African Journal of Biotechnology Vol. 7 (23), pp. 4219-4229, 3 December, 2008
Available online at http://www.academicjournals.org/AJB
ISSN 1684–5315 © 2008 Academic Journals

Full Length Research Paper

Molecular profiling of interspecific lowland rice populations derived from IR64 (*Oryza sativa*) and Tog5681 (*Oryza glaberrima*)

Marie Noelle Ndjiondjop[1]*, Kassa Semagn[2], Moussa Sie[1], Mamadou Cissoko[1], Blandine Fatondji[1] and Monty Jones[3]

[1]Africa Rice Center (WARDA), 01 BP 2031, Cotonou, Benin.
[2]International Maize and Wheat Improvement Center (CIMMYT), P. O. Box 1041, Nairobi, Kenya.
[3]Forum for Agricultural Research in Africa, PMB CT 173 Cantonments, Accra, Ghana.

Accepted 24 October, 2008

Several lowland NERICAs (New Rice for Africa) were derived from crosses between IR64 (an *Oryza sativa* subsp. *indica* variety) and Tog5681 (an *Oryza glaberrima* variety) that possess useful traits adapted to lowland conditions in West Africa. The proportion of parental genomic contribution and extent of genetic differences among these sister lines is unknown at the molecular level. The objectives in this study were therefore to determine, with 60 SSR markers that cover 1162 cM of the rice genome, the frequency and magnitude of deviations from the expected parental contributions among 21 BC_2F_{10}, 17 BC_3F_8 and 10 BC_4F_8 lines and determine patterns of their genetic relationships. The estimated average *O. glaberrima* genome coverage was 7.2% (83.5 cM) at BC_2F_{10}, 8.5% (99.3 cM) at BC_3F_8 and 8.1% (93.8 cM) at BC_4F_8 lines. The *O. sativa* parent accounted for 73.2% (851.3 cM) at BC_2F_{10}, 82.6% (959.5 cM) at BC_3F_8 and 78.2% (908.6 cM) at BC_3F_8. Non-parental alleles were detected at all 3 backcross generations but the frequency of such alleles at BC_2 (8.8%) was twice that of BC_3F_8 (3.4%) and nine times that of BC_4F_8 (0.9%). Both cluster and principal component analyses revealed two major groups irrespective of the level of backcross generations and the proportion of parental genome contribution.

Key words: African rice, introgression, microsatellite, NERICA, *Oryza glaberrima*, rice.

INTRODUCTION

In areas of high population density, African farmers are intensifying cropping on fragile upland soils leading to severe environmental degradation and loss in production potential. The development of appropriate lowland rice technologies is one of the strategies to help farmers to reduce pressure on degraded uplands by shifting cultivation to the lowlands. To improve African rice production, breeders at the Africa Rice Center (WARDA) and the National Agricultural Research Systems (NARS) have continued in developing interspecific rice varieties by crossing the two cultivated species, Asia rice (*Oryza sativa*) and Africa rice (*Oryza glaberrima*). Various studies have shown that very useful agronomic traits exist within the traditional landraces of *O. glaberrima*. These include (i) rapid and profuse vegetative growth coupled with droopy lower leaves (Jennings et al., 1979) that contribute to weed competitiveness (Fofana et al., 1995); (ii) moderate to high levels of resistance to blast (Silue and Notteghem, 1991), rice yellow mottle virus (Attere and Fatokun, 1983; John et al., 1985), rice gall midge (Sauphanor, 1985; Alam, 1988) and nematodes (Reversat and Destombes, 1995); (iii) reasonably good levels of tolerance to soil acidity, iron toxicity and drought (Sano et al., 1984; Jones et al., 1994). The grain of *O. glaberrima* also has good aroma and taste. However, cultivation of *O. glaberrima* is being increasingly replaced by *O. sativa* varieties due to low yield-performance, high shattering and susceptibility to lodging (Jones et al., 1997).

Interspecific hybridization is an important tool to transfer desirable genes between species (Roy, 1984; Katiyar et al., 1998), generate morpho-physiological variation

*Corresponding author. E-mail: m.ndjiondjop@cgiar.org.

(Choudhary and Joshi, 2001), elucidate intergenomic relationships (Mizushima, 1980; Attia et al., 1987; Choudhary et al., 2000), and develop synthetic amphidiploids (Olsson, 1960). Non-homologous pairing between different genomes (Mizushima, 1980; Choudhary et al., 2000) provides immense opportunities for generating variability and broadening the gene pool through interspecific hybridization.

Earlier efforts to transfer useful genes between *O. sativa* and *O. glaberrima* have been frustrated by high spikelet sterility in the progeny. In the early 1990s, WARDA developed over 3000 interspecific BC_2 inbred lines from a cross between an upland *O. sativa tropical japonica* variety, WAB 56-104, as the recipient parent, and an *O. glaberrima* variety, CG14, as the donor parent. Several upland NERICA (New Rice for Africa) varieties were released for production in different parts of Africa. The most popular upland NERICA varieties combine the best traits of both parents: high yield presumably derived from the *O. sativa* parent and the ability to thrive in harsh environments from the *O. glaberrima* parent (Jones et al., 1997; www.warda.org). The NERICA varieties offer real hope for improving the productivity, profitability, and sustainability of rice farming in sub-Sahara Africa. WARDA and NARS breeders then began making crosses between *O. sativa indica* (IR64) and *O. glaberrima* (Tog5681) to create interspecific lines adapted to lowland conditions. This work have resulted in the generation of many stable and high yielding interspecific lowland NERICA varieties that are well adapted to lowland rice production systems of the region. The proportion of parental genomic contribution and extent of genetic differences among these lowland sister lines is unknown at the molecular level.

Three recent advances have rapidly changed the face of rice genomic research: the use of microsatellite or simple sequence repeats (SSRs), the availability of functional markers such as expressed sequence tags (ESTs), and the completion of the rice genomic sequence (International Rice Genome Sequencing Project, 2005). Microsatellites are considered ideal markers for genetic studies because they are co-dominant, multiallelic, highly polymorphic even in closely related individuals, high abundance and uniform distribution in plant genomes (Morgante and Olivieri, 1993; Rafalski and Tingey, 1993; Sharma et al., 1995; Brondani et al., 1998). The sequence information and map positions of rice SSR markers are publicly available (http://www.gramene.org) and more rice SSR markers are being developed to tag any possible polymorphic parents (Temnykh et al., 2000; McCouch et al., 2002; International Rice Genome Sequencing Project, 2005). Many research groups have used SSR markers for various purposes, including estimation of the proportion of donor genome in a recurrent parent background (Bernardo et al., 2000; Heckenberger et al., 2005b; Ndjiondjop et al., 2006; Semagn et al., 2007). Semagn et al. (2007) used SSR markers to examine the proportion of introgression and extent of genetic differences among 70 BC_2 interspecific upland rice lines developed by crossing a *tropical japonica* variety (WAB 56-104) as the recurrent parent to an *O. glaberrima* variety (CG 14) as the donor parent. That study revealed an average introgression of 6.3% from the *O. glaberrima* parent and a recovery of 87.4% of the *O. sativa* recurrent parent. The objectives of the present study were to (a) estimate the extent of introgressions from the donor, and the degree and rate of recovery of the recurrent parent alleles among 21 BC_2F_{10}, 17 BC_3F_8 and 10 BC_4F_8 interspecific lines developed by crossing *O. glaberrima* (Tog5681) as donor parent and *O. sativa* subsp. *indica* (IR64) as recurrent parent, and (b) determine the genetic relationships among these 48 lines and evaluate the potential breeding value of the lines that have not yet been released as variety.

MATERIAL AND METHODS

Plant materials

Sixty lowland NERICA lines, developed by crossing an *O. sativa* subsp. *indica* variety IR64 as recurrent parent and *O. glaberrima variety* Tog5681 as donor parent, were selected by WARDA Varietal Nomination Committee. Forty eight of these interspecific lines (Table 1) were used in this study. IR64 is an irrigated improved Asian variety with several desirable agronomic traits, such as high yield, short growth duration and plant height adapted to irrigated conditions. Tog5681 is a variety from Nigeria and it has low yield potential due primarily to grain shattering and susceptibility to lodging (Jones et al., 1997). However, Tog5681 has several useful traits, including long panicle length, high weed-competitiveness ability as a result of early vigor and high tiller number, resistance to rice yellow mottle virus (Ndjiondjop et al., 1999), and resistance to nematodes *Heterodara sacchari* (Lorieux et al., 2000). All lines in this study had gone either through two generations (21 lines), three generations (17 lines) or four generations (10 lines) of backcrossing followed by eight to ten generations of selfing for fixing the lines (Table 1).

Genotyping and data analyses

DNA was extracted from 2 weeks old seedlings using a Cetyl trimethylammonium bromide (CTAB) protocol as described by Saghai-Maroof et al. (1984). One hundred and forty SSR primers were selected according to their position on the genetic map by Temnykh et al. (2001) and screened for polymorphism between parents. Sixty polymorphic markers (Table 2) were then used for genotyping the 48 interspecific lines. Polymerase chain reactions (PCR), PCR product separation, data scoring and statistical analyses were performed as described by Semagn et al. (2007). Four different statistical analyses were performed. First, the proportion of the donor and recurrent parental genome was calculated using the software package GGT, 2007 edition (http://www.dpw.wau.nl/pv/pub/ggt; Van Berloo, 1999). Map distances in centiMorgans (abbreviated as cM) between markers were used as the basis for estimating parental contribution per chromosome and line. Second, statistical differences in the proportion of parental genome contributions, heterozygosity and non-parental alleles among the 3 backcross generations were examined using one-way analysis of variance (ANOVA) and tested with LSD for pairwise comparisons using SPSS version 12 (SPSS, Inc.). Third, simple matching coefficients (the ratio of number of matches to total number markers) were calculated as a measure of

Table 1. Pedigree and parental contribution for 21 BC_2F_{10}, 17 BC_3F_8 and 10 BC_4F_8 interspecific lines developed using an *O. sativa* variety IR64 and an *O. glaberrima* variety Tog5681. Lines released as varieties are indicated in boldface, with variety name given in parenthesis in the first column. Expected donor/recurrent parent contributions with Mendelian inheritance for BC_2, BC_3 and BC_4-derived inbred lines are 12.5/87.5, 6.25/93.75% and 3.13/96.88%, respectively.

Line name (variety name)	WARDA designation	Pedigree	Level of backcross	Genome composition (%)				
				Tog 5681	IR64	Heterozygotes	Non parental	Missing data
WAS 122-IDSA 10-WAS 1-1 FKR 1	NERICA-L-1	TOG5681/IR64/2* IR64	BC_2	11.7	83.7	0.0	0.0	0.0
WAS 122-IDSA 10-WAS 6-1 FKR 1	NERICA-L-2	TOG5681/IR64/2* IR64	"	6.6	60.0	0.0	28.4	29.5
WAS 122-IDSA 11-WAS 11-4-FKR 1	NERICA-L-3	TOG5681/IR64/2* IR64	"	5.7	75.7	4.2	13.3	13.2
WAS 122-IDSA 11-WAS 8-2	NERICA-L-4	TOG5681/IR64/2* IR64	"	7.7	89.6	2.7	0.0	0.0
WAS 122-IDSA 12-WAS B-FKR 1	NERICA-L-5	TOG5681/IR64/2* IR64	"	6.2	90.1	2.2	0.0	0.0
WAS 122-IDSA 13-WAS 10-FKR 1	NERICA-L-6	TOG5681/IR64/2* IR64	"	10.8	84.8	0.0	0.0	0.0
WAS 122-IDSA 13-WAS 13-3-3 FKR 1	NERICA-L-7	TOG5681/IR64/2* IR64	"	16.2	69.7	0.0	3.2	5.1
WAS 122-IDSA 14-WAS B- FKR 1	NERICA-L-8	TOG5681/IR64/2* IR64	"	7.9	61.9	3.5	4.1	3.5
WAS 122-IDSA 10-WAS-3-1-TGR 3	NERICA-L-9	TOG5681/IR64/2* IR64	"	7.1	85.8	0.0	0.0	0.0
WAS 122-IDSA 10-WAS-7-2-FKR1-TGR 8	NERICA-L-10	TOG5681/IR64/2* IR64	"	7.7	86.1	0.9	0.0	0.0
WAS 122-IDSA 11-WAS -10-2-TGR 60	NERICA-L-11	TOG5681/IR64/2* IR64	"	7.4	89.9	0.0	0.0	0.0
WAS 122-IDSA 11-WAS -B-IER-11-19	NERICA-L-12	TOG5681/IR64/2* IR64	"	1.6	73.8	0.0	1.6	1.1
WAS 122-IDSA -13-WAS 10- WAB-B-TGR 5	NERICA-L-13	TOG5681/IR64/2* IR64	"	6.6	68.7	0.0	4.7	4.7
WAS 122-IDSA -1-WAS 2-WAB 1-TGR 6	NERICA-L-14	TOG5681/IR64/2* IR64	"	5.0	60.4	0.0	22.6	24.3
WAS 122-IDSA -1-WAS -2	NERICA-L-15	TOG5681/IR64/2* IR64	"	3.4	66.1	0.0	21.6	23.4
WAS 122-IDSA -1-WAS -2-B-1-TGR 132	NERICA-L-16	TOG5681/IR64/2* IR64	"	7.6	75.2	4.1	7.4	7.3
WAS 122-IDSA -1-WAS -2-WAB2-TGR 7	NERICA-L-17	TOG5681/IR64/2* IR64	"	1.1	52.6	0.0	10.5	10.4
WAS 122-IDSA -1-WAS -4-B-1-TGR 121	NERICA-L-18	TOG5681/IR64/2* IR64	"	10.9	87.5	0.8	0.0	0.0
WAS 122-IDSA -1-WAS -6-1 (FKR62N)	**NERICA-L-19**	**TOG5681/IR64/2* IR64**	**"**	**8.2**	**62.0**	**0.0**	**24.3**	**26.0**
WAS 122-IDSA -1-WAS -B (FKR60N and N2)	**NERICA-L-20**	**TOG5681/IR64/2* IR64**	**"**	**8.0**	**57.6**	**0.0**	**19.9**	**21.7**
WAS 122-IDSA-1-B-IER-18-6 (Niger)	**NERICA-L-49**	**TOG5681/IR64/2* IR64**	**"**	**3.4**	**57.0**	**0.0**	**22.3**	**23.5**
Mean for BC_2				**7.2**	**73.2**	**0.9**	**8.8**	**9.2**
WAS 161-B-1-1-FKR 1	NERICA-L-26	TOG5681/IR64/3* IR64	BC_3	8.8	82.0	0.0	7.6	1.6
WAS 161-B-2-B-1	NERICA-L-27	TOG5681/IR64/3* IR64	"	10.3	81.1	0.9	4.0	3.6
WAS 161-B-2-B-2	NERICA-L-28	TOG5681/IR64/3* IR64	"	8.1	82.3	0.0	0.0	9.5
WAS 161-B2-B3	NERICA-L-29	TOG5681/IR64/3* IR64	"	8.6	80.3	0.0	0.0	11.1
WAS 161-B-2-B-4	NERICA-L-30	TOG5681/IR64/3* IR64	"	7.6	85.9	0.0	3.2	3.3
WAS 161-B-4-1-FKR 1	NERICA-L-31	TOG5681/IR64/3* IR64	"	11.7	81.3	0.0	2.2	4.8
WAS 161-B-4-B-1-TGR 51	NERICA-L-32	TOG5681/IR64/3* IR64	"	8.1	87.6	0.0	0.0	4.8
WAS 161-B-4-B-2	NERICA-L-33	TOG5681/IR64/3* IR64	"	10.3	87.8	0.0	0.0	1.9
WAS 161-B--6-3-FKR 1	NERICA-L-34	TOG5681/IR64/3* IR64	"	6.0	90.4	0.0	1.6	2

Table 1. Contd.

WAS 161-B--6-4-FKR 1	NERICA-L-35	TOG5681/IR64/3* IR64	"	11.1	82.2	0.0	0.7	6
WAS 161-B-6-B1	NERICA-L-36	TOG5681/IR64/3* IR64	"	6.5	82.1	0.0	4.5	6.9
WAS 161-B-6-B-4	NERICA-L-37	TOG5681/IR64/3* IR64	"	11.1	84.1	0.0	2.2	2.6
WAS 161-B-6-B-B-1-B	NERICA-L-38	TOG5681/IR64/3* IR64	"	8.5	85.1	0.0	4.1	2.2
WAS 161-B-6-WAB-B-TGR 16 (Niger)	**NERICA-L-39**	**TOG5681/IR64/3* IR64**	**"**	**12.3**	**80.3**	**1.2**	**3.1**	**3.1**
WAS 161-B-9-1-FKR 1	NERICA-L-40	TOG5681/IR64/3* IR64	"	3.3	85.3	0.0	6.1	5.4
WAS 161-B-9-3 (FKR56N)	**NERICA-L-41**	**TOG5681/IR64/3* IR64**	**"**	**4.4**	**62.0**	**7.6**	**16.4**	**9.6**
WAS 161-IDSA-3-WAS-B-IER-2-4 (N1)	**NERICA-L-42**	**TOG5681/IR64/3* IR64**	**"**	**8.6**	**83.8**	**0.8**	**2.2**	**4.6**
Mean for BC_3				**8.5**	**82.6**	**0.6**	**3.4**	**4.9**
WAS 191-10-3-FKR 1	NERICA-L-48	IR64/TOG5681/4*IR64	BC_4	10.6	79.0	0.0	1.6	8.8
WAS 191-10-4-FKR 1-TGR 123	NERICA-L-50	IR64/TOG5681/4*IR64	"	10.1	74.2	0.0	0.0	15.6
WAS 191-10-WAB-B-TGR 23	NERICA-L-51	IR64/TOG5681/4*IR64	"	1.1	64.5	0.0	0.0	34.4
WAS 191-1-5-FKR 1	NERICA-L-52	IR64/TOG5681/4*IR64	"	7.6	84.1	0.0	0.0	8.3
WAS 191-4-10	NERICA-L-54	IR64/TOG5681/4*IR64	"	11.4	82.9	0.0	0.0	5.7
WAS 191-8-1-FKR 1	NERICA-L-55	IR64/TOG5681/4*IR64	"	7.6	90.7	0.0	0.0	1.7
WAS 191-8-3	NERICA-L-56	IR64/TOG5681/4*IR64	"	7.6	90.7	0.0	0.0	1.7
WAS 191-9-B-2	NERICA-L-57	IR64/TOG5681/4*IR64	"	13.3	53.4	4.1	5.3	23.8
WAS 191-9-WAB-B-TGR 24	NERICA-L-58	IR64/TOG5681/4*IR64	"	6.5	71.7	3.3	0.0	18.5
WAS 191-9-3- FKR-1 (FKR58N)	**NERICA-L-60**	**IR64/TOG5681/4*IR64**	**"**	**4.9**	**90.5**	**0.0**	**1.9**	**2.7**
Mean for BC_4				**8.1**	**78.2**	**0.7**	**0.9**	**12.1**

genetic similarity between the 48 lines and used to generate dendrogram using the complete-link method of SAHN clustering. Fourth, principal component analysis (PCA) was used to investigate the overall variation and patterns of relationship among the lines. Both cluster and principal component analyses were performed using NTSYS-pc for windows, version 2.0, Exeter Software (Rohlf, 1998).

RESULTS

Marker polymorphism and parent contribution

An initial polymorphism survey conducted using DNA from the two parents (IR64 and Tog5681) revealed a polymorphism level of 77.9% (109 out of the 140 SSR markers). The number of polymorphic markers per chromosome varied from 7 on chromosome 10 to 11 on chromosomes 7, and the overall average was 9.1 polymorphic markers per chromosome (data not shown). Sixty (Table 2) out of the 109 polymorphic markers were selected for genotyping the 48 interspecific lines and the average number of markers used for genotyping was 5 per chromosome.

Knowledge on the proportion of donor parent contribution among lines and chromosomes provides useful information for selection and variety development. In our study, *O. glaberrima* DNA varied from 1.1 to 16.2% of the genome at $BC_2F_{10,}$ 3.3 to 12.3% at BC_3F_8 and 1.1 to 13.3% at BC_4F_8 (Table 1). The average proportion of the genome containing *O. glaberrima* alleles at BC_2F_{10}, BC_3F_8 and BC_4F_8 was 7.2% (83.5 of 1,162.1 cM), 8.5% (99.3 of 1,162.1 cM) and 8.1% (93.8 cM), respectively. The introgression of *O. glaberrima* DNA across chromosomes was highly variable. There was no *O. glaberrima* DNA on (i) chromosomes 3, 5 and 11 at BC_2F_{10}, (ii) chromosomes 3, 4, 5, 9, 11 and 12 at BC_3F_8, and (iii) chromosomes 4, 5, 9, 11 and 12 at BC_4F_8 (Figure 1). Chromosome 6 contained the highest *O. glaberrima* DNA at all the 3 backcross generations (54.8% at BC_2F_{10}, 70.9% at BC_3F_8 and 88.1% at BC_4F_8).

O. sativa genome varied from 52.6 to 90.1% at BC_2F_{10}, from 62.0 to 90.4% at BC_3F_8 and from 53.4 to 90.7% at BC_4F_8. The average proportion of the *O. sativa* parent was 73.3% (851.3 of 1,162.1 cM) at BC_2F_{10}, 82.6% (959.5 cM) at BC_3F_8

Table 2. Genetic map distance (Temnykh et al., 2001) of the 60 SSR markers used in the present study.

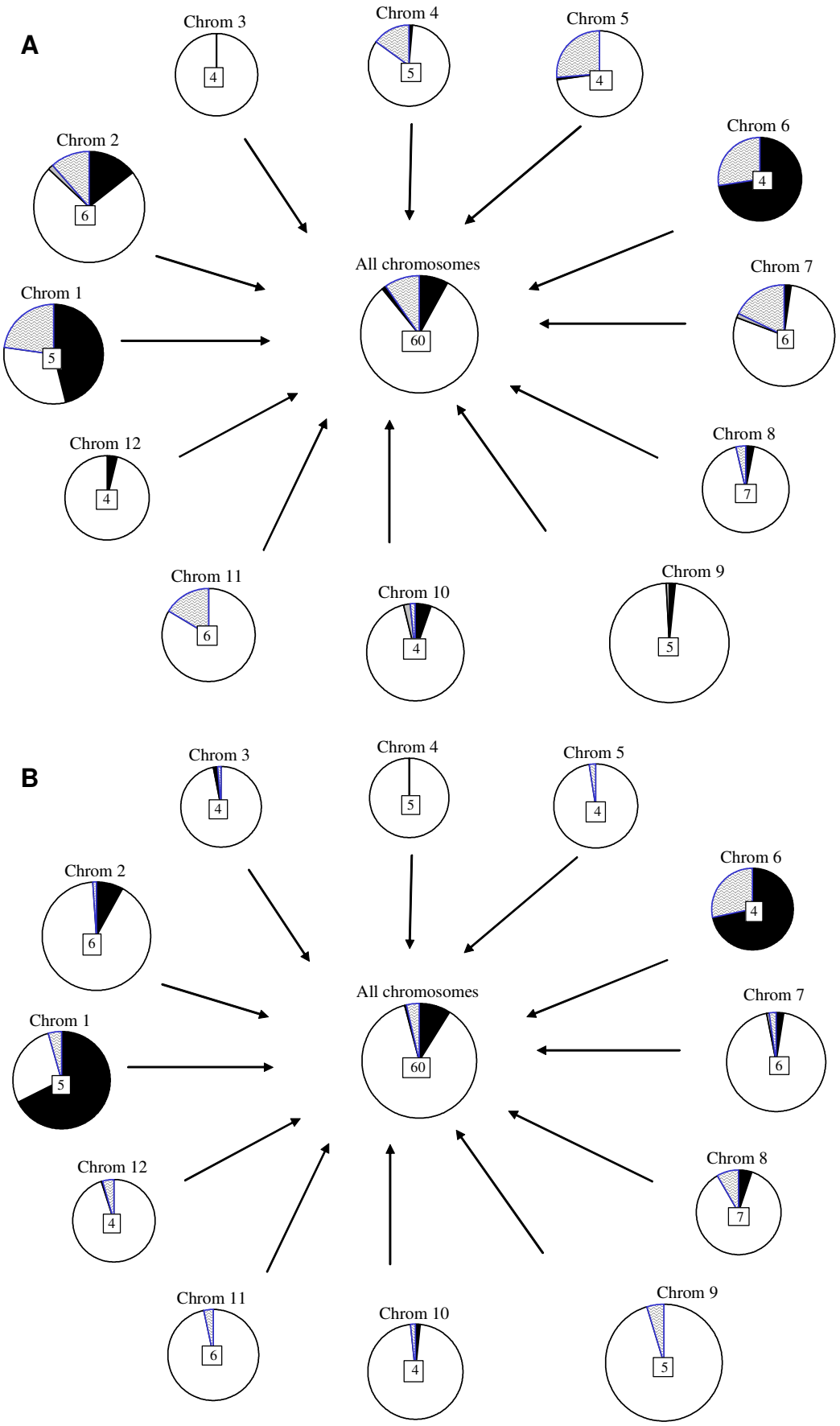
Chromosome	Marker name	Map distance (cM)	Chromosome	Marker name	Map distance (cM)
1	RM84	26.4	7	RM180	27.5
1	RM220	28.4	7	RM501	34.7
1	RM259	54.2	7	RM11	47.0
1	RM486	153.5	7	RM429	96.9
1	RM315	165.3	8	RM152	9.4
2	RM555	34.7	8	RM25	52.2
2	RM424	67.5	8	RM72	60.9
2	RM475	92.5	8	RM404	69.0
2	RM573	143.7	8	RM223	80.5
2	RM425	168.1	8	RM230	112.2
2	RM208	188.4	8	RM433	116.0
3	RM60	0.0	9	RM409	45.6
3	RM007	64.0	9	RM257	66.1
3	RM554	100.6	9	RM553	76.7
3	RM16	131.5	9	RM160	82.4
4	RM471	53.8	9	RM215	99.4
4	RM564	73.1	10	RM474	0.0
4	RM119	76.1	10	RM239	25.2
4	RM470	115.5	10	RM467	46.8
4	RM348	135.4	10	RM294	87.1
5	RM13	31.4	11	RM167	37.5
5	RM164	80.1	11	RM479	50.6
5	RM26	122.7	11	RM536	55.1
5	RM480	130.6	11	RM209	73.9
6	RM508	0.0	11	RM229	77.8
6	RM589	3.2	11	RM224	120.1
6	RM204	25.1	12	RM20	3.2
6	RM253	37.0	12	RM558	10.0
7	RM481	0.8	12	RM247	32.3
7	RM125	24.8	12	RM17	109.1

and 78.2% (908.6 cM) at BC_4F_8 (Table 1). The highest proportion of *O. sativa* alleles was observed on chromosome 9 both at BC_2F_{10} (95.2%) and BC_4F_8 (94.2%) and chromosome 5 at BC_3F_8 (97.5%). The highest frequency of heterozygosity per line was 4.2% at BC_2F_{10}, 7.6% at BC_3F_8 and 4.1% at BC_4F_8. Heterozygosity at BC_2F_{10}, BC_3F_8 and BC_4F_8 accounted for 0.9% (10.2 of 1,162.1 cM), 0.6% (7.2 cM of 1162.1 cM) and 0.7% (8.6 of 1,162.1 cM) of the total genome, respectively. A total of 5 chromosomes at BC_2F_{10}, 3 chromosomes at BC_3F_8 and 2 chromosomes at BC_4F_8 contained heterozygous loci. The frequencies of non-parental alleles among lines derived from BC_2F_{10}, BC_3F_8 and BC_4F_8 were 0-28.4, 0-16.4 and 0-5.3%, respectively. The average frequency of overall non-parental alleles per line was 8.8% (101.8 of 1,162.1 cM) at BC_2F_{10}, 3.4% (39.4 of 1,162.1 cM) at BC_3F_8 and 0.9% (10.2 of 1,162.1 cM) at BC_4F_8 (Table 1). Non-parental alleles at least in one line were observed in a total number of 9, 11 and 5 chromosomes at BC_2F_{10}, BC_3F_8 and BC_4, respectively (Figure 1). The highest proportion of non-parental alleles was observed on chromosome 5 (23.2%) at BC_2F_{10} and chromosome 6 both at BC_3F_8 (28.0%) and BC_4F_8 (10.9%).

ANOVA revealed significant differences among the 3 backcross generations only for the recurrent parent genome ($p < 0.04$) and non-parental alleles ($p < 0.02$). Pairwise comparison of means using LSD indicated the presence of significantly higher ($p = 0.011$) recurrent parent genome at BC_3F_8 than BC_2F_{10}. In contrast, the proportion of non-parental alleles at BC_2F_{10} was significantly higher than those of BC_3F_8 ($p = 0.027$) and BC_4F_8 ($p = 0.07$).

Genetic similarity and relationship

Similarity indices and patterns of relationships among


A
Chrom 3
4
Chrom 4
5
Chrom 5
4
Chrom 2
6
Chrom 6
4
All chromosomes
60
Chrom 1
5
Chrom 7
6
Chrom 12
4
Chrom 8
7
Chrom 11
6
Chrom 10
4
Chrom 9
5
B
Chrom 3
4
Chrom 4
5
Chrom 5
4
Chrom 2
6
Chrom 6
4
All chromosomes
60
Chrom 1
5
Chrom 7
6
Chrom 12
4
Chrom 8
7
Chrom 11
6
Chrom 10
4
Chrom 9
5

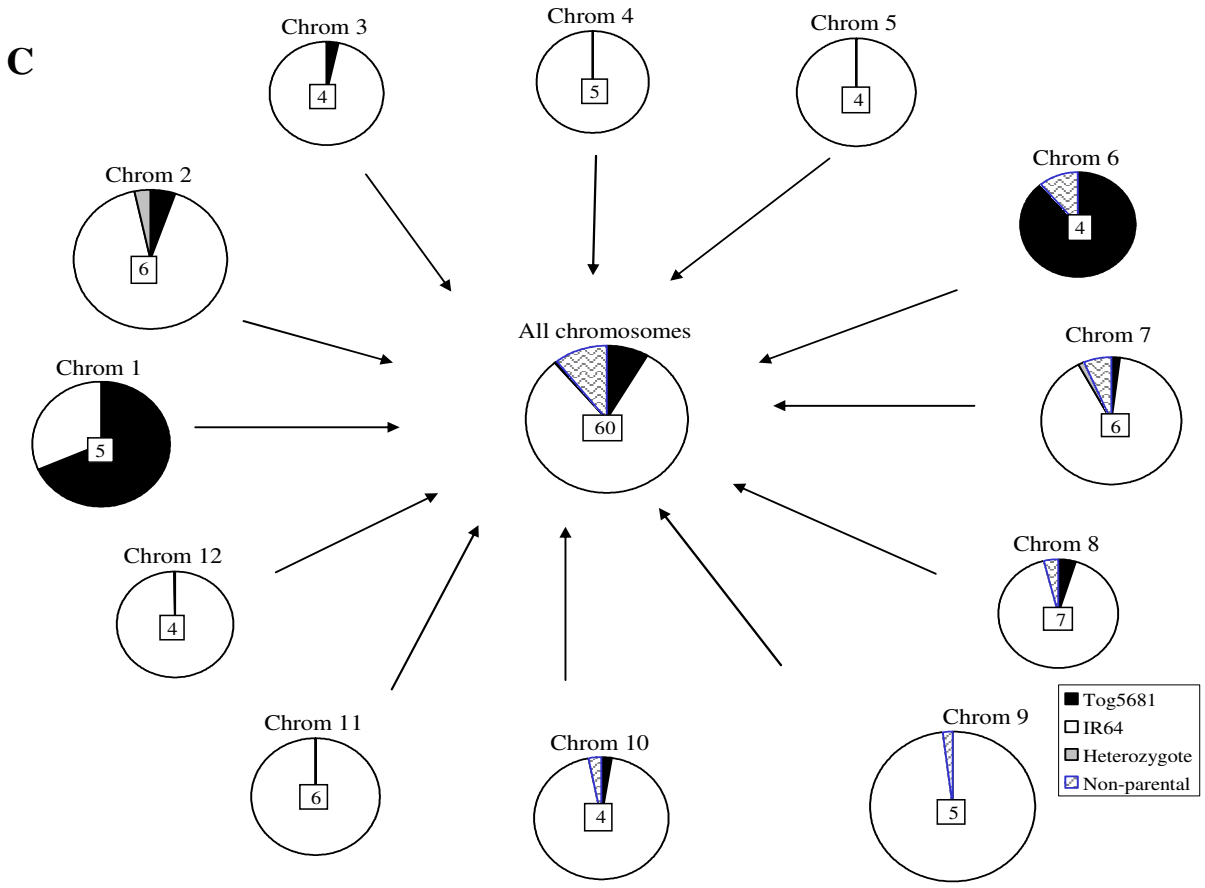


Figure 1. Pie charts for 12 rice chromosomes depicting the proportion of genome introgression among (a) 21 BC_2F_{10}, (b) 17 BC_3F_8, and (c) 10 BC_4F_8 interspecific lowland rice population derived from Tog5681 (donor) and IR64 (recurrent) parents. The pie charts were plotted from the graphical genotyping analyses outputs. Numbers in the center of the pies correspond to the number of SSR markers used in the study.

lines from cluster and Principal component analyses are useful to evaluate the potential breeding value of the lines that have not yet been released as variety. The lowest genetic similarly (55.6%) was obtained between NERICA-L-2 and NERICA-L-57. The highest genetic similarly was 100.0% and it was observed between 9 pairs of lines: NERICA-L-11 and NERICA-L-12, NERICA-L-28 and NERICA-L-32, NERICA-L-28 and NERICA-L-51, NERICA-L-28 and NERICA-L-56, NERICA-L-31 and NERICA-L-37, NERICA-L-51 and NERICA-L-52, NERICA-L-51 and NERICA-L-55, NERICA-L-51 and NERICA-L-56, and NERICA-L-55 and NERICA-L-56 (data not shown). These similarities were also evident in the cluster analysis performed using the simple matching coefficients derived from SSR markers. The phenogram produced two major groups and six sub-groups (Figure 2). All except the first sub-group were represented by at least 1 released variety. The first five principal components (PCs) from principal component analysis explained 54.4% of the variations. A plot of PC1 (22.9%) and PC2 (9.5%) from the principal component analysis revealed the two major groups (Figure 3) in same way as the cluster analysis. There were, however, two differences between the cluster and PCA: (i) the six sub-groups observed in the cluster analysis were not evident in the PCA; and (ii) NERICA-L-17 in group 2 and NERICA-L-3, NERICA-L-39 and NERICA-L-41 in group-1 appeared to be distant from all others within the same group.

DISCUSSION

Hundreds of microsatellite markers have been incorporated into rice genetic maps constructed using both intraspecific and interspecific populations (Chen et al., 1997; Cho et al., 2000; Lorieux et al., 2000; Temnykh et al., 2001). For many studies, large numbers of markers are not required. Rather, a well-distributed, highly informative and robust set of markers would be of particular value. Most of the microsatellite markers used in the present study were well-distributed along the rice chromosomes. However, there were a few intervals with uneven distribution of markers, primarily due to the lack of polymorphic markers within those intervals.

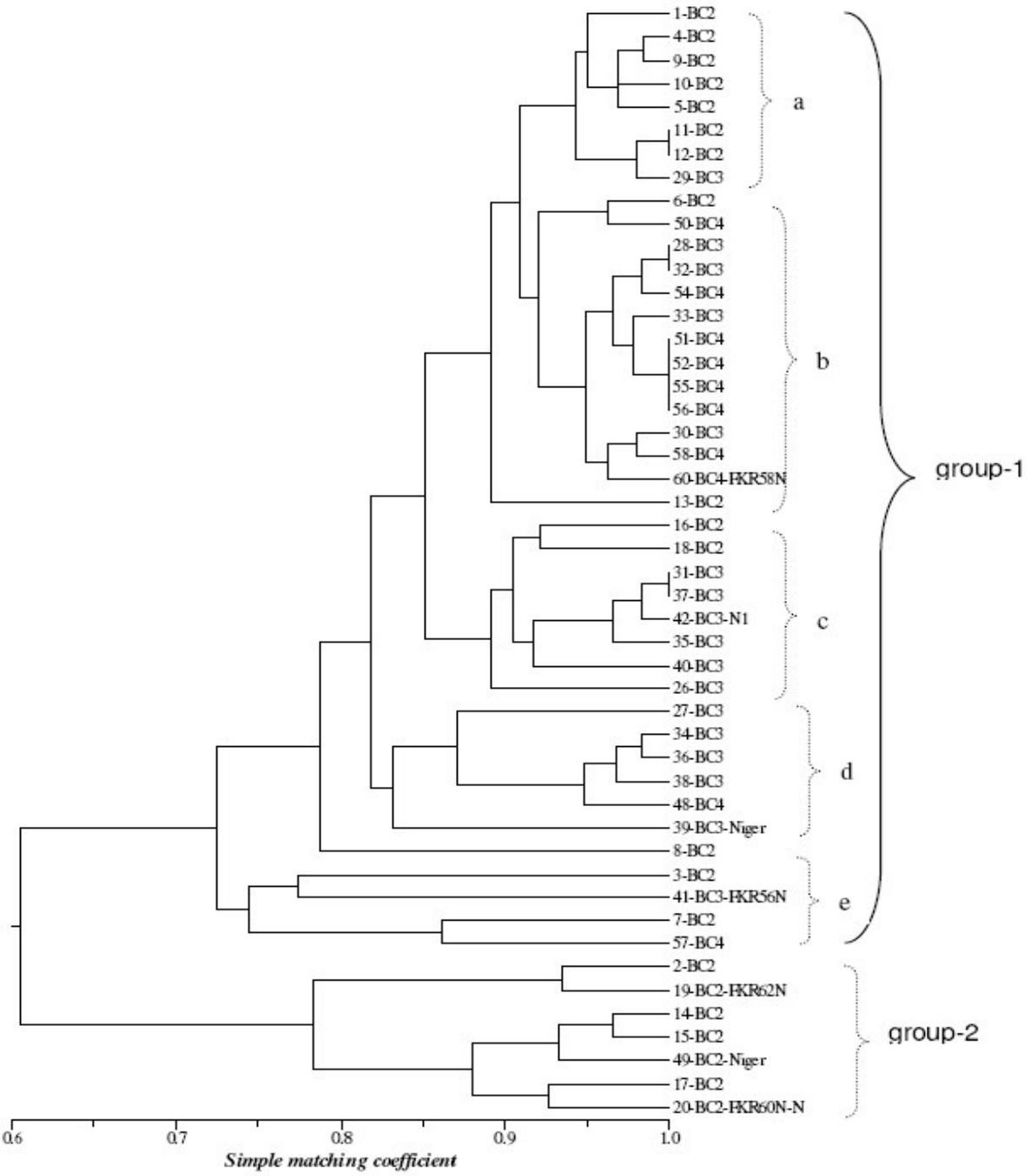


Figure 2. Dendrogram of the 48 interspecific lines using simple matching coefficient derived from 60 microsatellite markers. Each line is named with a prefix number to indicate WARDA designation as shown in Table 1 followed by level of backcross (BC_2, BC_3 or BC_4). The 7 released varieties are indicated either with their variety name or country of release as suffix; FKR56N, FKR58N, FKR60N, and FKR62N all are released in Burkina Faso; two varieties with N1 and N2 as suffix are released in Mali.

Molecular markers provide a useful means for estimating parental contributions to inbred progeny (Lorenzen et al., 1995; Visscher, 1996; Bernardo et al., 1997, 2000; Heckenberger et al., 2005a; Frisch and Melchinger, 2006; Semagn et al., 2007). Lorenzen et al. (1995), for example, used restriction fragment length polymorphism (RFLP) markers to detect significant differences between estimated and expected parental contributions in 4 out of 24 soybean (*Glycine max*) varieties. Bernardo et al. (1997) used RFLP markers to determine the frequency and magnitude of deviations from the expected parental contribution among F_2- and BC_1-derived maize (*Zea*

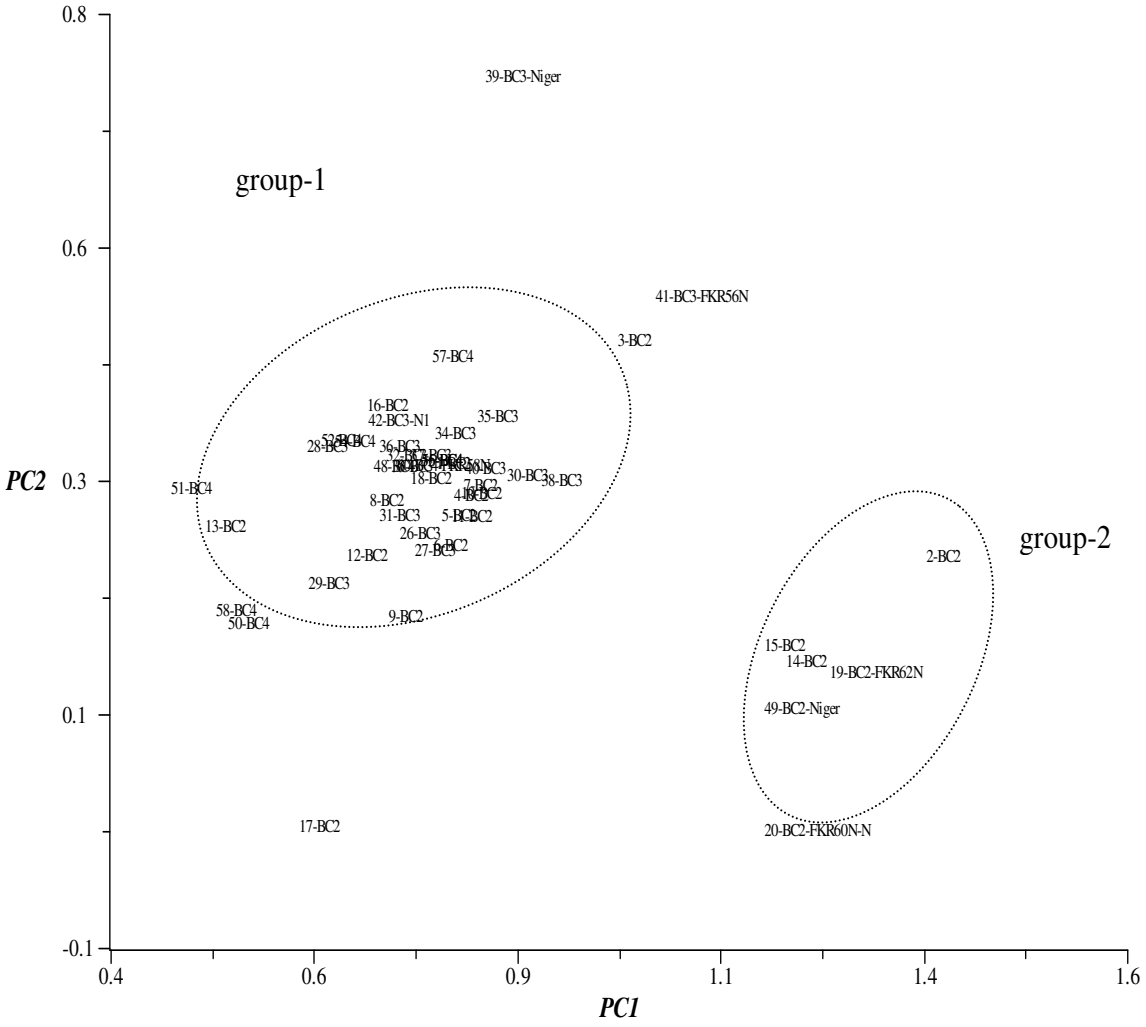


Figure 3. Principal component analysis of the 48 interspecific lines genotyped with 60 microsatellite markers. Each line is named with a prefix number to indicate WARDA designation as shown in Table 1 followed by level of backcross (BC_2, BC_3 or BC_4). The 7 released varieties are indicated either with their variety name or country of release as suffix; FKR56N, FKR58N, FKR60N, and FKR62N all are released in Burkina Faso; two varieties with N1 and N2 as suffix are released in Mali.

mays) inbreds. Among 34 BC-derived maize inbreds, eight had estimates that deviated significantly from the expected contribution of 75% from the recurrent parent. Bernardo et al. (2000) reported the superiority of SSR markers over RFLP markers in estimating parental contribution. In the present study, the average proportions of genome containing the *O. glabberima* parent at BC_2, BC_3 and BC_4 were 7.2, 8.5 and 8.1%, respectively, and were not statistically different. However, the *O. glabberima* genome at BC_2 was significantly ($p < 0.05$) lower than what would be expected (12.5%) for a random set of BC_2 lines with Mendelian inheritance while those of BC_3 and BC_4 showed significantly higher *O. glaberrima* genome than the expected (6.25% at BC_3 and 3.13% at BC_4).

In another study, Semagn et al. (2007) used microsatellite markers for estimating the contribution of each parent among 70 BC_2 interspecific lines developed for upland climatic conditions. The average recurrent parent (*O. sativa* variety WAB56-104) genome in the upland interspecific lines was consistent with what was expected for a random set of BC_2 lines (87.5%) but the average proportion of introgressed *O. glaberrima* genome (6.3%) was half of what would be expected at BC_2 generation (12.5%). Therefore, the estimated donor parent genome at BC_2 in the present study is in agreement with Semagn et al. (2007). However, *O. glaberrima* genome both at BC_3 and BC_4 as well as the *O. sativa* genome at all 3 backcross generations disagrees with Semagn et al. (2007). About 82% of the BC_3 lines and 90% of the BC_4 lines (Table 1) showed an introgression higher than expected for random lines at the same generation. On the other hand, three BC_2F_{10} lines and none of the BC_3F_8 and BC_4F_8 lines contained *O. sativa* alleles greater than the expected value at the same generation. There are

two possible explanations for such deviations between the estimated and expected parental contribution. First, intensive selection during selfing at both BC_3 and BC_4 generations might have been done in favor of the *O. glaberrima* parent for a number of traits which could accumulate several genes from the donor parent. Second, selection and genetic drift during inbreeding might have caused differences between the actual and expected proportions of the genome derived by an inbred from each of its parents (St. Martin, 1982; Lorenzen et al., 1995; Visscher, 1996; Bernardo et al., 1997, 2000; Heckenberger et al., 2005b; Frisch and Melchinger, 2006). The extent of recovery of recurrent parent genome may depend on the difference between the agronomic performances of the starting donor and recipient parents. The donor *O. glaberrima* parent was of poor agronomic performance and only used to introduce a particular trait to the *O. sativa* parent. In wide crosses and/or with undesirable linkages, therefore, a greater number of backcrosses may be necessary to increase the recovery of the recurrent parent genome other than the target regions. Usually 6 backcrosses with selection for type in the early generations have proved sufficient although Young and Tanksley (1989) found one cultivar developed after 11 backcrosses still containing the entire chromosome arm carrying the gene from the donor parent, which is the case on chromosome 6 in this study.

The average genome containing non-parental alleles per line was the lowest at BC_4F_8 (0.9%) and the highest at BC_2F_{10} (8.8) while those of the BC_3F_8 were intermediate (3.4%; Table 1). Non-parental alleles have been reported in other studies (Smith et al., 1997; Bernardo et al., 2000; Semagn et al., 2007; Liu et al., 2008). Smith et al. (1997) gave several possible reasons for the existence of non-parental bands (alleles). First, residual heterozygosity may be present in the parental inbreds. A problem with marker analysis of inbreds and their parents is that DNA samples are taken from plants grown from remnant seed stocks. Ideally, DNA samples should be taken from the actual plants used to make the original cross from which an inbred was developed. Residual heterozygosity may therefore cause marker genotypes to differ between remnant seed stocks and the actual plants used to make the cross. Second, contamination by stray pollen may have occurred during inbred development. Third, seed stocks of inbreds may have changed genetically over time though spontaneous mutations at SSR loci or physical mixing with seed from another inbred parent or undocumented outcrossing during generation advance. However, the reduction in non-parental alleles with increasing number of backcross generations in this study suggests that contamination by stray pollen at the early generation of selfing as the main cause for non-parental alleles in this study. The highest frequencies of non-parental alleles were observed on chromosome 5 at BC_2 and chromosome 6 both at BC_3 and BC_4 (Figure 1). The reason for the high frequency of non-parental alleles on chromosome 5 is unclear but chromosome 6 is known to contain a sporo-gametophytic sterility gene in the vicinity of the waxy starch synthase gene (Sano, 1990; Lorieux et al., 2000; Heuer and Miezan, 2003). Sterility association with this locus could affect the production of viable pollen within a line and would tend to promote outcrossing.

The 3 released lowland NERICA (FKR60N, FKR62N and NERICA-L-49) were selected among lines that belong to group-2 (Figures 2 and 3). All the other five subgroups in the cluster analysis that consisted of a total of 41 lines were represented only by 4 released varieties, suggesting that the other lines may provide an opportunity for selection and additional varietal development. However, the possibility for further selection and varietal development within group-1 will be highly dependent on the availability of reliable morpho-agronomic data from both multi-location trials and participatory varietal selections. This study is the first attempt to characterize the introgression of chromosomal segments among interspecific lines developed for the lowland climatic conditions in Africa and provides valuable information for breeders.

ACKNOWLEDGEMENT

This work was supported by USAID project.

REFERENCES

Alam M (1988). Evaluation of rice cultivars for resistance to *Diopsis longicornis* (Diptera: Diopsidae). J. Econ. Entomol. 81: 934-936.

Attere A, Fatokun C (1983). Reaction of *O. glaberrima* accessions to rice yellow mottle virus. Plant Dis. 67: 420-421.

Attia T, Busso C, Robbelen G (1987). Digenomic triploids for an assessment of chromosome relationships in the cultivated diploid *Brassica* species. Genome 29: 326-330

Bernardo R, Murigneux A, Maisonneuve JP, Johnsson C, Karaman Z (1997). RFLP-based estimates of parental contribution to F_2- and BC_1-derived maize inbreds. Theor. Appl. Genet. 94: 652-656.

Bernardo R, Romero-Severson J, Ziegle J, Hauser J, Joe L, Hookstra G, Doerge RW (2000). Parental contribution and coefficient of coancestry among maize inbreds: pedigree, RFLP, and SSR data. Theor. Appl. Genet. 100: 552-556.

Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998). Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. Theor. Appl. Genet. 97: 816-827.

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. 100: 563-567.

Cho YG, Ishii T, McCouch SR (2000). Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100: 697-712.

Choudhary BR, Joshi P (2001). Genetic diversity in advanced derivatives of *Brassica* interspecific hybrids. Euphytica 121: 1-7.

Choudhary BR, Joshi P, Ramarao S (2000). Interspecific hybridization between *Brassica carinata* and *B. rapa*. Plant Breed. 119: 417-420.

Fofana B, Koupeur T, Jones MP, Johnson DE (1995). The development of rice varieties competitive with weeds. Proceedings of Brighton Crop Protection Conference: Weeds, Brighton, UK, pp. 187-192.

Frisch M, Melchinger AE (2006). Marker-based prediction of the parental genome contribution to inbred lines derived from biparental crosses. Genetics, 174: 795-803.

Heckenberger M, Bohn M, Frisch M, Maurer HP, Melchinger AE (2005a). Identification of essentially derived varieties with molecular markers: an approach based on statistical test theory and computer

simulations. Theor. Appl. Genet. 111: 598-560.
Heckenberger M, Bohn M, Melchinger AE (2005b). Identification of essentially derived varieties obtained from biparental crosses of homozygous lines. I. SSR data from maize inbreds. Crop Sci. 45: 1132-1140.
Heuer S, Miezan KM (2003). Assessing hybrid sterility in *Oryza glaberrima* x *O. sativa* hybrid progenies by PCR marker analysis and crossing with wide compatibility varieties. Theor. Appl. Genet. 107: 902-909.
International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. Nature, 436: 793-800.
Jennings PR, Coffman WR, Kauffman HE (1979). Rice improvement. International Rice Research Institute, Los Banos, Laguna, Philippines.
John V, Thottapilly G, Ng N, Allury K, Gibbons J (1985). Varietal reaction to rice yellow mottle virus resistance. FAO Plant Prot. Bull 33: 109-111.
Jones M, Heinrichs E, Johnson D, Riches C (1994). Characterization and utilization of *Oryza glaberrima* in the upland rice breeding programmed. WARDA Annual report 1993, pp 3-13, WARDA, Côte d'Ivoire.
Jones MP, Dingkuhn M, Aluko GK, Semon M (1997). Interspecific *Oryza sativa* L. x *O. glaberrima* Steud. progenies in upland rice improvement. Euphytica, 92: 237-246.
Katiyar RK, Chamola R, Chopra VL (1998). Tetralocular mustard, Brassica juncea: New promising variability through interspecific hybridization. Plant Breed. 117: 398-399.
Liu G, Bernhardt JL, Jia MH, Wamishe YA, Jia Y (2008). Molecular characterization of the recombinant inbred line population derived from a japonica-indica rice cross. Euphytica 159: 73-82.
Lorenzen LL, Boutin S, Young N, Specht JE, Shoemaker RC (1995). Soybean pedigree analysis using map-based molecular markers. I. Tracking RFLP markers in cultivars. Crop Sci. 35: 1326-1336.
Lorieux M, Ndjiondjop MN, Ghesquière A (2000). A first interspecific *Oryza sativa* x *O. glaberrima* microsatellite-based genetic linkage map. Theor. Appl. Genet. 100: 593-601.
McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Magirang R, Li Z, Xing Y, Zhang Q, Kono I, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res. 9: 199-207.
Mizushima U (1980). Genome analysis in Brassica and allied genera. In: Tsunoda S, Hinata K, Gomez-Campo C (eds) Brassica crops and wild allies, Japan Sci Soc Press, Tokyo, pp. 89-106.
Morgante M, Olivieri AM (1993). Hypervariable microsatellites in plants. Plant J. 3: 175-182.
Ndjiondjop MN, Albar L, Fargette D, Fauquet C, Ghesquière A (1999). The genetics basis of the high resistance to RYMV in two cultivated rice *Oryza sativa* and *O. glaberrima*. Plant Dis. 83: 931-935.
Ndjiondjop MN, Semagn K, Cissoko M, Tsunematsu H, Jones MP (2006). Genetic relationships among rice varieties based on expressed sequence tags and microsatellite markers. Asian J. Plant Sci. 5: 429-437.
Olsson G (1960). Species crosses within the genus *Brassica*. I. Artificial *Brassica juncea* Coss. Hereditas 46: 171-223.
Rafalski JA, Tingey SV (1993). Genetic diagnostics in plant breeding: RAPD's, microsatellites and machines. Theor Appl Genet 8:275-280
Reversat G, Destombes D (1995). Resistance to *Heterodera sacchari* in rice. Nematologica 41: 333-334.
Rohlf FJ (1998). NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System. Version 2.0, Exeter software, New York.
Roy NN (1984). Interspecific transfer of *Brassica juncea* type high blackleg resistance to *B. napus*. Euphytica 33: 295-303.
Saghai-Maroof MA, Soliman K, Jorgensen RA, Allard RW (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Nat Acad Sci (USA) 81: 8014-8018.
Sano Y (1990). The genic nature of gamete eliminator in rice. Genetics 125: 161-191.
Sano Y, Sano R, Morishima H (1984). Neighbour effects between two occurring rice species, *Oryza sativa* and *O. glaberrima*. J Appl Ecol 21: 245-254.
Sauphanor B (1985). Some factors of upland rice tolerance to stem borers in West Africa. Insect Sci. Appl. 6: 429-434.
Semagn K, Ndjiondjop MN, Lorieux M, Cissoko M, Jones M, Mccouch S (2007). Molecular profiling of an interspecific rice population derived from a cross between WAB 56-104 (*Oryza sativa*) and CG 14 (*Oryza glaberrima*). Afr. J. Biotechnol. 6: 2014-2022.
Sharma PC, Winter P, Bünger T, Hüttel B, Weingand F, Weising K, Kahl G (1995). Abundance and polymorphism of di, tri and tetranucleotide tandem repeats in chickpea (*Cicer arietinum* L.). Theor. Appl. Genet. 90: 90-96.
Silue D, Notteghem J (1991). Resistance of 99 *Oryza glaberrima* varieties to blast. Int. Rice Res. News 16: 13-14.
Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997). An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree. Theor. Appl. Genet. 95: 163-173.
St. Martin SK (1982). Effective population size for the soybean improvement program in maturity groups 00 to IV. Crop Sci. 22: 151-152.
Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch SR (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.
Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000). Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet 100: 697-712.
Van Berloo R (1999). GGT: Software for the display of graphical genotypes. J. Hered. 90: 328-329.
Visscher PM (1996). Proportion of the variance in genetic composition in backcrossing programs explained by molecular markers. J Hered 87: 136-138.
Young ND, Tanksley SD (1989). RFLP analysis of the size of chromosomal segments retained around the *tm-2* locus of tomato during backcross breeding. Theor. Appl. Genet. 77:.353-359.