic mapping of econo-mically important traits (Yamamoto et al., 1999; Brondani et al., 2002; Zhang et al., 2001; Sasaki et al., 2002), population studies (Barbier, 1989b Gao et al., 2000a,b; Gao et al., 2001), germplasm analysis (Glaszmann, 1987; Ni et al., 2002; Parsons et al., 1999; Garris et al., 2003; Gao et al., 2000a,b; Gao et al., 2001) and assessments of the level and structure of genetic diversity in cultivars of interest and on natural populations of *Oryza sativa* and other wild relatives (Olufowote et al., 1997; Yang et al., 1994; Beló, 2001; Gao et al., 2000a,b; Gao et al., 2001; Ni et al., 2002; Garris et al., 2003, 2005). The combined PCR of multiple markers in a sin-gle reaction based on fluorescently labeled microsatellite marker panels greatly increases the capacity of semiautomated genotyping of a large number of accessions, such as those conserved in germ plasm collections. This strategy has been confirmed as fast and highly informative in the characterization of rice genetic resources (Beló, 2001; Pessoa, 2004).

In this study, three panels combining 16 fluorescently labeled microsatellite markers were used for semiautomated genotyping of a germplasm collection and analyzed its genetic structure. Pairwise estimates of genetic distance and genetic diversity parameters were used to establish genetic relationships between accessions. Models of population structure based on genetic relationships were then tested with estimates of F-statistics.

MATERIALS AND METHODS

Plant material and DNA extraction

A sample of 298 accessions of rice (Table 1), mostly composed of landraces collected in the Brazilian territory, were evaluated in this study. This sample also included nine thermosensitive genic male sterility (TGMS) rice lines (TGMS70977_12, TGMS70977_26, TGMS68940, TGMS70987, TGMS68935, TGMS71018, TGMS70989, TGMS68944 and TGMS68945) and several rice varieties commonly used in breeding programs. The *indica* accessions BR-IRGA 417 (ARR1) e BR-IRGA422 (ARR3) were used in all size fragment analysis gels to control the allele sizing variation between each gel electrophoresis run.

About 150 mg of leaves of at least 10 individuals of each accession were used for DNA extraction based on the CTAB protocol (Ferreira and Grattapaglia, 1998). The leaves were dried into 1.5 ml eppendorf tube using liquid nitrogen and ground inside the tubes, followed by addition of 70 ul of the pre-heated extraction buffer (2% acetyldimethyl riethylammonium bromide – CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCL pH=8.0, 1% polyvinylpyrrolidone, 0.2% 2-mercaptoethanol). The solution was incubated for 40 min at 60°C with occasional swirling. Then, 600 uL of chloroform: isoamyl alcohol (24:1) was added to each sample, followed by centrifugation at 12.000 rpm for 10 min. The aqueous phase was transferred to a new tube, and the DNA was precipitated with 0.6 volume of cold isopropanol. A DNA pellet was obtained through centrifugation at 7.500 rpm for 5 min. The supernatant was then discarded and the pellet was washed in two steps using 1 ml of cold 70% ethanol followed by 100% ethanol, and then dried in a vacuum-speed centrifuge (Centrivap Concentrator - LABCONGO) for 10 min. The DNA pellet was then eluted with 5 ul of TE buffer (10 mM Tris-HCl, ph=8.0, 1mM EDTA) containing RNAse (10 $ug.ml^{-1}$) and kept at 37°C for 120 min for RNA digestion. DNA concentration was estimated by 1% agarose gel electrophoresis using 10 ng of λ DNA (GibcoBRL) as a standard. DNA was diluted in TE buffer to a final concentration of 3 $ng.ul^{-1}$.

Genotyping using fluorescent markers

Three multiplex panels consisting of 16 fluorescent-labelled microsatellite loci, developed by Bélo (2000) and Pessoa Filho (2004), were used in this study (Table 2). Simultaneous PCR amplification were carried out in a final volume of 15 ul containing 6 ng of genomic DNA, 0.4 mM of each dNTP, 0.2% $ug.ul^{-1}$ BSA, 3 mM $MgCl_2$, and 2 U Taq DNA Polimerase (Phoneutria-Brazil). For multiplex Panel A, consisting of 5 marker loci, primer concentrations were 0.2 uM (OS19 and RM248) and 0.13 uM (RM252, RM224 and OG44); for multiplex Panel B, consisting of 6 marker loci, primers concentrations were 0.13 uM (OG101, OG05 and OG81), 0.2 uM (OG106), 0.23 uM (OG61) and 0.1 uM (RM263); and for multiplex Panel C, consisting of 5 loci, primer concentration were 0.13 uM (RM335, RM420, RM418, RM259 and RM475). Reactions were

Table 1. Rice accessions of the Rice Germplasm Bank (Embrapa, Brazil) used in the analysis.

#	GEL #	Germplasm bank #	Common name	Collection year	Source
1	AC1480	CA880070	3 MESES BRANCO	-	CNPAF
2	571_AL	CA960010	90 DIAS	2000	AL
3	298_RO	CA830032	A.BOLINHA/CATETINHO	1984	RO
4	A1729B	-	A1729B	-	CNPAF
5	A1744B	-	A1744B	-	CNPAF
6	A1802C	-	A1802C	-	-
7	A1809B	-	A1809B	-	-
8	A1860B	-	A1860B	-	-
9	AC1802b	-	AC1802b	-	-
10	ACC5	CA780002	AGULHA (1)	-	CNPAF
11	ACC310	CA790064	AGULHA (2)	-	CNPAF
12	ACC516	CA790300	AGULHA (2A)	-	CNPAF
13	AC1012	CA830126	AGULHA (3)	-	CNPAF
14	538_MT	CA880075	AGULHA DA TERRA	1989	MT
15	107_CE	CA780283	AGULHA DOURADO	1983	CE
16	ACC222	CA780342	AGULHINHA	-	CNPAF
17	ACC543	CA790326	AGULHINHA	-	CNPAF
18	AC878	CA820073	AGULHINHA (6)	-	CNPAF
19	AC235	CA780366	AGULHINHA (8)	-	CNPAF
20	482_MG	CA870120	AGULHINHA BRANCO	2000	MG
21	AC1257	CA860048	AGULHINHA-DO-BREJO	-	CNPAF
22	AC1928	BAG 94	AKITAKOMACHI	-	EPAGRI
23	AC1193	CA850050	AMARELAO	-	CNPAF
24	AC205	CA780313	AMARELAO PRECOCE	-	CNPAF
25	573_AL	CA960017	AMARELINHO	2000	AL
26	576_AL	CA960029	ANÃO	2000	AL
27	AC202	CA780313	ANAO (5)	-	CNPAF
28	ACC105	CA780134	ANAO (8)	-	CNPAF
29	AC189	CA780285	ANAO-DO-FIM	-	CNPAF
30	577_AL	CA960030	ANAOZINHO	2000	AL
31	-	-	ARR1	-	CNPAF
32	-	-	ARR3	-	CNPAF
33	AC1420	CA870158	ARROZ 4 MESES	-	CNPAF
34	AC1426	CA870168	ARROZ 51	-	CNPAF
35	AC1439	CA870191	ARROZ 9 ANOS	-	CNPAF
36	AC913	CA830002	ARROZ AMARELO	-	CNPAF
37	294_SC	CA830003	ARROZ AMARELO E BRANCO	1984	SC
38	549_MG	CA890001	ARROZ ANÃO	2000	MG
39	277_ES	CA810039	ARROZ BARRIGA BRANCA	1997	ES
40	491_GO	CA870153	ARROZ DA TERRA	1999	GO
41	51_MA	CA780157	ARROZ DE DEUS OU ARROZ JUDIANO	1983	MA

Table 1. Continued

42	377_MS	CA850023	ARROZ DE MAIO	2000	MS
43	CC802A	CA810060	ARROZ MINEIRO	-	CNPAF
44	433_GO	CA870007	ARROZ PELUDO	1996	GO
45	689_GO	CNA0005564	ARROZ PIAUI	2000	GO
46	AC1809	GEN 1239	ARROZ PRETO	-	CNPAF
47	AC1218	CA860003	ARROZ PRETO (4)	-	CNPAF
48	493_MG	CA870162	ARROZ TRÊS MESES	1999	GO
49	AC1971	PB08	ARROZ VERMELHO	-	CPAMN
50	AC1964	PB01	ARROZ VERMELHO (1)	-	CPAMN
51	AC1965	PB02	ARROZ VERMELHO (2)	-	CPAMN
52	AC1969	PB06	ARROZ VERMELHO (3)	-	CPAMN
53	AC263	CA790003	ARROZ VERMELHO (5)	-	CNPAF
54	AC1967	PB04	ARROZ VERMELHO (7)	-	CPAMN
55	AC1974	PB11	ARROZ VERMELHO (8)	-	CPAMN
56	AC1975	PI01	ARROZ VERMELHO (9)	-	CPAMN
57	AC1968	PB05	ARROZVERMELHO (6)	-	CPAMN
58	AC1976	PE01	ARROZVERMELHO4	-	CPAMN
59	ACC338	CA790098	BACABA	-	CNPAF
60	257_MG	CA800128	BARRIGA BRANCA	1994	MG
61	A1722a	-	BASMATI 370	-	CNPAF
62	BG902	-	BG90-2	-	CNPAF
63	AC1409	CA870142	BICO GANGA	-	CNPAF
64	AC1811	RG033	BILLCLINTON	-	IRGA
65	AC1813	RG037	BLUE BONNET 50	-	IRGA
66	AC1812	RG036	BLUEBELLE	-	IRGA
67	AC1813	RG037	BLUEBONNET 502	-	IRGA
68	46_MA	CA780148	BRANCO PRECOCE	1983	MA
69	AC1345	CA870040	BRANCO-DE-BREJO	-	CNPAF
70	ACC250	CA780391	BRANQUINHO	-	CNPAF
71	ACC17	CA780017	BRANQUINHO 90 DIAS	-	CNPAF
72	662_CLB	CNA0004482	BRASILEIRO	1983	CLB
73	BRIRGA	-	BR-IRGA409	-	CNPAF
74	AC1814	RG047	BRS-BOJURU	-	IRGA
75	ACC403	CA790168	BURITI	-	CNPAF
76	62_RS	CA780180	CACHINBO DO SECO	1983	RS
77	AC1816	RG053	CACHINHO	-	IRGA
78	59_RS	CA780171	CACHINHO	1983	RS
79	625_BZL	CNA0000982	CALORO	2000	BZL
80	AC1817	RG055	CALORO 7985	-	IRGA
81	615_RS	CNA0000777	CALOURO	2000	RS
82	574_AL	CA960020	CANARINHO	2000	AL
83	AC1436	CA870186	CANAROXA	-	CNPAF
84	318_RO	CA830125	CANELA CURTA / PINRABNAO	1984	RO

Table 1. Continued

85	ACAC73	CA780091	CANELA-DE-ACO	-	CNPAF
86	ACC78	CA780099	CANELA-DE-FERRO	-	CNPAF
87	ACC577	CA790376	CANELA-DE-FERRO (6)	-	CNPAF
88	670_CLB	CNA0004573	CANILLA	1983	CLB
89	673_CLB	CNA0004600	CANUTO	1983	CLB
90	AC1818	RG057	CAPI-93	-	IRGA
91	AC48	CA780063	CAQUI	-	CNPAF
92	554_PB	CA910002	CAQUI	1996	PB
93	AC1819	RG059	CARNAROLI	-	IRGA
94	AC208	CA780319	CAROLINA (1)	-	CNPAF
95	AC462	CA790237	CAROLINA (2)	-	CNPAF
96	569_AL	CA960006	CAROLINA BRANCO	2000	AL
97	695_GO	CNA0006548	CATALAO 101	1994	GO
98	44_MA	CA780139	CATARINA	1983	MA
99	ACC449	CA790222	CATETO	-	CNPAF
100	575_AL	CA960027	CHATINHO	2000	AL
101	668_CLB	CNA0004546	CHATO RAYADO	1983	CLB
102	AC1168	CA850015	CHIFRE-DE-VEADO	-	CNPAF
103	666_CLB	CNA0004503	CHILENO	1983	CLB
104	ACC87	CA780109	CHILILICA	-	CNPAF
105	AC1820	RG061	CHINSEIASAHI	-	IRGA
106	AC1806	GEN 1236	CHIQUINHO	-	CNPAF
107	CICA 8	-	CICA8	-	CNPAF
108	AC1822	RG068	CINIA 900	-	IRGA
109	AC1823	RG069	CINIA 937	-	IRGA
110	AC1734	-	CIWINI	-	CNPAF
111	AC719	CA800146	COME-CRU	-	CNPAF
112	ACC295	CA790044	COMECRU (13)	-	CNPAF
113	AC55	CA780071	COMUM	-	CNPAF
114	AC67	CA780084	CUTIAO BRANCO	-	CNPAF
115	ACC307	CA790061	CUTIAO VERMELHO	-	CNPAF
116	AC79	CA780101	DEABRIL (1)	-	CNPAF
117	AC627	CA800028	DEABRIL (2)	-	CNPAF
118	AC1824	RG225	DELLA	-	IRGA
119	AC1424	CA870166	DESCONHECIDO	-	CNPAF
120	AC1040	CA840032	DESCONHECIDO (4)	-	CNPAF
121	AC1890	RG640	DESCONHECIDO12 (OKAMINORI 2)	-	IRGA
122	AC1825	RG226	DIAMANTE	-	IRGA
123	AC1729	-	Diamante	-	CNPAF
124	AC1826	RG227	DIAMANTE 1078	-	IRGA
125	AC1827	RG229	DOBLE CAROLINA	-	IRGA
126	270_ES	CA810016	DOIDÃO	1998	ES
127	669_CLB	CNA0004560	DONATO	1983	CLB

Table 1. Continued.

128	AC59	CA780075	DOURADINHO	-	CNPAF
129	AC1828	RG230	DREW	-	IRGA
130	AC1829	RG231	DULAR 22169-76	-	IRGA
131	AC1830	RG232	DULAR CNA1010	-	IRGA
132	AC1536	CA950029	EC1	-	CNPAF
133	AC1831	RG233	EEA201	-	IRGA
134	AC1832	RG234	EEA301	-	IRGA
135	AC1833	RG235	EEA304	-	IRGA
136	AC1834	RG236	EEA401LC	-	IRGA
137	AC1835	RG237	EEA404	-	IRGA
138	AC1836	RG238	EEA405	-	IRGA
139	AC1837	RG239	EEA406	-	IRGA
140	AC1838	RG240	EEA406 (MBL)	-	IRGA
141	AC1839	RG243	ELONI	-	IRGA
142	A1580a	CA980002	EMGOPINHA	-	CNPAF
143	21_AL	CA780055	ESAV-36	1994	AL
144	AC1555	CA960024	ESAVE	-	CNPAF
145	AC1840	RG249	EUROSE	-	IRGA
146	F1	-	F1	-	CNPAF
147	AC1841	RG250	FANNY21501-24	-	IRGA
148	AC1842	RG251	FANNY22193-	-	IRGA
149	AC1512	CA940007	FARROUPILHA	-	CNPAF
150	AC1843	RG253	FARROUPILHA10	-	IRGA
151	AC1379	CA870097	FERRAO PRETO	-	CNPAF
152	AC1370	CA870086	FERRAO PRETO (3M)	-	CNPAF
153	AC1330	CA870021	FERRUJAO	-	CNPAF
154	AC1844	RG270	FRANCES	-	IRGA
155	ACC130	CA780176	FRONTEIRA	-	CNPAF
156	AC1845	RG271	FUKU60	-	IRGA
157	AC1846	RG272	FUKUBOZU	-	IRGA
158	99_CE	CA780266	GAVIAO	1983	CE
159	AC1318	CA860138	GUAIRA	-	CNPAF
160	AC1116	CA840131	GUAIRA (4)	-	CNPAF
161	AC1343	CA870042	GUAIRA BRANCO	-	CNPAF
162	AC21	CA780022	GUAPINHA	-	CNPAF
163	AC1847	RG280	GULFMONT	-	IRGA
164	AC1847	RG280	GULFMONT (2)	-	IRGA
165	ACC41	-	HIBRIDO	-	-
166	268_ES	CA810003	HÍBRIDO	1997	ES
167	AC616	CA800013	HIBRIDO (6)	-	CNPAF
168	AC1850	RG287	HINOHIKARI	-	IRGA
169	AC665	CA800071	HONDURAS	-	CNPAF
170	AC1738	-	HUAN-SEM-GO	-	CNPAF

Table 1. Continued.

171	ACC355	CA790117	IAC	-	CNPAF
172	AC1769	-	IAC 202	-	CNPAF
173	AC1768	-	IAC1246	-	CNPAF
174	AC1768	-	IAC165	-	CNPAF
175	AC1851	RG291	IACUBA18	-	IRGA
176	AC1852	RG292	IACUBA19	-	IRGA
177	AC1853	RG293	IACUBA20	-	IRGA
178	AC1855	RG295	ICTAQUIRIGUA	-	IRGA
179	AC31	CA780036	IGUAPE	-	CNPAF
180	AC1849	RG296	ILLABONG	-	IRGA
181	AC1859	RG300	INIA P10	-	IRGA
182	AC1860	RG301	INIA P11	-	IRGA
183	AC1856	RG297	INIACUARÓ	-	IRGA
184	AC1857	RG298	INIATACUARI	-	IRGA
185	AC1723	-	IR36	-	CNPAF
186	IR50	-	IR50	-	CNPAF
187	163_RS	CA780409	ITAQUI	1994	RS
188	AC1862	RG588	JACKSON	-	IRGA
189	274_ES	CA810030	JAGUARÃO OU TAQUARÃO	2000	ES
190	AC196	CA780296	JAPONES	-	CNPAF
191	ACC126	CA780169	JAPONES CLARO	-	CNPAF
192	AC700	CA800122	JAPONES-DA-VARZEA	-	CNPAF
193	578_AL	CA960033	JAQUARI	2000	AL
194	685_GO	CNA0004799	JARAGUA	1983	GO
195	AC1865	RG592	JEFFERSON	-	IRGA
196	AC1805	GEN 1235	JOÃO PACU	-	CNPAF
197	AC1866	RG593	JODON	-	IRGA
198	AC1867	RG594	JUMA 62	-	IRGA
199	AC1867	RG594	JUMA624	-	IRGA
200	AC1858	RG595	K39-96-1-1-1-2	-	IRGA
201	AC1868	RG596	KAMEJI	-	IRGA
202	AC1869	RG597	KANTO 51	-	IRGA
203	AC1872	RG601	KULINAR(GM)	-	IRGA
204	ACC285	CA790031	L.CORADIM (1484)	-	CNPAF
205	ACC350	CA790110	L.CORADIN (1560)	-	CNPAF
206	ACC553	CA790332	LAGEADO (1)	-	CNPAF
207	ACC850	CA820041	LAGEADO (2)	-	CNPAF
208	AC1038	CA840029	LAGEADO (3)	-	CNPAF
209	AC1135	CA840158	LAGEADO (4)	-	CNPAF
210	AC251	CA780392	LAGEADO (51)	-	CNPAF
211	AC366	CA790128	LAGEADO (62)	-	CNPAF
212	334_PI	CA840042	LAGEADO LIGÍTIMO	1996	PI
213	AC1876	RG609	LAGRUE	-	IRGA

Table 1. Continued.

214	AC1878	RG612	LEBONNET	-	IRGA
215	AC1484	CA880078	LEVANTA TESTA	-	CNPAF
216	345_PI	CA840069	LIGEIRINHO DESCONHECIDO	1985	PI
217	674_CLB	CNA0004601	LIGERITO	1983	CLB
218	324_PI	CA840022	MACABA MIUDO	1985	PI
219	ACC614	CA800011	MACARICO	-	CNPAF
220	ACC86	CA780108	MAIO	-	CNPAF
221	ACC210	CA780321	MANGOTE	-	CNPAF
222	AC63	CA780079	MARANHAO	-	CNPAF
223	234_MG	CA800041	MARELINHO/PAGA DÍVIDA	1994	MG
224	METI1B	-	METICA	-	CNPAF
225	AC1884	RG625	MIARA	-	IRGA
226	AC1743	-	MINAMIHATAMOCHI	-	CNPAF
227	579_AL	CA960036	MINEIRO	2000	AL
228	683_FLP	CNA0004763	MIRITI	1983	FLP
229	AC1453	CA880025	MISTURA	-	CNPAF
230	AC1929	BAG 98	MOCHIGOME	-	EPAGRI
231	ACC253	CA780396	MOGI	-	CNPAF
232	AC1885	RG626	MOLO	-	IRGA
233	AC1009	CA830122	MONTANHA	-	CNPAF
234	110_CE	CA780288	MOURIN AGULHA	1983	CE
235	55_MA	CA780166	MULATINHO	1983	MA
236	235_MG	CA800049	MUNDICERA	2000	MG
237	255_MG	CA800214	NANICO	1994	MG
238	AC1886	RG633	NIPPONBARE	-	IRGA
239	AC1742	-	NOURINMOCHI	-	CNPAF
240	AC1888	RG638	NP125	-	IRGA
241	AC1889	RG639	OBANASAWA (6)	-	IRGA
242	AC197	CA780297	OITENTAO	-	CNPAF
243	AC1891	RG641	OKAMINORI (3)	-	IRGA
244	AC1892	RG642	OOTORI	-	IRGA
245	AC1752	-	Oryzica1	-	CNPAF
246	AC1893	RG650	OU188	-	IRGA
247	AC628	CA800029	PAGA DIVIDA (L.HIPOTECA)	-	CNPAF
248	AC427	CA790196	PALHA MURCHA	-	CNPAF
249	AC626	CA800027	PANCHOLINA	-	CNPAF
250	AC1506	CA930003	PARAZINHO PALHA	-	CNPAF
251	AC879	CA820074	PAULISTA	-	CNPAF
252	47_MA	CA780143	PIRRACA	1983	MA
253	AC1898	RG669	PUNTAL(ÉFFEM)	-	IRGA
254	AC1902	RG678	QUILLA166002	-	IRGA
255	680_CLB	CNA0004637	RABO DE YEGUA	1983	CLB
256	AC372	CA790134	RABO-DE-BURRO	-	CNPAF

Table 1. Continued.

257	ACC83	CA780105	RECHORO	-	CNPAF
258	ACC129	CA780175	REETZ	-	CNPAF
259	AC1137	CA840160	REXORO	-	CNPAF
260	AC1905	RG682	RIKUTONORIN (11)	-	IRGA
261	AC1906	RG683	RIKUTOSHINRIKI	-	IRGA
262	AC1923	BAG 54	ROXO	-	EPAGRI
263	661_FLP	CNA0004319	SALUMPIKIT	1983	FLP
264	ACC99	CA780125	SANTA CATARINA	-	CNPAF
265	228_MG	CA800001	SANTA CATARINA	2000	MG
266	AC1907	RG691	SAN-YANG-AI (1)	-	IRGA
267	158_BA	CA780397	SAQUAREMA	1984	BA
268	AC1034	CA840018	SAQUAREMA (6)	-	CNPAF
269	AC1810	GEN 1240	SEUANTONIO	-	CNPAF
270	AC1910	RG700	SINALOAA80	-	IRGA
271	AC1912	RG717	TA-POO-CHO-Z	-	IRGA
272	40_MA	CA780128	TAQUARAO	1983	MA
273	AC1913	RG718	TASHIMINORI	-	IRGA
274	AC1916	RG721	TEXMONT	-	IRGA
275	-	-	TGMS68935	-	CNPAF
276	-	-	TGMS68940	-	CNPAF
277	-	-	TGMS68944	-	CNPAF
278	TGMS68945	-	TGMS68945	-	CNPAF
279	-	-	TGMS70977_12	-	CNPAF
280	-	-	TGMS70977_26	-	CNPAF
281	-	-	TGMS70987	-	CNPAF
282	-	-	TGMS70989	-	CNPAF
283	-	-	TGMS71018	-	CNPAF
284	AC1808	GEN 1238	TIRIRICA (11)	-	CNPAF
285	ACC81	CA780103	TIRIRICA (13)	-	CNPAF
286	AC1744	-	TOMOEMOCHI	-	CNPAF
287	AC54	CA780070	TORO	-	CNPAF
288	24_PB	CA780061	TRES POTES	1983	PB
289	43_MA	CA780136	VASSOURINHA	1983	MA
290	236_MG	CA800050	VENES BRANCO	1994	MG
291	39_MA	CA780127	VENEZ ROXO	1983	MA
292	AC1566	CA960041	VERMELHAÇÃO	-	CNPAF
293	AC1580	CA980001	VERMELHO	-	CNPAF
294	AC1918	RG730	XIANGJING834	-	IRGA
295	ACC365	CA790127	ZEBU	-	CNPAF
296	AC1919	RG734	ZENITH	-	IRGA
297	AC1861	RG736	ZHAOTONMAXAIGU	-	IRGA
298	AC1920	RG737	ZHEN-GUI-AI-1	-	IRGA

performed on a GeneAmp PCR system 9600 (Perkin-Elmer) using the following profile: (i) a hot start of 94°C for 5 min, (ii) 30 amplification cycles consisting of 1 min at 94°C, 1 min at 52°C, 2 min at 72°C, and (iii) a final extension step of 7 min at 72°C. An aliquot of 5 ul of the amplification product was combined with 3 ul of loading buffer (98% formamide, 10 mM EDTA-blue dextran) and 2 ul of an internal size standard (Brondani and Grattapaglia, 2001), followed by denaturation at 95°C for 5 min. A total of 1.5 ul of the sample was loaded on 4% Long Ranger polyacrylamide gels in 1X TBE buffer for 2.5 h with constant 30 W power on an ABI Prism 377 automatic DNA sequencer (Applied Biosystems).

Data analysis

Microsatellite fragment sizing was performed using the GeneScan software version 3.1.2. (Applied Biosystems). The size of amplified fragments was assigned using the Genotyper software version 2.5.2 (Applied Biosystems) based on the internal standard fragment sizes. Allele binning was performed by rounding of the Genotypes assigned values to the nearest base-pairs integer to give a base pair estimate for the allele. Because most of the loci used in this study harboured dinucleotide motifs, the binning process sometimes resulted in intermediate values for the assigned alleles. A correction was performed so that all values would follow the expected size for dinucleotide motif loci. For this purpose, the most frequent fragment allele was considered as a reference for the expected values of the other alleles at the marker locus.

Genetic diversity analysis

The allele frequencies, total number of the observed alleles, allele size range per locus, the number of different alleles for each locus, the observed gene diversity (GDo) and expected gene diversity (GDe) values were performed using the Cervus software version 2.0 (Marshall et al., 1998). The expected gene diversity measurements were performed following Nei's unbiased estimator which is a reflection of allele diversity and their frequencies, with a correction for the number of sampled individuals, given by $[GDe]=(1-\Sigma_i p_i^2)(2n)/(2n-1)$, where p_i the frequency of the *ieth* allele for each locus and n the number of analyzed samples (Nei, 1987). The Polymorphism Information Content (PIC) was also estimated as the probability that an individual is informative with respect to the segregation of its inherited alleles (Botstein et al., 1980). The probability of identical genotypes, defined as $PI= \Sigma\ pi^4 + \Sigma\ (2pipj)^2$ was estimated for the selected loci individually, and later, for all 16 loci as a whole (Sefc et al., 1999).

Genetic distances among rice accessions based on microsatellite marker polymorphism were estimated by shared allele distance in pairwise comparisons. The estimates were based on the sum of the proportion of common alleles between two rice accessions examined across loci divided by twice the number of tested loci (Bowcock et al., 1994; Goldstein et al., 1995). The genetic distance diagonal matrix was calculated using the parameter [(-ln(Ps)] of the web-based Genetic Distance Calculator for shared allele distance (http://www2.biology.ualberta.ca/jbrzusto/sharedst.php). The diagonal matrix was then submitted to cluster analysis using the Neighbour-Joining Method (Saitou and Nei, 1987) performed by the NTSYSpc software version 2.10 (Rohlf, 1998). Two- and three-dimensional plots were obtained using Dcenter and Eigen procedures in NTSYSpc (Rohlf, 1998) to better visualize the accessions distribution.

Germplasm collection genetic structure analysis

In order to analyze the genetic structure of the germplasm collection, the main clusters of accessions defined by the genetic distance analysis were evaluated as subpopulations. The estimates were performed using the bootstrapping methodology of GENETIX 4.05 software (Belkhir, 2001). The following statistics were estimated (Weir and Cockerham, 1984): the inbreeding coefficient (F_{IS}) was calculated by $F_{IS}=(H_S-H_I)/H_S$ and it was used to estimate the mean reduction in heterozygosity of an individual due to non-random mating within a subpopulation. The fixation index was calculated by $F_{ST}=(H_T—H_S)/H_T$ and it was used to estimate the extent of genetic differentiation among subpopulations (clusters). The overall fixation index (F_{IT}) was calculated by $F_{IT}=(H_T-H_I)/H_T$ and it was used to estimate the mean reduction of heterozygosity of an individual relative to the total population (germplasm collection). The accessions of the collection were subdivided according to the results of the Neighbour Joining analysis: (a) two main clusters and (b) nine clusters.

The genetic structure of the germplasm collection was also analyzed according to a contrast between an *a priori* model of population structure based on the clusters defined by the genetic distance analysis and an unknown *a priori* model using the software STRUCTURE version 2.1 (Pritchard et al., 2004). The program allows for the determination of the coefficient of ancestry (Q) and genetic divergence (F_{ST}) through the Markov Chain Monte Carlo (MCMC) method. It is a model based program which allows the comparison of models of population structure determined by the user based on previously available data (ex. cluster analysis) with models that include an unknown number of subpopulations. The structure analysis was set up for admixture ancestry model, correlated allele frequency, and 20.000, 50.000 and 2000.00 MCMC repeats after equal period of burning (Pritchard et al., 2004).

RESULTS AND DISCUSSION

Microsatellite multiplex genotyping

The automated genotyping using the three multiplex panels of fluorescent microsatellite markers (Figure 1) yielded 305 unique alleles (Table 3). The average allele number per locus was 19.06, ranging from 7 for marker RM420 to 30 for marker RM250. The most frequent allele (≥ 40 counts) represented, on average, 54.47% of all alleles in each marker locus and it ranged from 22.27% in locus OG106 to 93.72% in locus RM420.

Almost all loci presented allele size ranges within the expected, according to data collected in previous studies (Beló, 2001; Pessoa, 2004). The exceptions were loci OG61 and OG81, dinucleotide motifs, for which the allele size ranges extended for 64 and 32 bp against the expected 58 and 18 bp, respectively. The observed allele size ranges for loci RM335 and RM418, trinucleotide markers, were also different from the expected (66 instead of 73 bp expected and 63 instead of 52 bp expected, respectively). The observed range of microsatellite alleles on a per locus basis was larger than those reported in previous studies using other types of markers such as isozymes (Glaszmann, 1987; Second, 1982), RFLPs (Wang and Tanskley, 1989; Zhang et al., 1992) and microsatellites (Beló, 2001). However, 6 markers out of 16 loci with size range larger than that expected (Beló, 2001) were lower than the figures reported by Pessoa Filho (2004). This could have been caused by two factors: (a) a much higher genotyping

Table 2. Fluorescent-labelled microsatellite markers used in three multiplex panels.

Panel	Loci	Fluorescent dye	Color	Expected size range*	Chromosome	Motif
A	RM252	TET	Green	194-274	4q	(ag)n
A	RM224	TET	Green	124-162	11q	(aag)8(ag)13g(ag)
A	OG44	6-FAM	Blue	152-172	3q	(aag)8(ag)22(gt)(gc)6
A	RM248	6-FAM	Blue	72-108	7q	(ag)n
A	OS19	HEX	Yellow	174-198	6p	(ag)n
B	OG101	TET	Green	95-161	2p	(ag)n
B	OG106	TET	Green	178-254	9p	(ag)n
B	OG81	6-FAM	Blue	71-89	1	(ag)n
B	RM263	6-FAM	Blue	147-195	2q	(ag)n
B	OG05	HEX	Yellow	170-194	3p	(ag)n
B	OG16	HEX	Yellow	96-154	5q	(ag)n
C	RM335	6-FAM	Blue	104-156	4p	(ctt)25
C	RM420	6-FAM	Blue	183-201	7q	(aaat)7
C	RM418	6-FAM	Blue	253-311	7q	(att)21
C	RM259	NED	Yellow	155-176	1p	(ct)17
C	RM475	NED	Yellow	345-387	3q	(attg)30

*Pessoa Filho (2004).

accuracy in allele determination using fluorescently based genotyping in DNA sequencers compared to polyacrilamide gel electrophoresis (PAGE) stained with silver nitrate, as described by Beló (2001) and (b) an effect of sample size in the study described here and that developed by Pessoa Filho (2004).

Microsatellites have been ideal markers for characterrizing genetic diversity in cultivated rice at both the intervarietal (Yang et al., 1994; Xiao et al., 1996; Beló, 2001; Pessoa, 2004) and intra-varietal (Olufowote et al., 1997) levels. In addition, many microsatellite markers reliably amplify alleles in a range of closely related non-domesticated rice species (Wu and Tanksley 1993; Panaud et al., 1995; Harrington 2000; Moncada et al., 2001; Ishii et al., 2001) and sometimes in other more-distantly related grass species (Ishii and McCouch 2000; Chen et al., 1997). The extent of allelic diversity, detected at rice microsatellite loci, depends on the specific loci being assayed, the number and diversity of the genotypes sampled and, to a certain degree, the sensitivity of the technique used to detect small molecular-weight differences.

Database of allelic frequencies and diversity analysis

Polymorphism at the 16 marker loci genotyped with the three multiplex panels was evaluated by estimates of the number of alleles detected on each locus (k), the observed gene diversity (GDo), expected gene diversity (GDe) and PIC values (Table 3).

The average number of alleles per locus was 19, ranging from 7 alleles for marker RM420 (tetranucleotide motif) to 30 alleles for marker RM250 (dinucleotide motif) (Figure 3). Recent reports (Blair et al., 2002; Ni et al., 2002; Pessoa, 2004), although sampling different marker loci, show similar variations in the number of alleles per locus, usually higher in dinucleotide repeat markers and lower in tetranucleotide repeat markers. It has to be emphasized the size and diversity of the sample of rice accessions used in this type of study have to be similar for comparison purposes.

Expected gene diversity (GDe) values were high for most of the SSR markers, with an average value of 0.892, ranging from 0.806 for RM420 to 0.943 to RM250. The average GDe value (0,892) was higher than that reported by Pessoa Filho (2004) for a collection twice as large as the one studied here (0,786). A significant correlation between GDe values and the number of alleles per locus ($r = 0.750$, $P < 0.001$), and allele size ranges (Figure 2; $r = 0.550$, $P < 0.001$) was observed, corroborating previous studies (Pessoa, 2004).

Large microsatellite repeat motifs (ex. tetranucleotides) tend to have smaller number of alleles per locus than small repeat motifs (ex. dinucleotides), but in general their allele size ranges seem to be larger (Figure 3) (Hall et al., 1996; Pessoa, 2004). Since the number of tri- and tetranucleotide repeat markers used in this study was limited, it was not possible to address the question of independence of the size of the microsatellite repeat and the allele range size.

PIC values were also high for most of the SSR markers, with an average value of 0.881, ranging from 0.777 for RM420 to 0.939 to RM250. The average PIC value (0,881) was similar to those reported by Blair et al. (2002) and Ni et al. (2002). The 16 microsatellite markers were used to estimate the probability of identical genotypes (Sefc et al., 1999), based on the allelic frequency data-

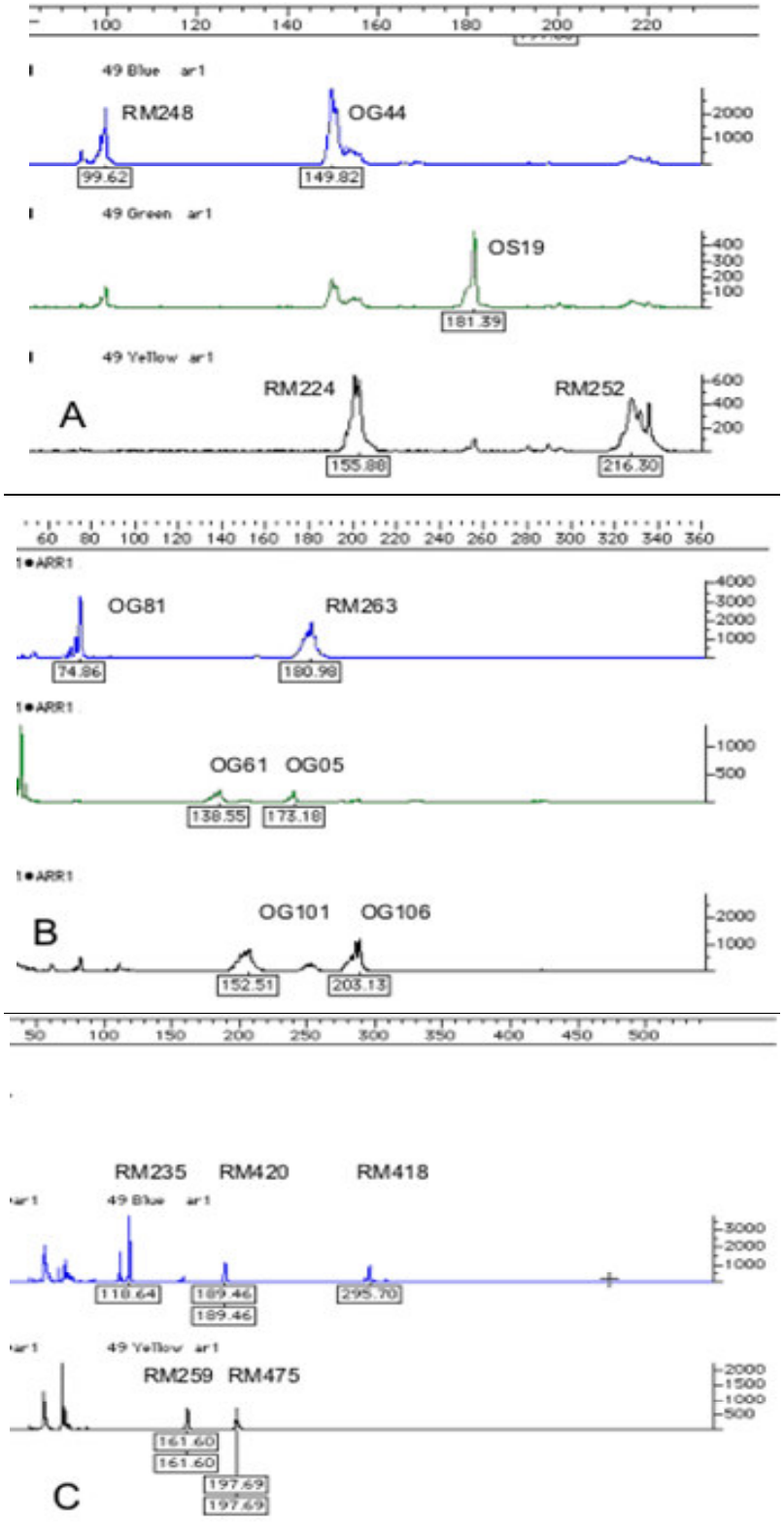

Figure 1. Electropherograms depicting the genotype of accession BR-IRGA 417 (ARR1) obtained with 16 microsatellite markers analyzed with three multiplex panels. Horizontal axis represents estimates of product sizes in base-pairs and the vertical axis indicates fluorescence intensity measured by an ABI Prism 377 DNA Sequencer.

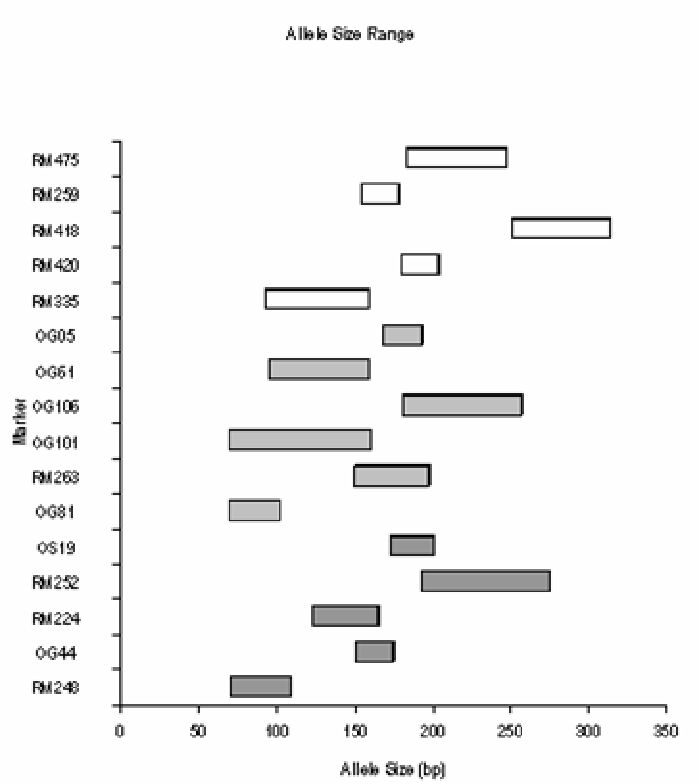

Figure 2. Allele size range of 16 microsatellite markers of three multiplex panels (gray dark and white – pentaplex, Beló, 2001; gray light; – hexaplex; Pessoa Filho, 2004).

Table 3. Descriptive statistics of 16 microsatellite marker loci genotyped in multiplex panels for a collection of 298 rice accessions.

Marker locus	k	N	GDo	GDe	PIC
RM248	17	272	0.250	0.900	0.890
OG44	13	283	0.201	0.877	0.863
RM224	19	275	0.193	0.881	0.868
RM250	30	271	0.188	0.943	0.939
OS19	15	258	0.174	0.904	0.894
OG81	16	265	0.140	0.837	0.817
RM263	20	274	0.186	0.871	0.857
OG101	26	260	0.262	0.938	0.933
OG106	29	247	0.146	0.942	0.937
OG61	27	218	0.220	0.932	0.926
OG05	13	201	0.214	0.897	0.886
RM335	21	239	0.163	0.904	0.895
RM420	7	247	0.045	0.806	0.777
RM418	21	235	0.149	0.913	0.905
RM259	13	237	0.143	0.904	0.894
RM475	18	202	0.173	0.835	0.820
Average	**19.06**	**-**	**0.1779**	**0.892**	**0.881**
Total	**305**	**-**	**-**	**-**	**-**

K, number of allele; N, number of accessions analyzed; GDo, observed gene diversity; GDe, expected gene diversity; PIC, Polymorphism Information Content.

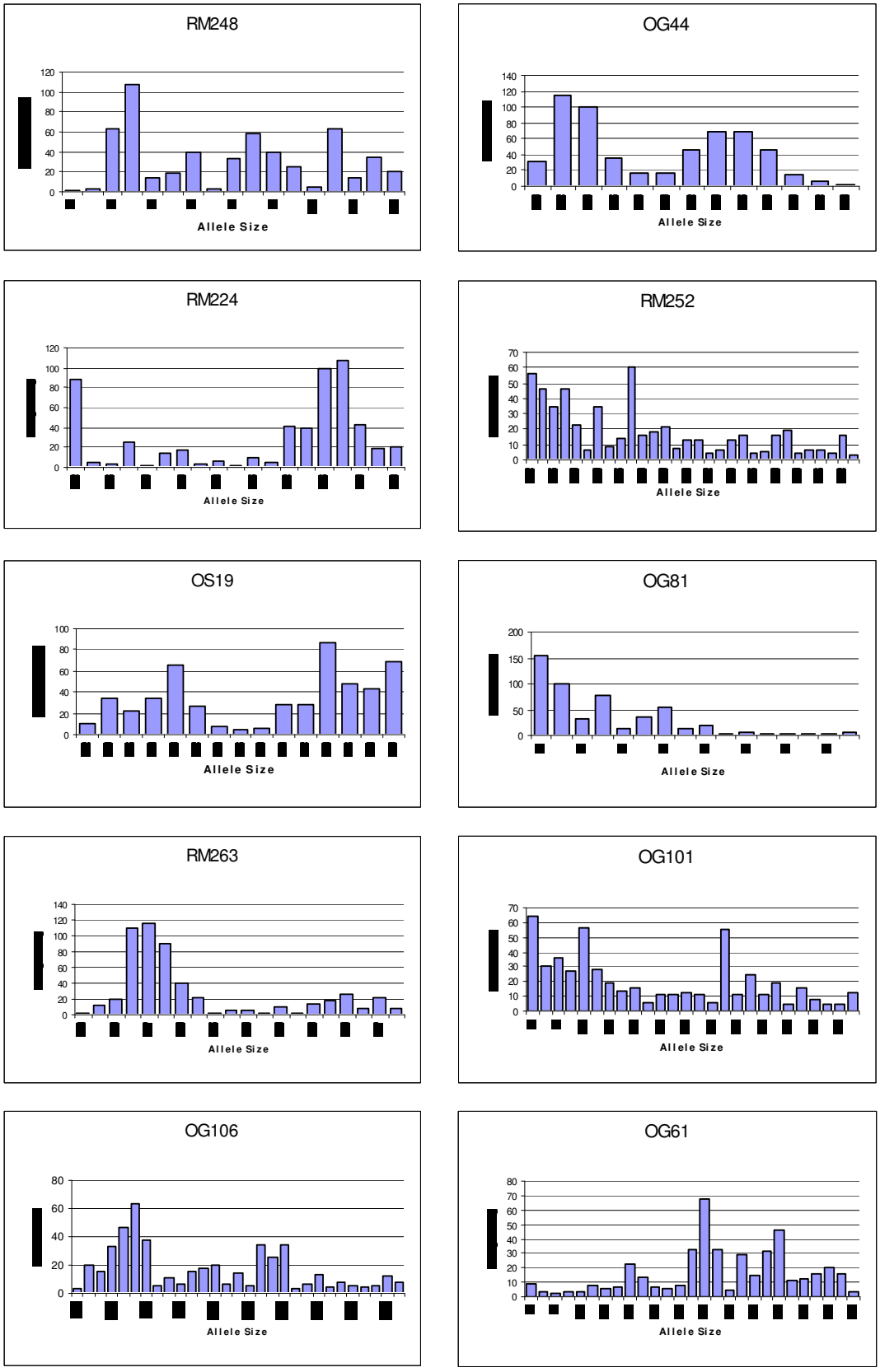

Figure 3. Bar histogram of allele frequency of 16 microsatellite markers where the Y-axis scale varies between individual graphs.

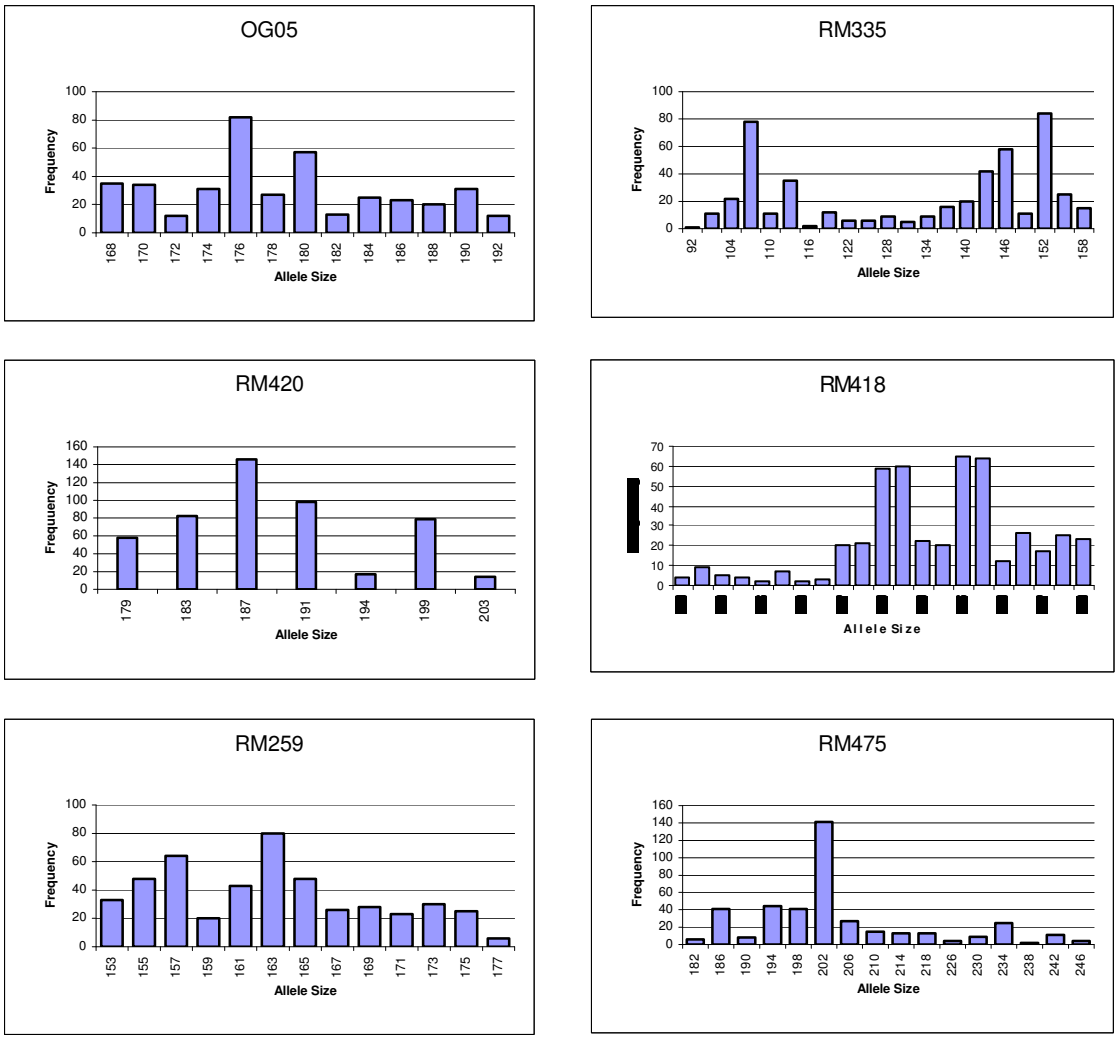

Figure 3. Bar histogram of allele frequency of 16 microsatellite markers where the Y-axis scale varies between individual graphs. (cont.).

base (Table 4a,b). The probability of identical genotypes was estimated as 1,05 x 10^{-23}. This indicates that the three multiplex panels of microsatellite markers can be efficiently used to differentiate rice individuals and accessions.

Cluster analysis of germplasm accessions

The diagonal genetic distance matrix, based on shared allele distance estimates, was used for cluster analysis of rice accessions based on the Neighbour-Joining method. The dendrogram allowed for the identification of groups of rice accessions with closer genetic relationship. The resultant genetic distance dendrogram (Figure 4) shows the accessions distributed into two main groups, one comprising 55 accessions (Cluster 1) and another composed of 243 accessions (Cluster 2). Cluster 1 includes accessions that can be classified as indica-janponica hybrids since they do not fall neither within indica nor japonica groups. Also, these accessions are not closely related to the typical *japonica* accessions included in the study (ex. Tomoe Mochi, IAC 166) as well as typical *indica* accessions (ex. IR-50, Cica 8, BG 90-2). It has to be noticed that in a similar study (Pessoa, 2004), these 55 accessions formed a cluster apart from *indica* and *japonica* typical accessions.

The second main group (Cluster 2), the largest one, is composed of *indica* and *japonica* rice subdivided in smaller clusters. In general, accessions of known *japonica* genetic background clustered with other *japonica* types. This same is true for indica types. This was important to have a first indicative of the genetic background of most of the traditional accessions in the collection. Some examples call attention. Nine accessions of red rice were genotyped for the first time with the multiplex panels of

Table 4a. Allele frequency database for rice accessions using 16 microsatellite markers.

	RM248		OG44		RM224		RM263		OG81		OS19		OG05		RM259	
#	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq
1	70	0,0037	150	0,0565	122	0,1618	149	0,0036	69	0,2943	172	0,0194	168	0,0871	153	0,0696
2	76	0,0055	152	0,2032	124	0,0091	151	0,0237	71	0,1887	174	0,0659	170	0,0846	155	0,1013
3	80	0,1158	154	0,1784	126	0,0055	153	0,0383	73	0,0585	176	0,0426	172	0,0299	157	0,135
4	82	0,1985	156	0,0636	128	0,0455	155	0,2026	75	0,1453	178	0,0678	174	0,0771	159	0,0422
5	84	0,0257	158	0,03	130	0,0036	157	0,2135	77	0,0245	180	0,1279	176	0,204	161	0,0907
6	86	0,0349	160	0,0283	132	0,0255	159	0,1661	79	0,066	182	0,0523	178	0,0672	163	0,1688
7	88	0,0717	162	0,0795	134	0,0309	161	0,073	81	0,1038	184	0,0155	180	0,1418	165	0,1013
8	90	0,0055	164	0,1219	138	0,0055	163	0,0401	83	0,0245	186	0,0097	182	0,0323	167	0,0549
9	92	0,0607	166	0,1201	140	0,0127	165	0,0055	85	0,0358	188	0,0116	184	0,0622	169	0,0591
10	94	0,1066	168	0,0795	144	0,0036	167	0,0128	87	0,0075	190	0,0543	186	0,0572	171	0,0485
11	96	0,0717	170	0,0247	146	0,0182	169	0,0109	89	0,0132	192	0,0562	188	0,0498	173	0,0633
12	98	0,046	172	0,0106	150	0,0073	171	0,0055	91	0,0075	194	0,1686	190	0,0771	175	0,0527
13	100	0,0092	174	0,0035	152	0,0745	173	0,0201	93	0,0057	196	0,093	192	0,0299	177	0,0127
14	102	0,1158	X	X	154	0,0709	177	0,0055	95	0,0038	198	0,0833	X	X	X	X
15	104	0,0257	X	X	156	0,1818	179	0,0274	97	0,0057	200	0,1318	X	X	X	X
16	106	0,0643	X	X	158	0,1945	181	0,0347	101	0,0151	X	X	X	X	X	X
17	108	0,0386	X	X	160	0,0764	183	0,0474	X	X	X	X	X	X	X	X
18	X	X	X	X	162	0,0345	185	0,0146	X	X	X	X	X	X	X	X
19	X	X	X	X	164	0,0382	187	0,0401	X	X	X	X	X	X	X	X
20	X	X	X	X	X	X	197	0,0146	X	X	X	X	X	X	X	X
21	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
22	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
23	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
24	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
25	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
26	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table 4a. Continued

27	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Table 4b. Allele frequency database for rice accessions using 16 microsatellite markers (cont.).

	RM475		RM420		RM418		RM335		RM252		OG106		OG61		OG101	
#	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq	Allele	Freq
1	182	0,0149	179	0,1174	250	0,0085	92	0,0021	192	0,1033	180	0,0061	94	0,0206	93	0,1231
2	186	0,1015	183	0,166	253	0,0191	101	0,023	194	0,0849	200	0,0405	96	0,0069	95	0,0577
3	190	0,0154	187	0,2955	259	0,0106	104	0,046	196	0,0627	202	0,0324	98	0,0046	97	0,0692
4	194	0,1089	191	0,1984	263	0,0085	107	0,1632	198	0,0849	204	0,0668	100	0,0069	99	0,0519
5	198	0,1015	194	0,0344	265	0,0043	110	0,023	200	0,0424	206	0,0951	102	0,0069	101	0,1096
6	202	0,349	199	0,1599	267	0,0149	113	0,0732	202	0,0129	208	0,1275	104	0,0183	103	0,0538
7	206	0,0668	203	0,0283	271	0,0043	116	0,0042	204	0,0646	210	0,0769	106	0,0138	105	0,0365
8	210	0,0371	X	X	274	0,0064	119	0,0251	214	0,0166	212	0,0121	108	0,0161	107	0,0269
9	214	0,0322	X	X	277	0,0426	122	0,0126	216	0,0258	214	0,0223	110	0,0528	111	0,0308
10	218	0,0322	X	X	280	0,0447	125	0,0126	218	0,1107	216	0,0142	114	0,0321	113	0,0115
11	226	0,0099	X	X	283	0,1255	128	0,0188	220	0,0295	218	0,0324	116	0,0161	117	0,0212
12	230	0,0223	X	X	286	0,1277	131	0,0105	222	0,0332	220	0,0364	118	0,0138	121	0,0212
13	234	0,0619	X	X	289	0,0468	134	0,0188	224	0,0406	222	0,0405	120	0,0183	127	0,0231
14	238	0,005	X	X	292	0,0426	137	0,0335	226	0,0148	224	0,0142	130	0,0757	129	0,0212
15	242	0,0272	X	X	295	0,1383	140	0,0418	228	0,024	226	0,0304	134	0,156	131	0,0115
16	246	0,0099	X	X	298	0,1362	143	0,0879	230	0,024	228	0,0121	136	0,0757	135	0,1058
17	X	X	X	X	301	0,0255	146	0,1213	232	0,0074	230	0,0688	138	0,0092	137	0,0212

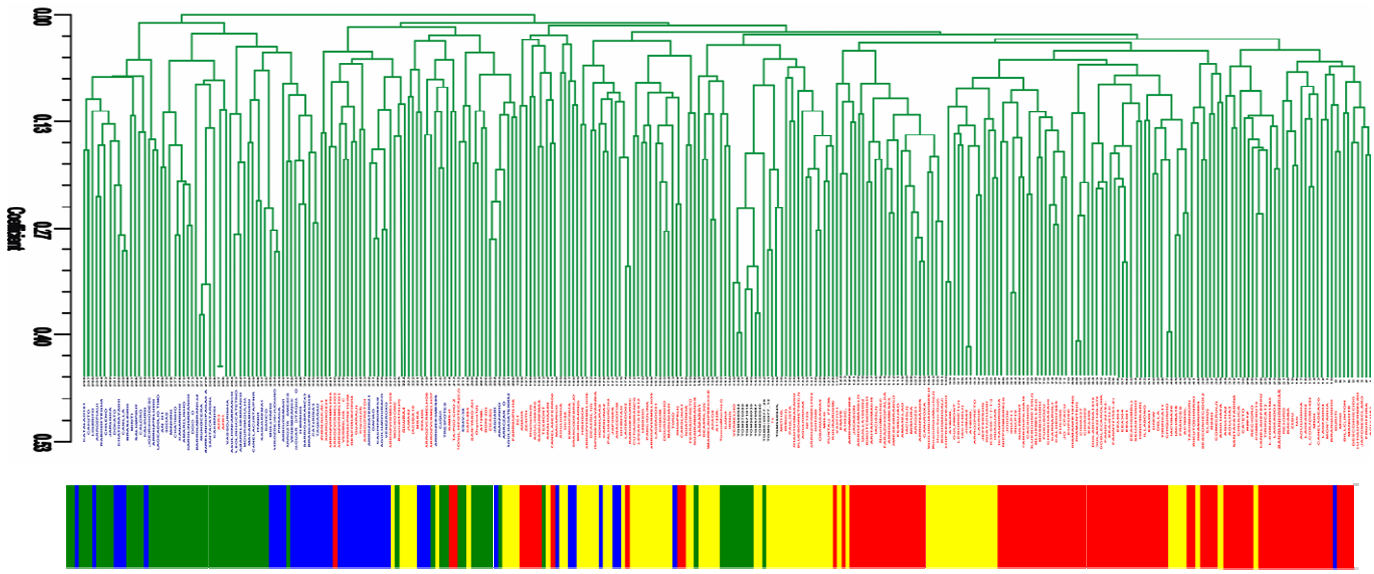

Figure 4. Genetic structure analysis of 298 accessions of rice *O. sativa* L. based on DNA polymorphism at 16 microsatellite loci. The dendrogram reflects the cluster analysis performed using the neighbour-joining method (Saitou and Nei, 1987) based on the shared allele coefficient while the coloured bar reflects the genetic structure analysis based on the Bayesian clustering method performed according to Pritchard et al. (2000a).

Table 5. Estimates of Wright's F parameters for the germplasm collection. Clusters of accessions were defined according to Neighbour-Joining analysis of genetic distance data based on microsatellite polymorphism; 95% confidence intervals were estimated based on bootstrap analysis (1000x).

Wright's F	Two clusters	Nine clusters
F_{IS}	0.786 (0.763 - 0.811)	0.773 (0.747 - 0.798)
F_{IT}	0.802 (0.782 - 0.827)	0.794 (0.772 - 0.820)
F_{ST}	0.077 (0.051 - 0.103)	0.095 (0.073 - 0.121)

microsatellite markers. These accessions have been collected in small villages of North eastern Brazil, where traditionally the farmers give more value to red rice than to white rice for culinary reasons. The accessions Arroz Vermelho 1, Arroz Vermelho 2, Arroz Vermelho 3 and Arroz Vermelho 9 form a cluster with Ciwini and Basmati 370, known as *indica* types (Beló, 2001; Fatima et al., 2002). Basmati 370 is aromatic rice, with fine quality and extra fine grains and peculiar elongation after cooking. However, this cluster also includes a variety called Mochi Gome, which is a common name for brown, glutinous (sticky) varieties of rice grown in Japan. The other Arroz Vermelho accessions also cluster with indica types in other groups. For example, Arroz Vermelho (7) and Arroz Vermelho (8) cluster with Oryzica 1, Cica 8 e BG 90-2, while Arroz Vermelho (4) clusters with the TGMS indica lines. The accessions Arroz Vermelho 5 and Arroz Vermelho 6 belong to a separate cluster which did not include any well known *indica* types and is composed mostly of accessions with very common names. It could be observed, though, that the traditional varieties called Arroz Vermelho are very diverse among themselves and could be readily distinguished based on multiloci - genotyping analysis.

All thermosensitive genic male sterility (TGMS) lines clustered together (lines TGMS70977_12, TGMS70977_26, TGMS68940, TGMS70987, TGMS68935, TGMS71018, TGMS70989, TGMS68944 and TGMS68945), forming a group with the *indica* lines IR 36, IR50 and Metica 1. The TGMS68935 line was crossed with IR50 (not tgms) in order to develop segregant populations that were used to map TGMS controlling loci in the rice genome.

It was interesting to notice that accessions with the same common name sometimes cluster together (ex. Agulha (1), Agulha (2) and Agulha (3) or Carolina (1) and Carolina (2)) while others (ex. Caqui or DeAbril) do not. This is a clear indication that genotypes with the same common name are genetically different, especially when traditional varieties are being considered.

Genetic structure of the collection

With the aim to analyze the cluster consistency given by the Neighbour-Joining method, the level of genetic differentiation between clusters of accessions obtained with the pairwise distance data were estimated based on Wright's *F* parameters (Table 5). The average genetic differentiation (F_{ST} = 0.08616) between clusters (Weir and Cockerham, 1984) was considered moderate (Wright, 1978), estimated as 0.077 when the two main clusters defined by the Neighbour-Joining method were considered and 0.095 when nine clusters were considered.

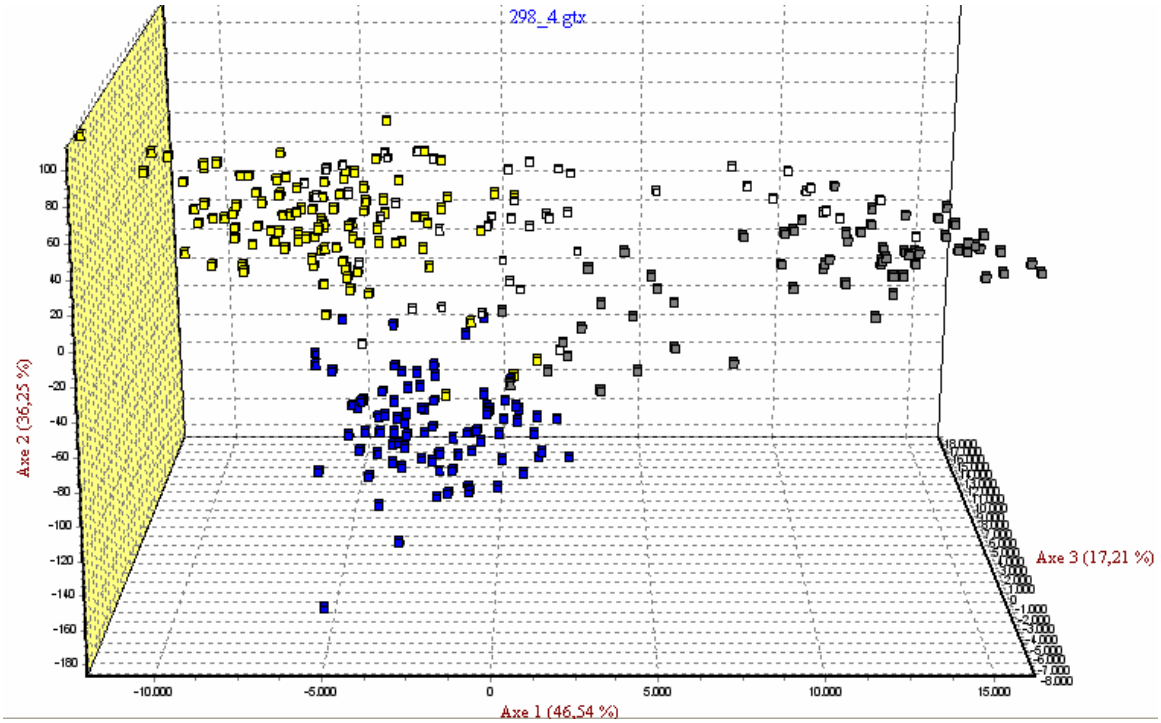

Figure 5. Classical factorial analysis of 298 accessions of rice *O. sativa* L. based on DNA polymorphism at 16 microsatellite loci. Genetic distance estimates were based on the shared allele coefficient and cluster analysis was performed using the Neighbour-Joining method (Saitou and Nei, 1987).

The clustering of the accessions as obtained by distance analysis, therefore, accounted for only about 8% of the genetic differentiation.

The estimates of Wright's F-statistics (Weir and Cockerham, 1984) revealed a high value of inbreeding coefficient when two or nine subpopulations (clusters) are considered in the collection (F_{IS}=0.786 and 0.773, respectively) (Table 5), as well as a relatively high overall fixation index (F_{IT}= 0.779). On the other hand, the F_{ST} values indicate moderate genetic differentiation when two or nine subpopulations are considered to compose the collection (F_{ST}= 0.077 and 0,095, respectively). This is a clear indication that the extent of inbreeding within clusters is high, but the extent of genetic differentiation among clusters is only moderate.

Further analysis was performed using the software Structure 2.1 (Pritchard et al., 2000b; Falush et al., 2003) to infer the level of genetic differentiation between subpopulations (clusters) defined according to the Neighbour-Joining analysis (two and nine clusters) and without a specific *a priori* model. The best resultant K values (number of subpopulations) were provided by the matrices without a specific *a priori* model and defined based on two clusters of the Neighbour-Joining analysis. For both matrices, a best-fit probability of 1.00 was obtained for K=4, indicating a genetic differentiation of the collection into four subpopulations (Figure 4).

The proportion of shared ancestry among the four genetically defined subpopulations detected in the total of 298 accessions show an interesting pattern. The F-statistics for K=4 were F_{IS}= 0.637; F_{ST}= 0.035 and F_{IT}= 0.650 (Weir and Cockerham, 1984). Again, the data indicated that the extent of inbreeding within clusters is high, but genetic differentiation among clusters is only moderate. The four groups comprised 63, 45, 106 and 81 accessions, respectively. Group 1 shared 20.63% of ancestry with group 2, 9.52% with group 3 and less than 5.0% with groups or mixed groups 2-3, 2-4, 3-4 and 4. Group 2 shared 11.11% ancestry with Group 1, 13.33% with Group 3 and 8.89% with Group 4. Group 3 shared 10.38% ancestry with Group 2, 8.49% with Group 4 and less than 5.00% with groups or mixed groups 1, 1-2, 2-4, 4-1 and 4-2. Group 4 shared 11.11% ancestry with Group 1, 12.00% with Group 2, 12.35% with Group 3, and less than 5.00% with groups or mixed groups 1-2, 1-3 and 2-3. A Classical Factorial Analysis (Belkhir, 2001) allowed for the spatial separation of the four subpopulations depicted by the MCMC approach (Figure 5).

The genetic stratification suggested by Neighbour-Joining method (Saitou and Nei, 1987) does not correspond (r= 0.75, P >0.10) to the same subpopulations (clusters) suggested by the MCMC method (Figure 4) (Pritchard et al., 2000b; Falush et al., 2003). This was expected due to differences of methodologies, since the Neighbour-Joining method groups accessions based on genetic distances (that is, shared allele distances) and the MCMC

approach groups them according to the probability of two alleles having common ancestry. Each method has its own strengths and constraints, and there is no single or simple strategy to address effectively various complex issues related to choice of distance measure(s), clustering methods, determination of optimal number of clusters or analysis of individual, and combined data sets by means of various statistical tools. However, empirical data generated in recent years by different strategies has provided an enhanced understanding of the above issues, and reasonably effective means of analysing genetic diversity at various levels (individuals, populations, or species).

Different levels of genetic structure have been previously documented in natural populations (Barbier, 1989b; Gao et al., 2000a,b; Gao et al., 2001) and germplasm collections of rice (Glaszmann, 1987; Ni et al*.,* 2002; Parsons et al*.,* 1999; Garris et al., 2003; Gao et al., 2000a,b; Gao et al., 2001). In natural populations of an autogamous species such as rice, one is expected to find a greater proportion of diversity among subpopulations rather than within subpopulations. Germplasm collections, however, artificial by nature and several forces, including genetic drift and man driven artificial selection, are playing a role in the way they are structured. High values of F_{ST} were estimated in wild rice, indicating strong differentiation in natural populations of *O. officinalis* (F_{ST} = 0.882; Gao et al., 2001), *O. rufipogon* f (F_{ST} = 0.310; Barbier 1989b; Gao et al., 2000a) and *O. granulata* (F_{ST} = 0.859; Gao et al., 2000b). High values of fixation index (F_{ST} > 0.85) have also been reported in studies of rice varieties (ecotypes of *O. sativa*) from different countries showing diversity for resistance to *Xanchomonas oryzae pv oryzae* (Garris et al., 2003). Likewise, a diverse germplasm composed of rice (*O. sativa)* accessions from several parts of the world showed values of F_{ST} ranging from 0.20 to 0.43 (Garris et al., 2005). It has to be stressed that the collection analyzed here is not representative of the genetic diversity of rice and the level of differentiation observed is considered moderate. But it is interesting to notice the possibility of detecting population structure with a small set of microsatellite markers distributed in three multiplex panels, corroborating the usefulness of highly informative markers in population genetic analysis.

With the recent development and use of model-based clustering methods based on Bayesian statistics (Pritchard, 2001; Thornsberry et al., 2001), the possibility of carrying out association studies in crop plants for identifying genes for agronomically important but complex traits have been enhanced. There is still a distinct need for developing comprehensive and user-friendly statistical packages that facilitate an integrated analysis of different data sets for generating reliable information about genetic relationships, germplasm diversity, and favourable allele variation. Equally important, and perhaps more challenging, is the concerted and planned utilization of germplasm in crop breeding programs on the basis of knowledge obtained from studies on genetic diversity (Mohammadi and Prasanna, 2003).

Conclusions

- The automated genotyping using multiplex panels of fluorescent microsatellite markers were efficient in analysing rice accessions since they yield, in average, a large number of allele per locus with a broad size range between the lower and the higher number of alleles. That genotyping method associated with microssatellite markers made it possible to reproduce some early analysis and, also, to detect new alleles as the sample size of rice accessions was increased.
- The Genebank of Embrapa Rice and Beans is conserving highly diverse rice germplasm accessions which hold potential for use in breeding programs once those accessions showing high levels of genetic diversity. The expected gene diversity was significantly correlated with the number of alleles per locus and the alleles size range.
- The three panels of microsatellite markers can be efficiently used to differentiate rice individuals and accessions based on the probability of identical genotypes estimated from the allelic frequency database.
- Rice germplasm are distinguished in two different main groups of accessions and, in general, accessions of known *japonica* genetic background clustered with other *japonica* types while accessions of known *indica* genetic background clustered with other *indica* types. This gives a first indication of the genetic background of most of the traditional accessions in the collection. Also, the cluster gives a clear indication that several genotypes with the same common name are genetically different, especially when traditional varieties are being considered.
- Different softwares used for analysis of genetic structure of rice accessions are suitable for confirming the high degree of inbreeding of rice species due to its sexual reproduction system but, they separate the accessions into different subpopulations due to differences of methodologies. Therefore, there is still a distinct need for developing comprehensive and user-friendly statistical packages that facilitate an integrated analysis of different data sets for generating reliable information about genetic relationships, germplasm diversity, and favourable allele variation. Equally important, and perhaps more challenging, is the concerted and planned utilization of germplasm in crop breeding programs on the basis of knowledge obtained from studies on genetic diversity.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to André Beló and Pessoa-Filho to make available the microsatellite multiplex panels and the Embrapa-Genetic Resources and Biotechnology for the grant.

REFERENCES

Barbier P (1989b). Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations. Japanese Journal Genetics 64:273–285

Barrett BA, Kidwell KK (1998.) AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci. 38: 1261–1271.

Belkhir K, Borsa P, Chilhi L, Raufaste N, Bonhomme F(2001). Genetix 4.02, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Belo A (2001). Avaliação de recursos genéticos para produção de híbridos de arroz (Oryza sativa L.) Tese de Mestrado, UFSC, Florianópolis, Brasil.

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

Botstein D, White RL, Skolnickm M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Gene. 32: 314–331.

Bowcock M, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994). High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368: 455-457

Brondani C, Rangel N, Brondani V, Ferreira E (2002). QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. Theor. Appl. Genes. 104(6-7): 1192-1203.

Brondani RPV, Grattapaglia D (2001). Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. BioTechniques 31: 793-800.

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. Gene. 95: 553-567.

Cox TS, Murphy JP, Rodgers DM (1986). Changes in genetic diversity in the red winter wheat regions of the United States. Proc. Natural. Academy. Science. (USA). 83: 5583–5586.

Falush D, Stephens M, Pritchard JK (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Gene. 164: 1567-1587.

Ferreira ME, Grattapaglia D (1998). Introdução ao uso de marcadores moleculares em análise genética. 220pp. Embrapa –SPI.

Frankel OH, Brown AHD (1984). Plant genetic resources today: A critical appraisal. Pp. 249–257 in Crop Genetic Resources: Conservation and Evaluation, Holden JHW, Williams JT, eds. London: Allen and Unwin.

Gao LZ, Ge S, D-Y Hong DY (2001). High Levels of Genetic Differentiation of Oryza officinalis Wall. ex Watt. From China. J. Hered. 92(6): 511-516.

Gao LZ, Ge S, Hong DY (2000a). Allozymic diversity and genetic structure of common wild rice *Oryza rufipogon* Griff. in China. Theor. Appl. Gene. 101: 494–502.

Gao LZ, Ge S, Hong DY (2000b). Low levels of genetic diversity within populations and high differentiation among populations of a wild rice, *Oryza granulata* Nees et Arn. ex Watt. from China. Int. J. Plant Sci. 161: 691–697.

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.).Gene. 165: 759-769.

Garris AJ, Tai TH, Coburn J, Kresovich S, Mccouch S (2005). Genetic structure and diversity in *Oryza sativa* L. Genetics: Published Articles Ahead of Print, published on January 16, 2005 as 10.1534

Glaszmann JC (1987). Isozymes and classification of Asian rice varieties. Theoretical and Applied Genetics 74 21–30 (1987).

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995). An evaluation of genetic distances for use with microsatellite loci. Gene. 139: 463-471.

Hall JM, Leduc CA, Watson AR, Roter AH (1996). An approach to high-throughput genotyping. Genome Research 6:781–790

Harrington S (2000). "A survey of genetic diversity of eight AA genome species of Oryza using microsatellite markers." MS thesis Cornell University, Ithaca, NY

Hartl DL (1987). A Primer of Population Genetics. 2nd Edition. Sinauer Associates, Massachussetts. p. 305.

Ishii T, Mccouch SR (2000). Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. Theoretical and Applied Genetics 100:1257–1266

Ishii T, Xu Y, Mccouch SR (2001). Nuclear and cholorplast microsatellite variation in A-genome species of rice. Genome 44: 648–666.

Litt M, Luty JA (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hum. Gen. 44: 397-401.

Marshall TC, Slate J, Kruuk L, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7(5): 639-655.

Mohammadi SA, Prasanna BM (2003). Analysis of genetic diversity in crop
plants-salient statistical tools and considerations. Crop Sci. 43: 1235–1248.

Moncada P, Martinez CP, Borrero J, Chatel M, Gauch H, Guimaraes E, Tohme J, Mccouch SR (2001). Quantitative trait loci for yield and yield components in an *Oryza sativa* X *Oryza rufipogon* BC2F2 population evaluated in an upland environment. Theor. Appl. Gen. 102(1): pp. 41-42.

Nei M (1987). Molecular evolutionary genetics. Columbia University and heterosis data. Maydica. 37: 53–60.

Ni J, Colowit PM, Mackill DJ (2002). Evaluation of genetic diversity in rice subspecies using microsatellite markers. Crop Sci. 42: 601-607.

Olufowote JO, Xu Y, Chen X, Park WD 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.

Panaud O, Chen X, Mccouch SR (1995). Frequency of microsatellite sequences in rice (Oryza sativa L.). Genome 38: 1170–1176.

Parsons BJ, Newbury HJ, Jackson MT, Ford-Lloyd BV (1999). The genetic structure and conservation of aus, aman and boro rices from Bangladesh. Genetic Resources and Crop Evolution. 46: 587-598.

Pessoa FMACP (2004). Análise Genética de Variedades Tradicionais do Banco de Germoplasma de Arroz (Oryza sativa L.) com Marcadores Moleculares. Tese de Mestrado. Departamento de Biologia Molecular. Universidade de Brasília.

Pritchard JK (2001). Deconstructing maize population structure. Nature Gen. 28: 203-204.

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

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b). Association mapping in structured populations. American Journal of Human Genetics 67: 170-181.

Pritchard JK, Wen W (2004). Documentation for structure software: version 2. Department of Human Genetics. University of Chicago 920 E 58tjh CLSC 507 Chicago IL 60637, USA (http:\\pritch.bsd.uchicago.edu).

Rohlf FJ (1998). NTSYS-pc (Numerical Taxonomy and Multivariate performance. pp. 621–628. In: Buxton DR (ed.) Proceedings of the Analysis System). Version 1.70. Exeter, Setauket, NY. International Crop Science Congress, 1st, Ames, IA. July.

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

Sasaki S, Obara M, Kashiba K, Sato T, Yano M, Ebitani T, Yamaya T (2002). Linkage analysis and characterization for QTL on chromosome 2 that associated with cytosolic glutamine synthetase content and panicle weight in rice (Oryza sativa L.). Plant and Cell Physiology 43: S71-S71.

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

Sefc K, Regner F, Turetschek E, Gloessl J, Steinkellner H (1999). Identification of microsatellite sequences in *Vitis riparia* and their applicability for genotyping of different *Vitis* species. Genome 42: 376-373.

Smith JSC (1984). Genetic variability within U.S. hybrid maize

Multivariate analysis of isozyme data. Crop Sci. 24: 1041–1046.
Tautz, D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17: 6463-6471.
Wang ZY, Tanksley SD (1989). Restriction Fragment Length Polymorphism in Oryza sativa L. Genome 32: 1113-1118.
Weber JL, May PE (1989). Abundant class of human DNA polymorhisms which can be typed using the polymerare chain reaction. Am. J. Hum. Gen. 44: 388-396.
Weir BS, Cockerham CC (1984). Estimating *F*-statistics for the-analysis of population structure. Evolution 38: 1358–1370.
Wright S (1978). Evolution and the Genetics of Populations, Vol. IV. Variability Within and Among Natural Populations. University of Chicago Press, Chicago.
Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol. Gene Gen. 241: 225-235.
Xiao J, Li J, Yuan L, Mccouch SR, Tanksley SD (1996). Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor. Appl. Gen. 92: 637–643.
Yamamoto T, Kuboki Y, Lin SY, Sasaki T, Yano M (1999). Fine mapping of quantitative trait loci, *Hd-1*, *Hd-2* and *Hd-3*, controlling heading date of rice as single Mendelian factors. Theor. Appl. Gen. 97: 37-44.
Yang GP, Maroof M, Xu CG, Zhang Q, Biyashev RM (1994). Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gene Gen. 245: 187–194.
Zhang J, Zheng HGM, Aarti A, Pantuwan G, Nguyen TT, Tripathy JN, Sarial AK, Robin S, Babu RC, Nguyen BD, Sarkarung S, Blum A, Nguyen HT (2001). Locating genomic regions associated with components of drought resistance in rice comparative mapping within and across species. Theor. Appl. Gene. 10: 319-329.
Zhang QF, Maroof MAS, Lu TY, Shen BZ (1992). Genetic diversity and differentiation of indica and japonica rice detected by RFLP analysis. Theor. Appl. Gen. 83: 495–499.