ORIGINAL PAPER

Determination of heterotic groups for tropical *Indica* hybrid rice germplasm

**Fangming Xie · Zhizhou He · Manuel Q. Esguerra ·
Fulin Qiu · Vairamani Ramanathan**

Abstract

Key message   **Two heterotic groups and four heterotic
patterns were identified for IRRI hybrid rice germ-
plasm to develop hybrid rice in the tropics based on
SSR molecular data and field trials.**

Abstract   Information on heterotic groups and patterns is
a fundamental prerequisite for hybrid crop breeding; how-
ever, no such clear information is available for tropical
hybrid rice breeding after more than 30 years of hybrid rice
commercialization. Based on a study of genetic diversity
using molecular markers, 18 parents representing hybrid
rice populations historically developed at the International
Rice Research Institute (IRRI) were selected to form dial-
lel crosses of hybrids and were evaluated in tropical envi-
ronments. Yield, yield heterosis and combining ability
were investigated with the main objectives of (1) evaluat-
ing the magnitude of yield heterosis among marker-based
parental groups, (2) examining the consistency between
marker-based group and heterotic performance of hybrids,
and (3) identifying foundational hybrid parents in discrete
germplasm pools to provide a reference for tropical *indica*
hybrid rice breeding. Significant differences in yield, yield
heterosis and combining ability were detected among par-
ents and among hybrids. On average, the hybrids yielded
14.8 % higher than the parents. Results revealed that inter-
group hybrids yielded higher, with higher yield heterosis
than intra-group hybrids. Four heterotic patterns within
two heterotic groups based on current IRRI B- and R-line
germplasm were identified. Parents in two marker-based
groups were identified with limited breeding value among
current IRRI hybrid rice germplasm because of their low-
est contribution to heterotic hybrids. Heterotic hybrids are
significantly correlated with high-yielding parents. The
efficiency of breeding heterotic hybrids could be enhanced
using selected parents within identified marker-based heter-
otic groups. This information is useful for exploiting those
widely distributed IRRI hybrid rice parents.

Communicated by T. H. Tai.

F. Xie (✉) · Z. He · M. Q. Esguerra · F. Qiu
International Rice Research Institute, DAPO Box 7777,
Metro Manila, Philippines
e-mail: f.xie@irri.org

Z. He
Tropical Crops Genetic Resources Institute, Chinese Academy
of Tropical Agricultural Sciences, Danzhou 571737, Hainan,
China

M. Q. Esguerra
Louisiana State University, 104 M.B. Sturgis Hall,
Tower Drive, Baton Rouge, LA 70803, USA

V. Ramanathan
Pioneer Overseas Corporation, 3rd Floor, Bab Mill Center,
#6-3-1009/1100, Rajbhavan Road, Hyderabad 500082, India

Introduction

The success of hybrid rice is based on the availability of
genetic variation and efficient selection strategies for germ-
plasm that make it possible to exploit hybrid vigor com-
mercially. Accurate assessment and assignment of parental
lines into heterotic groups, defined as related or unrelated
genotypes from the same or different populations, which
display similar combining ability and heterotic response
when crossed with genotypes from other genetically dis-
tinct germplasm groups (Melchinger and Gumber 1998),

have become a prerequisite for a successful commercial hybrid breeding program as shown in maize (Melchinger and Gumber 1998; Reif et al. 2005), sunflower (Reif et al. 2013), sorghum (Menz et al. 2004), triticale (Fischer et al. 2010), and rice (Lu and Xu 2010). Traditionally, heterotic groups are evaluated by combining ability analysis, which involves a multi-environment evaluation of parents and hybrids; however, advances in molecular marker technology have made it possible to combine information on parental pedigree and field trials with molecular marker data to detect and establish heterotic groups (Melchinger 1999).

Hybrid rice has been used in commercial production for four decades. Currently, there are about 17 million hectares of hybrid rice production in China and another 4 million hectares in other countries, mainly in Bangladesh, India, Indonesia, Myanmar, the Philippines, the US and Vietnam. Heterotic rice hybrids are generally derived from distant parents by geographic origin or different ecotypes (Yuan 1977; Lin and Yuan 1980). In the earlier stage of hybrid rice development in China, two heterotic groups, that is, early season *indica* from southern China and mid- or late-season *indica* from Southeast Asia were identified for three-line hybrid rice based on wild abortive (WA) male sterile cytoplasm (Yuan 1977). More heterotic groups were studied and identified for three-line hybrids derived from other male sterile cytoplasm and for two-line hybrid rice based on thermo-sensitive genic male sterility (Wang and Lu 2007; Lu and Xu 2010). For other types of rice hybrids, however, such as tropical *indica*, and temperate and tropical *japonica*, no clear information is available for a definition of heterotic groups. Parents of tropical *indica* hybrid rice are still categorized by fertility reaction (restorer or maintainer of male sterility). The lack of a systematic study aimed at heterotic groups could be one of main reasons for observed low yield heterosis in tropical hybrid rice resulting from the unpredictable combination of parents.

The International Rice Research Institute (IRRI) has been one of the major developers of hybrid rice in the world, with many hybrid rice products (parents, breeding lines and hybrids) produced and disseminated. IRRI-bred hybrid rice germplasm has been playing a major role in hybrid rice programs in public and private organizations globally. The long-term goal of IRRI's hybrid rice breeding program is to develop and disseminate broad-based heterotic germplasm, parents and hybrids fitting multi-environments with high yield, good quality, multiple resistances and tolerance to diseases and insects, and other production-required traits. There is, therefore, a crucial need not only to understand and verify the heterotic groups and patterns of IRRI hybrid rice germplasm but also to provide a guideline reference to maximize germplasm potential useful for increasing heterosis. In a previous study, we analyzed the genetic diversity and germplasm structure of hybrid rice parents historically developed at IRRI using simple sequence repeats (SSR) and single nucleotide polymorphism (SNP) markers (He et al. 2012). Both types of marker data revealed a consistent germplasm structure with six groups within two major clusters. The objectives of this study were to (1) evaluate the magnitude of yield heterosis among parents grouped by SSR markers; (2) examine the consistency between marker-based group and heterotic performance of hybrids; and (3) identify foundational male and female hybrid parents in discrete germplasm pools and construct a set of core heterotic groups to provide a reference for tropical *indica* hybrid rice breeding.

Materials and methods

Plant materials

A sample of 18 WA-cytoplasmic three-line hybrid rice parents, including 12 maintainer (B) lines and 6 restorer (R) lines, was selected from a distance-based cluster generated from 207 SSR markers (Table 1) (He et al. 2012). The parent selection was based on lines (1) representing the original 6 groups clustered from the SSR markers; (2) covering a maximum of the allelic variation which was 59 % of the original 168 hybrid rice parents; and (3) popularity applied in tropical hybrid rice breeding and production. A diallel mating design was used among the parents without reciprocals to develop 153 F_1 hybrids in the wet season (WS) of 2009 and the dry season (DS) of 2010. Two commercial inbred varieties (IR72, IRRI 123) and three commercial hybrids (IR75217H, IR78386H and Pioneer001) were included as checks.

Field experiments

A total of 176 entries (153 hybrids, 18 parents, and 5 checks) were evaluated in field trials in five environments (year, season and location): (1) 2010WS and 2011DS, Los Baños, Philippines, 14°10′N/121°13′E; (2) 2010WS and 2011DS, General Santos, Philippines, 6°5′N/125°14′E; and (3) 2011DS, Hyderabad, India, 17°22′N/78°28′E. All entries were grown in a randomized complete block design with two replications. Forty 21-day-old seedlings were transplanted in 4 rows with 10 plants per row at spacing of 20 cm × 20 cm. Field management followed local recommendations for the two different cropping seasons. Data on days to heading were recorded at the heading stage. At ripening stage, five healthy plants in the central rows were harvested and measured for six agronomic traits: plant height, number of panicles per plant, number of spikelets per panicle, spikelet fertility, 1,000-grain weight, and grain

Table 1 Parents and their yield means (g plant^{-1}) and GCA across environments

Code	Parent	Group	Yield mean over environments*	GCA	Parent fertility category
Parent					
P04	IR64R	G2	37.2a	2.67	Restorer
P02	IR78369B	G1	33.2ab	−0.40	Maintainer
P10	IR70368B	G4	33.0abc	0.84	Maintainer
P07	IR69712-154-2-3-1-3R	G3	32.4abcd	2.42	Restorer
P09	IR73971-87-1-1-1-1R	G3	32.3abcd	1.99	Restorer
P08	IR71921-4B-B-23-2-1R	G3	31.8abcd	1.71	Restorer
P01	IR62829B	G1	29.8bcde	−1.42	Maintainer
P06	IR72998-78-1-3-2R	G2	29.7bcde	1.14	Restorer
P17	IR79156B	G6	28.6bcde	−0.16	Maintainer
P03	IR80561B	G1	28.4bcde	−0.76	Maintainer
P14	IR72795B	G5	27.8bcde	−0.49	Maintainer
P18	IR80151B	G6	27.7bcde	0.82	Maintainer
P05	IR72102-4-159-1-3-3R	G2	27.6bcde	−1.40	Restorer
P16	IR58025B	G6	27.2cde	−0.77	Maintainer
P15	IR73323B	G5	26.8de	0.06	Maintainer
P11	IR72793B	G4	25.6e	−0.94	Maintainer
P13	IR68280B	G5	24.3ef	−0.31	Maintainer
P12	V20B	G4	19.5f	−5.00	Maintainer
	Mean		29.0		
Check					
	IR78386H		41.5		Hybrid CK
	IR75217H		36.6		Hybrid CK
	Pioneer001		35.8		Hybrid CK
	IRRI 123		31.9		Inbred CK
	IR72		31.3		Inbred CK

* Different letters mean significant difference at 0.05 level by Duncan's multiple range test

yield per plant. All hybrids had normal fertilities. To focus on the main objectives, we analyzed only the data on grain yield per plant, yield heterosis and yield combining ability.

Statistical analyses

The genetic diversity between each pair of parents was measured as Cavalli-Sforza chord genetic distance (GD) (Cavalli-Sforza and Edwards 1967) using PowerMarker v3.25 (Liu and Muse 2005) as previously described (He et al. 2012). A cluster based on the C.S. chord GD matrix of the 18 parents was generated using Darwin 5 software (Perrier and Jacquemoud-Collet 2006). The following statistical model was used for the analysis of variance (ANOVA):

$$Y = \mu + E + R + G + EGI + e$$

where Y observed value of hybrid yield from each test unit; μ population mean, E environmental (Env) effect, R replication effect within each environment, G genotype (parent or F_1 hybrid) effect, EGI interaction effect between each genotype and environment, and e residual effect. The environment and hybrid were treated as fixed factors and the replication-within-environment was considered to be a random factor. The significance of environmental variance was tested against replication-within-environment entity. For all other significance tests, an experimental error term was used. ANOVA was conducted for the yield of parents and hybrids for each and across all environments using the PROC GLM procedure (SAS Institute Inc 2012). General combining ability (GCA) effects of the parents and specific combining ability (SCA) of the crosses were estimated for each and across environments following Griffing's method 2 model 1 (Griffing 1956) using a R program (R Development Core Team 2011). Yield heterosis for each hybrid was calculated as (1) mid-parent heterosis (MPH) = $100 \times (F_1 - MP)/MP$, (2) better-parent heterosis (BPH) = $100 \times (F_1 - BP)/BP$, and (3) standard heterosis over inbred check (SDHI) or over hybrid check (SDHH) = $100 \times (F_1 - CK)/CK$; where F_1 hybrid yield, MP yield mean of both parents, BP yield of the better-yielding parent, and CK yield of the check variety, either the inbred (IRRI 123) or hybrid check (IR75217H). Although five checks were included in the study and the hybrid of IR78386H yielded higher than the hybrid of

IR75217H, only IRRI 123 and IR75217H were used for calculations of SDHI and SDHH because these two varieties have been long used in commercial rice production and as standard variety checks for inbred and hybrid yield trials hosted by IRRI, the Philippines Rice Research Institute, and elsewhere. Correlations between GD estimates and means of hybrid yield, yield heterosis and combining ability were calculated using PROC CORR of SAS (SAS Institute Inc 2012).

Results

Genetic distance and groups of parents

The genetic distance measured on the C.S. chord distance from 207 SSR markers for the original 168 parents ranged from 0 to 0.7593, with an average of 0.4469, and the GD of the 18 sampling parents ranged from 0.1246 to 0.6487, with an average of 0.4694, which was slightly higher than that of the original parent population, but the samples were considered, based on the criteria of parental selection, as fairly representative of the allelic variation and the cluster structure of the original population (He et al. 2012) (Figs. 1, 2). The sampling parents were clearly clustered into six groups (G) similar to those in the original parent population with two major clusters, i.e., one with three B-line groups (G4, G5 and G6), and another one with one B-line group (G1) and two R-line groups (G2 and G3). It is noted that G1 is clustered into the R-line sub-cluster and grouped closer to the G3 R-line groups than to the G2 R-line group, and G4 is
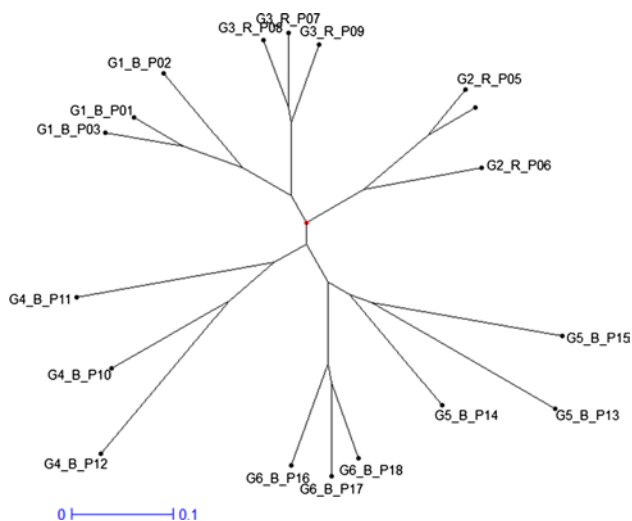

Fig. 1 Cluster of 18 WA-cytoplasmic three-line hybrid rice parents based on genetic distance calculated from 207 SSR markers. *Letter* and *number* combination refers to parent as marker-based group_fertility group_parent code
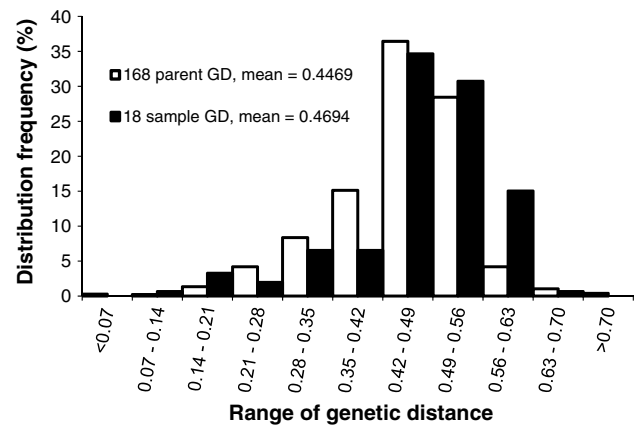

Fig. 2 Distribution of genetic distances of the original 168 and the 18 hybrid rice parents sampled for this study based on 207 SSR markers

closer genetically to the R-line groups than are the other two B-line groups (G5 and G6). Among the groups, the average GD (0.4942) of inter-groups was higher than that (0.2833) of intra-groups; however, the GDs in G4 (0.4296) and G5 (0.4145) groups had GD values similar to those in the inter-groups and were much higher than the GDs of other intra-groups (Table 2). The B × R groups had the highest allelic divergence with an average GD of 0.4868, higher than the GDs in B × B (0.4830) and R × R (0.3262) groups.

Parental and hybrid performance

ANOVA were performed to determine the statistical significance for the different sources of variation affecting parental and hybrid grain yields (Table 3). Difference of parental yielding was found insignificantly across the environments, but significantly for genotype ($P < 0.0001$) and GEI ($P < 0.01$), which accounted for 21 and 28 % of the total sums of squares, respectively. The average yield of the parents was 29.0 g plant^{-1}, ranging from 19.5 g plant^{-1} (V20B) to 37.2 g plant^{-1} (IR64R) across the environments (Table 1), and the average yield (31.8 g plant^{-1}) of R-lines was significantly ($P < 0.01$) higher than the average yield (27.7 g plant^{-1}) of B-lines.

Hybrid yields differed significantly for environment ($P < 0.01$), genotype ($P < 0.0001$), GEI ($P < 0.01$), and replication-within-environment ($P < 0.01$) with an average yield of 33.3 g plant^{-1} and a range of 21.1 to 43.6 g plant^{-1} (Table 3; Fig. 3a). The environment, genotype and GEI contributed for 23, 19 and 31 % of the total sums of squares to the variation, respectively. The environments in which the field experiments were conducted were geographically and seasonally different with diverse hybrid genotypes; thus, large effects due to environment and genotype were expected. The significant and relatively large percentage of

Table 2 Genetic distance among groups and performance of hybrid yield and yield heterosis

Hybrid category	Hybrid group	Genetic distance	Hybrid yield*	SCA	MPH	BPH	SDHI	SDHH
Summarized by hybrid group								
B × R	G3 × G5	0.5316	38.6a	3.99	34.9	19.4	22.0	8.6
B × R	G3 × G6	0.4853	36.6ab	1.68	25.2	11.2	15.4	3.2
R × R	G2 × G3	0.4082	35.7bc	0.04	15.0	2.8	12.7	1.2
B × R	G2 × G6	0.5011	35.4bcd	1.81	25.3	10.3	11.8	0.7
B × R	G2 × G5	0.5307	34.9bcd	1.57	24.4	7.0	10.4	0.9
R × R	G2 × G2	0.2298	34.8bcd	0.57	13.5	−3.8	11.5	1.0
B × R	G3 × G4	0.5043	34.0bcde	0.81	20.1	1.7	7.8	−2.0
B × R	G1 × G3	0.3719	33.6bcdef	−0.50	8.7	−1.3	5.6	−5.8
B × B	G4 × G6	0.5407	33.5bcdef	2.33	31.8	14.7	7.3	−2.2
R × R	G1 × G6	0.4806	32.9cdefg	0.82	17.3	5.0	4.0	−6.4
B × R	G1 × G2	0.4639	32.8cdefg	−0.02	7.2	−5.1	3.8	−7.1
R × R	G3 × G3	0.1758	32.7defg	−4.23	2.7	−5.9	1.3	−8.8
B × B	G1 × G5	0.5460	31.8efg	0.04	14.8	1.7	1.2	−8.6
B × B	G5 × G5	0.4145	31.7efg	−0.60	21.6	7.5	1.6	−7.1
B × B	G5 × G6	0.4461	31.2efg	−1.42	19.1	5.4	0	−8.6
B × B	G4 × G5	0.5532	31.1efg	0.23	21.4	2.7	−0.9	−10.7
B × B	G1 × G4	0.5439	31.0fg	0.72	11.2	−4.5	−1.6	−9.3
B × B	G1 × G1	0.2322	30.5g	−0.80	2.6	−7.8	−4.0	−13.5
B × B	G6 × G6	0.2170	30.2g	−2.73	17.0	4.9	−3.6	−10.5
B × R	G2 × G4	0.5058	29.9g	−1.92	6.4	−8.5	−5.8	−13.6
B × B	G4 × G4	0.4296	29.5g	0.10	15.4	−3.5	−5.8	−13.5
Summarized by parental fertility relationship								
R × R		0.3262	34.9a	−0.70	12.3 b	−0.3 a	10.2 a	−0.9 a
B × R		0.4868	34.5a	0.93	19.0 a	4.3 a	8.9 a	−1.9 a
B × B		0.4830	31.7b	0.19	18.3 a	3.5 a	0.8 b	−8.3 b
Summarized by inter- or inter-group								
Inter-group		0.4942	33.5a	0.68	18.8	4.2	6.2	−4.0
Intra-group		0.2833	21.6b	−1.28	12.1	−1.4	0.2	−8.7
Summarized by environment summarized by environment								
2010WS General Santos			39.3a		14.6	0.6	13.1	8.7
2011DS Los Baños			35.4b		13.8	2.5	11.1	0.5
2011DS General Santos			33.3c		29.5	8.9	−4.8	−28.5
2011DS Hyderabad			31.8d		10.4	0.2	16.7	15.7
2010WS Los Baños			27.0e		23.0	5.6	−8.3	−19.9

* Different letters mean significant difference at 5 % level

the total variation attributable to GEI suggests that hybrids responded differentially to environment for yield. The highest and the lowest yields of hybrids across environments were observed in the 2010WS at General Santos (39.3 g plant^{-1}) and at Los Baños (27.0 g plant^{-1}) (Table 2). Inter-group hybrids yielded significantly ($P < 0.01$) higher than intra-group hybrids. R × R and B × R hybrids yielded similarly, with average yields of 34.9 and 34.5 g plant^{-1}, respectively, and their yields were significantly ($P < 0.05$) higher than the yield (31.7 g plant^{-1}) of B × B hybrids. Among the individual hybrid groups, the highest yielding hybrid group was a B × R group (G3 × G5) which produced 38.6 g plant^{-1}, significantly ($P < 0.05$) higher than the yields in 19 of the other 20 hybrid groups. The G3 × G6 hybrid group, which is also a B × R group, was the second highest yielding group, but was only significantly higher than the yields in 12 of the other 19 hybrid groups.

Hybrid yield heterosis and combining ability

All four sources of effects, including environment, genotype, GEI and replication-within-environment, were

Table 3 Analysis of variance, including degrees of freedom (DF), mean squares (MS), and percent contribution to total sums of squares (%SS) over 5 environments for parent and hybrid yield, hybrid yield standard heterosis over inbred CK (SDHI) and hybrid CK (SDHH), and mid-parent heterosis (MPH) and better-parent heterosis (BPH)

Source	Parent yield			Hybrid Yield			SDHI		SDHH		MPH			BPH		
	DF	MS	SS (%)	DF	MS	SS (%)	SDHI	SS (%)	SDHH	SS (%)	DF	MS	SS (%)	DF	MS	SS (%)
Env	4	771	24	4	6,100**	23	3,7452	14	102,355	31	4	17,415***	4	4	3,885**	1
Genotype	17	159***	21	152	131***	19	1,285***	18	1,022***	12	152	1,746***	16	152	1,301***	16
GEI	68	54**	28	590	55***	31	542**	30	439**	20	590	1,029*	36	590	813*	38
Rep w/n Env	5	162**	6	5	201**	1	24,493***	11	49,538***	19	5	24,414***	7	5	17,163***	7
Error	85	31.2	21	712	39	28	405	27	345	19	711	873	37	709	707	39
Total	179			1,463							1,462			1,460		

*, **, *** Significant at the 0.05, 0.01 and <0.0001 probability levels, respectively

significant (from $P < 0.05$ to $P < 0.0001$) for MPH and BPH. GEI was the major variant source which contributed 36 and 38 % of the total sum of squares to the MPH and BPH variations, respectively (Table 3). The variation patterns of MPH and BPH were very similar, i.e., GEI was the major source contributing to the total sum of squares followed by genotype, and environment factor was the least affecting component for MPH and BPH. As for the yield heterosis over checks, only three major sources of variations (genotype, GEI and replication-within-environment) were statistically different (either $P < 0.01$ or $P < 0.0001$). The environment effect was not a significant contributor to the total variations of SDHI and SDHH, even though it took a relative large portion of the total variation in the sums of squares. GEI was still the most important factor affecting the heterosis over checks.

Comparing to the inbred check, the hybrids had a significant ($P < 0.05$) yield advantage of 4.3 g plant^{-1} (14.8 %) over their parents across environments, with an average of 18.1 % and a range of −29.6 to 53.2 % for MPH, and an average of 3.5 % and a range of −37.9 to 34.3 % for BPH across environments (Fig. 3b, c). High MPH and BPH were observed in the environments of the 2011DS at General Santos and the 2010WS at Los Baños (Table 2) because of the relatively low yields of the parents in these two environments. On average, the hybrids outyielded the inbred check by 5.6 %, with a range of −38.0 % to 36.4 %, but yielded less than the hybrid check by 4.5 %. The highest hybrid yield advantage over commercial varieties was observed at 2011DS, Hyderabad where the hybrids yielded 16.7 and 15.7 % more over the inbred and hybrid CKs, respectively, indicating a better environment than Los Banos and General Santos in the Philippines for hybrid rice extension. The inter-group hybrids had significantly ($P < 0.001$) higher yield and yield heterosis than the intra-group hybrids. Among the 9 hybrid groups that produced higher than the average yield (33.3 g plant^{-1}) of all experimental hybrids, one was from R × R, two were from B × B and the rest were all from B × R groups. The hybrids in the G3 × G5 group had the highest yield, yield heterosis and combining ability among the hybrid groups, followed by the G3 × G6 hybrid group. It was noted that the G4 × G6 hybrids had relatively high MPH and BPH, but they were non-competitive compared with the commercial checks. The R × R hybrids yielded high with low MPH and BPH due to their relatively high-yielding parents, and the B × R hybrids yielded insignificantly with the R × R hybrids, but significantly better than the R × R hybrids with high yield advantage over their parents and high combining ability.

Thirty-three hybrids (21.6 %) among the 153 experimental hybrids outyielded the inbred check by more than 15 %. The parents involved in those 33 heterotic hybrids
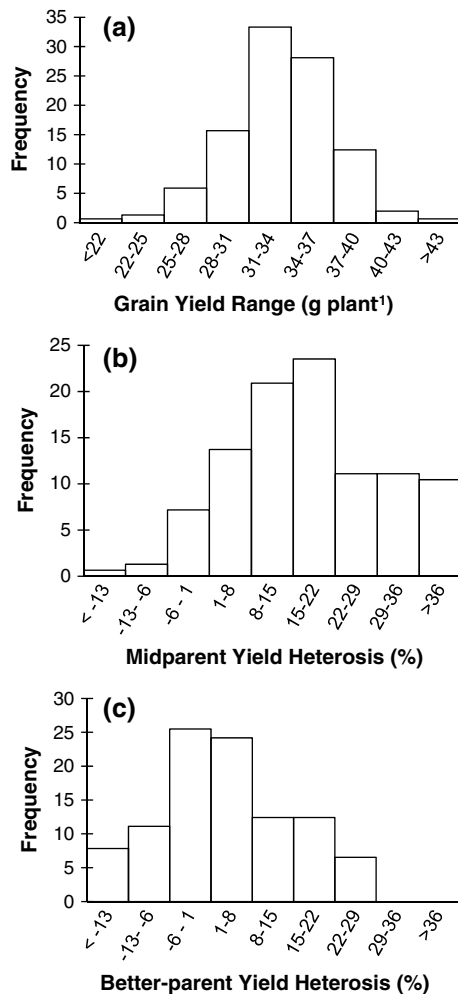
Fig. 3 Performance of hybrid grain yield (**a**), mid-parent heterosis (**b**), and better-parent heterosis (**c**)

were mainly from the groups of G3 (33.3 %), G5 (21.2 %), G2 (19.7 %), and G6 (13.6 %). The parents in the G1 and G4 groups contributed only 4.5 and 7.6 %, respectively, to those 33 heterotic hybrids, and there was no hybrid derived from the parents of IR62829B, IR70368B, and V20B contributed to those top-yielding hybrids.

The parent IR64R had the highest GCA for yield followed by IR69712-154-2-3-1-3R, and they belong to the R-line groups of G2 and G3, respectively (Table 1). With the exception of IR72102-4-159-1-3-3R, the R-lines had positive GCA for yield. Among the 12 B-lines, only 3 had positive GCA yield values. The lowest yielding GCA was the line V20B, which is in the G1 group and had the lowest parental yield. The crosses of G3 × G5, G3 × G6, G2 × G6 and G2 × G5 had high SCA values, hybrid yield and yield heterosis. On average, SCA values of the B × R were better than those of the B × B and R × R groups, and inter-group crosses were better than the intra-group crosses.

Association of hybrid yield heterosis and marker-based genetic distance

Correlations were performed for detecting association of hybrid performance with parental genetic distance based on SSR markers for each and across environments (Table 4). No significant correlations were found between GD and hybrid yield for each and across environments, but some significant correlations were found in particular environments and across environments either on all individual hybrids or on particular hybrid groups based on parental fertility categories. For all of the association tests, it was noted that GD were either insignificantly or significantly, but weakly associated with SCA, MPH and BPH. The highest correlation was the GD with MPH at 2010WS, General Santos in the B × R group ($r = 0.4221$, $P < 0.01$). However, it was observed that the correlations between GD and hybrid SCA and heterosis were increased greatly using marker-based groups rather than on the bases of individual hybrids or groups based on parent fertility categories. Across the five environments, the R^2 values of correlations of GD with SCA, MPH and BPH were increased almost twofold as compared to the R^2 values based on all individual hybrids or in the B × R group.

Discussion

Parental grouping based on molecular markers

The parents we selected were based on a study investigating genetic diversity of IRRI-bred hybrid rice parents and are considered to be a fair representation of the original parent population as they maintained the same cluster structure and similar allelic variation as those in the original population. In general, the GD among inter-group parents was higher than that among intra-group parents. However, parents of two B-line intra-groups (G4 and G5) had high genetic variation as they showed GD values similar to those of the parents in inter-groups and as shown in the cluster structure. The parents in the B × R groups had the highest genetic divergence, which is desirable for making heterotic hybrids within parental categories of current IRRI hybrid rice germplasm. The parental cluster shows that some IRRI B-lines (G1 and G4) were genetically close to the R groups (G2 and G3) (Fig. 1), revealing a high degree of genetic similarity with R-lines that resulted from insufficient attention given to genetic diversity in B-line breeding. Historically, the majority of IRRI B- and R-lines developed at earlier stages were derived directly from inbred breeding programs with many common ancestors shared and selected under the same environment with similar agronomic criteria without further breeding, which resulted in

Table 4 Correlation coefficient (r) between SSR marker-based genetic distance and hybrid yield and yield heterosis

Trait	All hybrids (n = 153)	B × B (n = 66)	B × R (n = 72)	R × R (n = 15)	Marker-based groups (n = 21)
2010WS, General Santos					
Yield	0.0603	0.1631	0.0440	0.1283	0.3262
SCA	0.2823**	0.3509*	0.1781	0.1351	0.7073**
MPH	0.3546***	0.4202**	0.4221**	0.0711	0.6305**
BPH	0.2858**	0.2703*	0.2829*	0.2150	0.6470**
2010WS, Los Baños					
Yield	0.0080	−0.1336	0.2274	0.2594	0.2314
SCA	0.1894*	0.0096	0.3709**	0.4342	0.6424**
MPH	0.1229	−0.1221	0.3648**	0.3214	0.1877
BPH	0.0881	−0.1282	0.3475**	0.2886	0.1069
2011DS, General Santos					
Yield	−0.1062	−0.0426	−0.0945	0.2830	−0.0998
SCA	0.1767*	0.1183	0.0771	0.3049	0.4202
MPH	0.1057	0.1233	0.1091	0.1372	0.3903
BPH	0.0239	0.0058	0.0433	0.1149	0.1786
2011DS, Hyderabad					
Yield	0.0635	0.2075	−0.0117	−0.2811	0.4493*
SCA	0.14683	0.3403**	0.1217	−0.2404	0.4734*
MPH	0.1571	0.2357	0.1632	−0.2934	0.3301
BPH	0.1230	0.2032	0.0988	−0.2014	0.3173
2011DS, Los Baños					
Yield	0.0729	0.1285	0.1224	0.1244	0.0337
SCA	0.1253	0.2138	0.0229	0.1639	0.1634
MPH	0.1969*	0.2062	0.1608	0.1456	0.1878
BPH	0.1577	0.1882	0.1056	0.1903	0.2137
Five environments combined					
Yield	0.0166	0.1021	0.0793	0.2205	0.1931
SCA	0.3261***	0.3936**	0.2446**	0.2841	0.6487**
MPH	0.3157***	0.2807*	0.3761**	0.2545	0.5525**
BPH	0.2318**	0.1742	0.2755*	0.3608	0.4281*

*, **, *** Significant at the $P < 0.05$, < 0.01 and < 0.0001 probability levels, respectively

a relatively high genetic uniformity among hybrid rice parents. This could be one of the reasons for the lower hybrid rice heterosis observed in the tropics than in the subtropics. In the last a few years, this issue has been addressed by separating the hybrid breeding from inbred breeding programs and by developing B- and R-line heterotic groups individually to maximize genetic diversity among hybrid rice parents.

Yield and yield heterosis

It is noteworthy that significant effects were detected for GEI in both parent and hybrid yields and heterosis due to their differential responses to environments modulated by various changes in climate, soil, and cultural practices across locations and seasons. It is an important component in hybrid rice breeding to evaluate GEI effect for

developing products adapted to different environments. It is difficult to develop a "universal high-yielding or high heterotic" hybrid for different environments in the tropics because of a more divergent and fluctuating environment as compared to the subtropics. The successful story of Shanyou 63 hybrid rice, which occupied about 50 % of 17 million hybrid rice hectares annually in China during the 1990s to 2000s, could never happen in the tropics. Some varieties have relatively high and consistent yield, or combining ability across environments, such as IR64R, because of their wide suitability in tropical environments. This kind of germplasm is suitable for developing hybrid rice products widely. Some varieties, however, always perform poorly in the tropics, especially those imported from the subtropical and temperate regions. Use of these parents directly in hybrid rice development in the tropics is questionable, even though they are elite parents in a particular cropping

region, such as V20B which was a widely used prominent hybrid rice parent in China from the 1980s to 2000s. A successful parent must have some degree of adaptedness to the area where the hybrid is grown (Troyer 2006). High yield potential and high combining ability of a parent could fail to meet the expectation of hybrid heterosis once the parent is moved to a poorly adapted cropping environment.

On average, our hybrids produced 14.8 % higher yield than their parents, with a wide range of variation in yield and heterosis, indicating a good opportunity to select heterotic hybrids for production. Based on an acceptable 15 % of SDHI (a general cutoff of hybrid rice yield advantage over an inbred rice variety for commercial production), there is a 21.6 % chance of our experimental hybrids being qualified as commercial hybrids without consideration of seed production, grain quality, and other factors. The opportunity for success is even higher if high-yielding parents are selected because of high and significant correlations between hybrid yield and MPH ($r = 0.6179$, $P < 0.0001$) and BPH ($r = 0.6607$, $P < 0.0001$) as revealed in this study.

Heterotic groups based on markers

Genetic diversity estimates are helpful in classifying germplasm into heterotic groups for hybrid crop breeding (Menz et al. 2004). Previously, the relative performance of inbred lines of known origin and pedigree was commonly used, which largely relies on breeders' empirical experience, to combine parents from different genetic backgrounds to develop heterotic hybrids. Molecular markers have been used in rice to assess the genetic relationships of rice ecotypes or sub-species (Garris et al. 2005; McNally et al. 2009; Zhao et al. 2010; Ali et al. 2011; Thomson et al. 2012) and hybrid rice parents (Xu et al. 2002); however, information is scarce on assessing heterotic groups among tropical rice inbred lines and populations, and no conclusive study has been conducted to clearly defined heterotic groups of tropical hybrid rice parents. Many of those studies investigating genetic diversity in rice with molecular markers were dealing with large pools of sub-species or ecotypes, such as *aus*, *indica*, aromatic, temperate *japonica*, and tropical *japonica* from rice germplasm collections, but with limited value to practical hybrid rice breeding due to the inability to produce yield heterosis owing, for example, to mass vegetative growth and partial fertility in hybrids between sub-species. Heterotic groups that are applied in breeding and production are different from the parental groups generated from germplasm collections based on molecular markers. It is still a challenge to find agreement between high, producible yield heterosis and high divergence among rice sub-species or ecotypes.

Our study generally agrees with the conclusion derived from many studies of hybrid crops on the association of marker-based GD and hybrid performance that the correlation of the two was too small to be used for predicting hybrid performance (Dudley et al. 1991; Xie 1993, Zhang et al. 1995; Saghai Maroof et al. 1997; Zhao et al. 2010). However, our present results showed that the association and prediction could be enhanced when parental groups are formed first by molecular markers, which may not predict the best hybrid combination, but it reveals a practical value of assigning existing and new hybrid rice germplasm into heterotic groups and increasing opportunities to develop desirable hybrids from the best heterotic groups, which is consistent with a previous study in maize (Lanza et al. 1997).

Heterotic groups in tropical *indica* hybrid rice

Previous studies related to heterotic groups in hybrid rice (Yuan 1977; Liu et al. 2002; Wang and Lu 2006a, b; Wang and Lu 2007; Xu et al. 2002) drew the following general conclusions: (1) two major heterotic groups in *indica* WA three-line hybrid rice, i.e., early season *indica* varieties from central and southern China and *indica* varieties from Southeast Asia, mostly from IRRI; (2) two major groups as represented by Xie Qinzhao and Zhenshan97 in the female pool, and two major groups as represented by IR24 and IR26, and some other breeding lines with IRRI variety ancestors, such as Ming Hui 63 (an IR30 offspring) in the male pool; and (3) R-lines are more divergent than B-lines (in China). However, a previous study showed that all WA R-lines used in China could be considered as one single group because they shared many ancestors and clustered closely together (Xie et al. 2012) with similar heterotic response to WA-based female parents. Our present study also shows that IRRI-bred B-lines are more divergent genetically than the R-lines.

Opportunities are generally high to obtain superior hybrids derived from parents from an inter-population rather than parents from an intra-population. All but one of the 33 heterotic hybrids identified in this study were from inter-population crosses. Hybrids with high MPH or BPH are not necessarily commercially competitive because of low-yielding parents, such as the hybrid of V20B × IR80151B, which had the highest MPH and BPH, but low SDHI (0.3 %) and SDHH (9.1 %). The ultimate goal of hybrid rice breeding is to produce a hybrid with high yield and high heterosis over the parents and commercial checks.

For the tropical hybrid rice parents generated from IRRI, two heterotic groups could be classified based on current B- and R-line germplasm categories, as represented by parents in G5 and G6 as females and parents in G2 and G3 as males (Table S1). The parents with low or no possibility of producing heterotic hybrids, such as those in the G1 and G4 groups,

have limited value for developing heterotic hybrids within current IRRI hybrid rice germplasm and they have to be further improved to combine with germplasm from other possible heterotic pools or to be changed to diverge from current R-line pools. G5 and G6 B-line groups combined with G2 and G3 R-line groups are our preferred choices of hybrids for current IRRI-bred hybrid rice germplasm. On average, the hybrids derived from these four crossing patterns produced 27.4, 11.9, and 14.9 % of MPH, BPH, and SDHI, respectively. We also tracked the pedigrees of all 11 IRRI-bred hybrids released in the Philippines for commercial production, and found that 6 parents were from G2 and G3 (36 %), 10 parents were from G5 and G6 (59 %), and only 1 parent was from G1, but none from G4. It should be noted that the core set of parents from this study generally fits tropical Asian environments based on current IRRI hybrid rice parents. Further heterotic groups could be changed or enhanced when new parents with new traits/germplasm are integrated and adapted to the targeted cropping region.

References

Ali ML, McClung AM, Jia MH, Kimball JA, McCouch SR, Eizenga GC (2011) A rice diversity panel evaluated for genetic and agro-morphological diversity between subpopulations and its geographic distribution. Crop Sci 51:2021–2035

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J Human Genet 19:233–257

Dudley JW, Saghai Maroof MA, Rufener GK (1991) Molecular markers and grouping of parents in maize breeding programs. Crop Sci 31:718–723

Fischer S, Maurer HP, Würschum T, Möhring J, Piepho HP, Schön CC, Thiemt EM, Dhillon BS, Weissmann EA, Melchinger AE, Reif JC (2010) Development of heterotic groups in *Triticale*. Crop Sci 50:584–590

Garris AJ, Tai TH, Coburn JR, Kresovich S, McCouch SR (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing system. Aust J Biol Sci 9:463–493

He ZZ, Xie FM, Chen LY, Dela Paz MA (2012) Genetic diversity of tropical hybrid rice germplasm measured by molecular markers. Rice Sci 19:193–201

Lanza LLB, Souza CL, Ottoboni LMM, Vieira MLC, Souza AP (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor Appl Genet 94:1023–1030

Lin SC, Yuan LP (1980) Hybrid rice breeding in China. In: Innovative approaches to rice breeding. IRRI, Philippines, pp 35–51

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Liu S, Chen H, Wang F, Zhu YG (2002) DNA polymorphism of main restorer lines of hybrid rice in China. Chin J Rice Sci 16:1–5

Lu ZM, Xu B (2010) On significance of heterotic group theory in hybrid rice breeding. Rice Sci 17:94–98

McNally KL, Childs KL, Bohnert R, Davidsond RM, Zhao K, Ulata VJ, Zeller G, Clark RM, Hoeng DR, Bureaug TE, Stokowski R, Ballinger DG, Frazer KA, Cox DR, Padhukasahasrame B, Bustamante CD, Weigel DW, Mackill DJ, Bruskiewich RM, Ratsch G, Buell RC, Leung H, Leach JE (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106:12273–12278

Melchinger AE (1999) Genetic diversity and heterosis. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. ASA, CSSA, and SSSA, Madison, pp 99–118

Melchinger AE, Gumber RK (1998) Overview of heterosis and heterotic groups in agronomic crops. In: Lamkey KR, Staub JE (eds) Concepts and breeding of heterosis in crop plants. CSSA, Madison, pp 29–44

Menz MA, Klein RR, Unruh NC, Rooney WL, Klein PE, Mullet JE (2004) Genetic diversity of public inbreds of Sorghum determined by mapped AFLP and SSR markers. Crop Sci 44:1236–1244

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/darwin

Reif JC, Hallauer AR, Melchinger AE (2005) Heterosis and heterotic patterns in maize. Maydica 50:215–223

Reif JC, Hao YS, Wurschum T, Gowda M, Hahn V (2013) Genomic prediction of sunflower hybrid performance. Plant Breed 132:107–114

R Development Core Team (2011) R: A Language and Environment for Statistical Computing. Vienna, Austria (http://www.R-project.org/)

Saghai Maroof MA, Yang GP, Zhang QF, Gravois KA (1997) Correlation between molecular marker distance and hybrid performance in U.S. southern long grain rice. Crop Sci 37:145–150

SAS Institute Inc (2012) SAS OnlineDoc® 9.3. SAS Institute Inc, Cary

Thomson MJ, Zhao K, Wright M, McNally KL, Rey J, Tung CW, Reynolds A, Scheffler B, Eizenga G, McClung A, Kim H, Ismail AM, Ocampo M, Mojica C, Reveche MY, Dilla-Ermita CJ, Mauleon R, Leung H, Bustamante C, McCouch SR (2012) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. Mol Breed 29:875–886

Troyer AF (2006) Adaptedness and heterosis in corn and mule hybrids. Crop Sci 46:528–543

Wang SJ, Lu ZM (2006a) Genetic diversity among parental lines of Indica hybrid rice (*Oryza sativa* L.) in China based on coefficient of parentage. Plant Breed 125:606–612

Wang SJ, Lu ZM (2006b) Evolvement and analysis of genetic diversity in indica hybrid rice (*Oryza sativa* L.) in China. Jiangsu J Agric Sci 22:192–198

Wang SJ, Lu ZM (2007) Study on heterosis groups of key parental lines in indica hybrid rice (*Oryza sative* L.) in China. J Nanjing Agric Univ 30:14–18

Xie FM (1993) Hybrid performance and genetic diversity as revealed by RFLPs in rice (*Oryza sativa* L.). Dissertation, Texas A&M University

Xie FM, Guo LB, Ren GJ, Hu PS, Wang F, Xu JL, Li XQ, Qiu FL, dela Paz MA (2012) Genetic diversity and structure of *indica* rice varieties from two heterotic pools of southern China and IRRI. Plant Genet Resour Charact Util 10:186–193

Xu WJ, Virmani SS, Hernandez JE, Sebastian LS, Redoña ED, Li ZK (2002) Genetic diversity in the parental lines and heterosis of the tropical rice hybrids. Euphytica 127:139–148

Yuan LP (1977) Practice and theory of breeding hybrid rice. Sci Agric Sin 10(1):27–31

Zhang Q, Gao YJ, Saghai Maroof MA, Yang SH, Li JX (1995) Molecular divergence and hybrid performance in rice. Mol Breed 1:133–142

Zhao K,Wright M, Kimball J, Eizenga G, McClung A, Kovach M, Tyagi W, Ali ML, Tung CW, Reynolds A, Bustamante CD, McCouch SR (2010) Genomic diversity and introgression in *O. sativa* reveal impact of domestication and breeding on the rice genome. PLoS One 5. doi:10.1371/journal.pone.0010780