Journal of Plant Biochemistry and Biotechnology
https://doi.org/10.1007/s13562-019-00529-9

ORIGINAL ARTICLE

Evaluation of genetic diversity of parental lines for development of heterotic groups in hybrid rice (*Oryza sativa L.*)

K. Sruthi[1] · B. Divya[3] · P. Senguttuvel[1] · P. Revathi[1] · K. B. Kemparaju[1] · P. Koteswararao[1] · R. M. Sundaram[4] · Vikram Jeet Singh[2] · E. Ranjith Kumar[2] · Prolay Kumar Bhowmick[2] · K. K. Vinod[2] · S. Gopala Krishnan[2] · A. K. Singh[2] · A. S. Hari Prasad[1]

Received: 26 April 2019 / Accepted: 14 August 2019

Abstract

Knowledge of genetic diversity and potential heterotic relationships among parental lines is of significant importance in hybrid rice breeding programs. In the present study, in order to understand the genetic diversity among 96 parental lines, they were characterized for their diversity with respect to their morphological traits (n = 12) and molecular markers using a set of 50 SSR markers. Morphological diversity was estimated using Mahalanobis D^2 statistics in terms of generalized group distance. Based on morphological diversity analysis, the 96 lines were grouped into 5 major and 13 monogenotypic clusters. In molecular marker analysis, the parental lines were consistently clustered into B (Maintainer group) and R (Restorer group) groups based on distance and model based approaches. Strong correspondence was observed between the pedigree of parental lines with molecular genotyping based grouping than morphological trait based grouping. From the results of the present investigation, it is evident that the available diversity among the two groups i.e., maintainer group (B) and restorer group (R) is sufficient for developing heterotic hybrids, but within the maintainer and restorer groups, the diversity is limited, the diversity among restorers was moderate, while it was low among the maintainers and hence efforts are needed for broadening their genetic base of parental lines for development and adoption of high-yielding hybrids.

Keywords Hybrid rice · Parental lines · Genetic diversity · SSRs

Abbreviations

SSR	Simple sequence repeats
GCP	Generation challenge programme
CTAB	Cetyl trimethyl ammonium bromide
PIC	Polymorphic information content
PCA	Principal component analysis
AMOVA	Analysis of molecular variance
UPGMA	Unweighted pair group method with arithmetic mean

Introduction

Rapid population growth may pose a threat of severe food shortage around the world, including India, in the near future due to shrinking of natural resources. This demands increase in production of staple cereal crops like rice in the realm of dwindling land and water resources and a rapidly changing climate. There is need to produce at least 40% more rice to feed ever increasing population globally. Therefore, enhancing rice productivity through large scale adoption of hybrid rice technology by the farming community is one of the potential approaches to bridge the productivity gap. So far, a total of 105 hybrids have been developed and released in India. However, 25 years of cultivation since the release of the first rice hybrid in India, the area planted under hybrid rice has reached only 3 million ha (6.8%) out of 44 million hectares under rice cultivation (Hari Prasad et al. 2018). The major reason behind the lower rate of adoption of hybrid rice technology by farming community is the low level of heterosis i.e. 15–20% (0.75–1.0 t/ha) (Virmani 1996), which may not be sufficient in high input productive areas where

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13562-019-00529-9) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article

Published online: 03 September 2019

inbred varieties are performing equivalent to hybrids in terms of yield. Lack of sufficient genetic diversity among the parental lines of hybrids could be one of the reasons for the constraints in breeding and adoption of hybrid rice technology in tropics (Xie et al. 2012). To raise the heterosis level further and also to broaden the genetic diversity among parents, research efforts are now focused more on breeding parental lines from indica/tropical japonica crosses (Hari Prasad et al. 2018). Large scale commercialization of hybrid rice technology depends on presence of genetic diversity among parental lines and further selection of diverse parents for developing highly heterotic hybrids. Hence, for successful promotion of hybrid technology, a systematic classification of parental lines into heterotic groups is needed along with a careful selection of parental lines (i.e. maintainer and restorer lines) on the basis of their genetic diversity for development of potential hybrids (Julfiquar et al. 1985). Encouraged from the great success witnessed through such studies in maize hybrids, rice breeders have initiated the heterotic group development in hybrid rice breeding (Xie et al. 2014; Wang et al. 2015; Beukert et al. 2017). This will allow the precise selection of potential parents for hybridization, thereby helping breeders to concentrate more on promising crosses between the groups. For development of heterotic gene pools, the first step is to study the genetic diversity pattern among the parental lines. Hence, this study was carried out with a major objective of understanding the level of genetic diversity present between the parental lines which are popularly used in hybrid rice breeding, both in terms of molecular and morphological characters.

Materials and methods

Plant material

Plant material comprising of 96 parental lines, including 31 maintainer lines and 65 restorer lines (developed at Indian Council of Agricultural Research-Indian Institute of Rice Research (ICAR-IIRR), Hyderabad, India and International Rice Research Institute (IRRI), Philippines) (ESM_1) were used for molecular and morphological diversity analysis. The molecular marker analysis was carried out at Genetics Division, ICAR-IARI, New Delhi and morphological characterization of parental lines was done at ICAR-Indian Institute of Rice Research (ICAR-IIRR), Hyderabad during *Kharif* 2016.

Selection of SSR markers

A panel of 50 SSR markers having extensive genome coverage used by International Rice Research Institute (IRRI), Philippines under generation challenge programme (GCP) for rice diversity analysis were selected. Information regarding markers with respect to chromosome number, position, annealing temperatures, allele size, forward and reverse sequence and repeat motifs were obtained from GRAMENE website (http://gramene.org/markers/microsat/50_ssr.html).

Phenotyping

Measurement of agro-morphological traits

Yield and yield attributing traits were recorded on three plants per line in each replication according to the descriptors prescribed by International Rice Research Institute (IRRI) (SES, IRRI, 2013). The traits namely, days to 50% flowering (DFF), plant height (PH), panicle length (PL), number of productive tillers (NPT), number of unproductive tillers (NUPT), total number of tillers (TT), single plant yield (SPY), spikelet fertility (SF), number of filled grains (NFG), number of unfilled grains (NUFG), total grains per panicle (TG), 1000 grain weight (TGW) were recorded during *kharif* 2016 (Wet season) as per prescribed protocols (SES, IRRI 2013) at research farm of ICAR-Indian Institute of Rice Research (ICAR-IIRR), Hyderabad (17°19′N and 78°29′) at an altitude of 549 m above mean sea level.

Genotyping

DNA isolation and PCR protocols

Seed material of 96 parental lines was sown in protray and leaf samples of 15-day old seedlings were collected for DNA isolation. Genomic DNA was isolated using CTAB procedure (Saghai-Maroof et al. 1984) with minor modifications. A set of 50 GCP-SSR markers were used for genotyping. The genomic DNA was subjected to PCR amplification using a programmable thermo cycler (Veriti Thermo Cycler, Applied Bio systems). The master mix was prepared in a 1.5 ml microcentrifuge tube by taking 5.3 µl of nuclease free water, 1.0 µl 10X buffer, 1 µl dNTPs, 1 µl (for both forward and reverse primers) and 0.2 µl of 2U/µl taq DNA polymerase. The initial denaturation cycle of 95 °C, for 5 min was followed by 35 cycles at 95 °C for 30 s, 50–69 °C for 30 s (annealing temperatures are based on http://archive.gramene.org/markers/), 1 min of extension at 72 °C with an additional step of 10 min at 72 °C was employed. After completion of PCR, the samples were collected and stored at − 20 °C. The SSR-PCR products were resolved in a 3.5% Agarose gel prepared in 1X TAE buffer stained with Ethidium Bromide and it was documented using gel documentation system (Bio-Rad, USA).

Scoring of SSR Bands Allele scoring was done manually following the expected minimum and maximum allele size

given in the GRAMENE website (http://gramene.org/markers/microsat/50_ssr.html). Only the clear and unambiguous amplified bands were scored. Amplicon size of each band, in each of the line or sample was identified using 50 bp DNA ladder and allele scoring was carried out based on the amplicon size data in the gel documentation system and converted into binary data. Presence of particular size allele for each marker was indicated as 1 and absence as 0, and the same was followed for all alleles at each locus for 39 polymorphic markers and a binary data matrix was generated. This genotypic data was used for molecular diversity analysis. To estimate the discriminatory power of a marker, the polymorphic information content for each SSR marker was calculated. PIC value measures the informativeness of the markers and assesses the allelic diversity in a population which accentuate the evolutionary pressure on the allele at a locus and the mutation occurred to the locus over a time period. Botstein et al. (1980) reported that PIC index can be used to evaluate the level of gene variation, the locus was considered of high diversity when PIC > 0.5; low diversity when PIC < 0.25; and the locus was of intermediate diversity, when PIC was between 0.25 and 0.5.

$$\mathrm{PIC} = 1 - \Sigma \mathrm{Pi}^2 - \Sigma\Sigma \mathrm{Pi}^2 \mathrm{Pj}^2,$$

where 'i' is the total number of alleles detected for SSR marker and 'Pi' is the frequency of the ith allele in the set of genotypes investigated and $j = i + 1$ (Botstein et al. 1980).

Data analysis

Morphological data was subjected to analysis for descriptive statistics using SAS enterprise guide 4.3 (SAS Institute Inc., Cary, NC, USA). Correlation and PCA analysis was carried out by using R programme (R Core Team 2012) with *cor* and *princomp* function respectively. To study the morphological diversity present among the hybrid rice parental lines and to study the genetic relationships, the data have been subjected to Mahalanobis D^2-Statistics (Rao, 1952). Morphological clustering was also performed by DARwin software ver 6.0.010 (Perrier and Jacquemoud-Collet 2006). To perform AMOVA analysis, GenALEx 6.5 v (Peakall and Smouse 2012) was used.

Cluster analysis

Molecular diversity parameters such as major allele frequency, allele number per locus, gene diversity, heterozygosity and polymorphic information content (PIC) were calculated using software POWERMARKER Ver3.25 (Liu and Muse 2005). There are two approaches to assess the genetic structure of a population, one is model based approach and another one is distance based approach. Model based approach was employed in STRUCTURE software ver 2.3.4 (Pritchard et al. 2000). The number of subpopulations denoted by K is obtained from STRUCTURE grouping. The optimum K value was estimated by plotting the mean estimate of the log posterior probability of the data [L (K)] against the K value using STRUCTURE HARVESTER (Earl 2012) and the clear peak point is considered as the optimum K value. Distance based clustering was performed by DARwin software ver 6.0.010 (Perrier and Jacquemoud-Collet 2006). To determine the clustering pattern among the set of parental lines dissimilarity matrix was used, based on Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbour- joining method. Confidence limits were tested by boot strapping for 1000 times to estimate the robustness of genotype clustering. To study the agreement or correlation between two distance matrices of phenotypic and genotypic data MANTEL test was carried out using R programme with ade4 package (Thioulouse et al. 1997; Dray and Dufour 2007).

Results

The present study was carried out at Hybrid Rice, Crop Improvement Section, ICAR-IIRR, Hyderabad and at Division of Genetics, Rice Section, IARI, New Delhi using 96 parental lines of hybrid rice, comprising 31 maintainers and 65 restorers.

Morphological characterization

Morphological characterization of hybrid rice parental lines revealed that a wide range of variation existed for traits such as NUFG (CV = 65.35%), NUPT (52.92%), TG (CV = 32.84%), NFG (CV = 31.22%), showing high coefficients of variation among the genotypes whereas traits such as DFF (CV = 6.63), PL (CV = 9.28%) and PH (CV = 9.82) showing the lowest (Table 1). Among the restorers, IR10198 (97 days) was the earliest flowering line and TCP585 was late flowering which has taken 134 days for 50% flowering. Among the maintainers, TCP771 (95 days) recorded earliest flowering and TCP1199 (129 days) recorded late flowering. The plant height varied from 64.17 cm (DR 714-1-2R) to 104.33 cm (BK64-116) for restorers and in maintainers the plant height varied from 71.17 cm (TCP731) to 99.33 cm (TCP765). In restorers, the panicle length ranged from 10.33 cm in IR72 to 22.83 cm for TCP341, RPHR-611-1 and RPHR-124 and in maintainers, it was from 16.5 cm in TCP1193 to 24.33 cm for TCP724. The number of productive tillers varied from 5.83 to 15.67 for restorers. The restorer line BK-64-116 (5.83) had registered with lowest number of productive tillers and IR48725 (15.67) with maximum number of

Table 1 Descriptive statistics for morphological traits studied for hybrid rice parental lines

Morphological trait	Mean (± SE)	Range	CV(%)
Days to 50% flowering (DFF)	114.67 (0.74)	94.5–133.5	6.63
Plant height (cm) (PH)	81.94 (0.77)	64.17–104.33	9.82
Panicle length (cm) (PL)	20.16 (0.18)	10.33–24.33	9.28
No. of productive tillers (NPT)	10.19 (0.22)	5.33–16	22.88
No. of unproductive tillers (NUPT)	1.51 (0.08)	0–4.5	52.92
Total no. of tillers (TT)	11.71 (0.23)	6.66–18.17	20.8
Single plant yield (g) (SPY)	10.82 (0.28)	2.53–20.14	26.29
1000 grain weight (g) (TGW)	18.84 (0.36)	11.25–27.85	19.68
No. of filled grains (NFG)	113.67 (3.43)	37.17–232.67	31.22
No. of unfilled grains (NUFG)	37.81 (2.39)	4–131.17	65.35
Total grains (TG)	151.49 (4.81)	48.33–375.5	32.84
Spikelet fertility (%)(SF)	75.98 (0.97)	38.38–96.06	13.24

productive tillers per plant. In maintainers, number of productive tillers varied from 5.33 (TCP 716) to 16 (TCP1216). The highest SPY was recorded in check US312 (20.14 g) while lowest in restorer TCP 343 (2.53 g) and maintainer TCP797 (4.88 g). Among the restorers, TCP341 (17.73 g) was recorded with highest SPY, while in maintainers, TCP1221 (15.45 g) had highest SPY. Restorer TCP314 (27.85 g) had maximum TGW, lowest TGW recorded in RP Bio 4919-363-5(11.75 g). Highest TGW among maintainers was observed in TCP876 (26.15 g) while TCP 766 (11.25 g) recorded least TGW.

Among restorers, maintainers and checks, the average performance is not significantly different for most of the traits except number of filled grains, number of unfilled grains and total number of grains whereas checks recorded highest values followed by restorers and maintainers (Fig. 1). Pearson correlation coefficients for 12 quantitative traits are shown in Table 2 and Fig. 2. The highest positive significant correlation was observed between NPT and TT ($r = 0.94$, $p < 0.01$), followed by NFG and TG ($r = 0.88$, $p < 0.01$) and TUFG and TG ($r = 0.74$, $p < 0.01$). Highest significant negative association was found between NUFG and SF% ($r = -0.81$, $p < 0.01$), followed by, TGW and TG ($r = -0.57$, $p < 0.01$), TGW and NFG ($r = -0.49$, $p < 0.01$) and TGW and NUFG ($r = -0.43$, $p < 0.01$). DFF exhibited positive significant association with TUFG and TG while significant negative association with TGW and SF%. PH showed significant positive association with PL, NFG and TG and significant negative association with NPT and TT. NPT was highly correlated with TT and significant negative correlation with NFG and TG. NUPT showed significant positive correlation with TT. TGW is having significant negative association with NFG, NUFG and TG and significant positive association with SF. Yield was significantly correlated with NPT ($r = 0.215$,

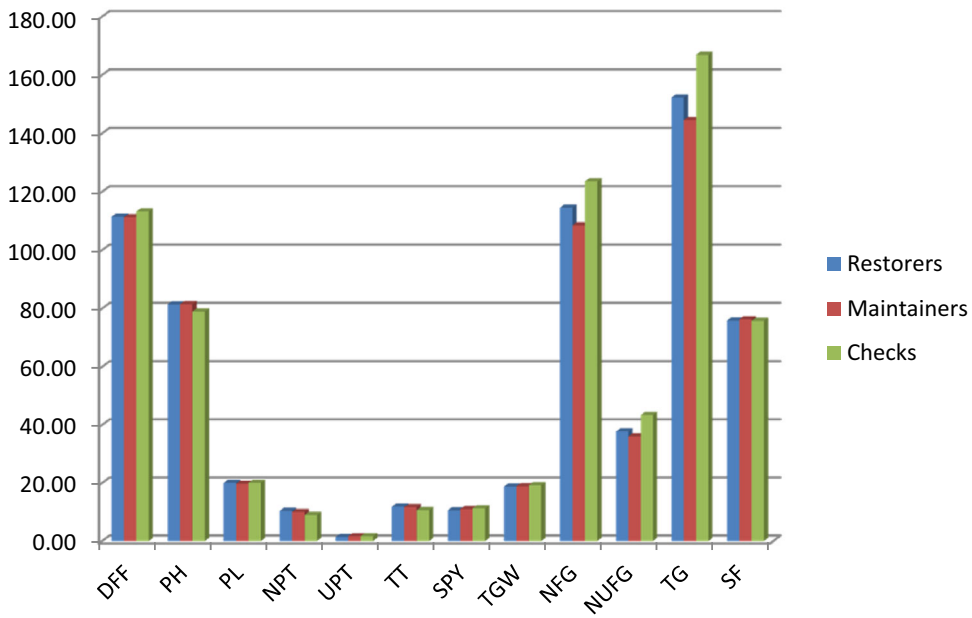

Fig. 1 Graph showing average performance of restorers, maintainers and checks for 12 phenotypic traits

Table 2 Correlation coefficients for 12 phenotypic traits

	DFF	PH	PL	NPT	NUPT	TT	TGW	NFG	NUFG	TG	SF%	SPY
DFF												
PH	0.179											
PL	0.015	**0.409****										
NPT	− 0.106	**− 0.294****	**− 0.294****									
NUPT	− 0.075	0.077	0.014	− 0.041								
TT	− 0.126	**− 0.256****	**− 0.277****	**0.944****	**0.290****							
TGW	**− 0.273****	− 0.179	0.180	− 0.106	− 0.021	− 0.108						
NFG	0.154	**0.333****	0.046	**− 0.292****	− 0.134	**− 0.323****	**− 0.495****					
NUFG	**0.465****	0.056	− 0.091	− 0.015	− 0.031	− 0.024	**− 0.436****	**0.345****				
TG	**0.341****	**0.265****	− 0.012	**− 0.216***	− 0.111	**− 0.242***	**− 0.570****	**0.885****	**0.743**			
SF	**− 0.418****	0.074	0.109	− 0.177	− 0.061	− 0.189	**0.198***	0.168	**− 0.818**	**− 0.286****		
SPY	− 0.115	0.107	0.017	**0.215***	− 0.156	0.153	0.051	**0.327****	− 0.086	**0.191***	**0.299****	

Bold values represent significant correlation coefficient values

DFF, Days to 50% flowering; PH, Plant height;PL, Panicle length; NPT, Number of productive tillers;NUPT, Number of unproductive tillers;TT, Total number of tillers; TGW, thousand grain weight; NFG, Number of filled grains; NUFG, Number of unfilled grains; TG, Total number of grains; SF%, Spikelet fertility %; SPY, Single plant yield

*, **Level of significance at 5% and 1% respectively and given in bold

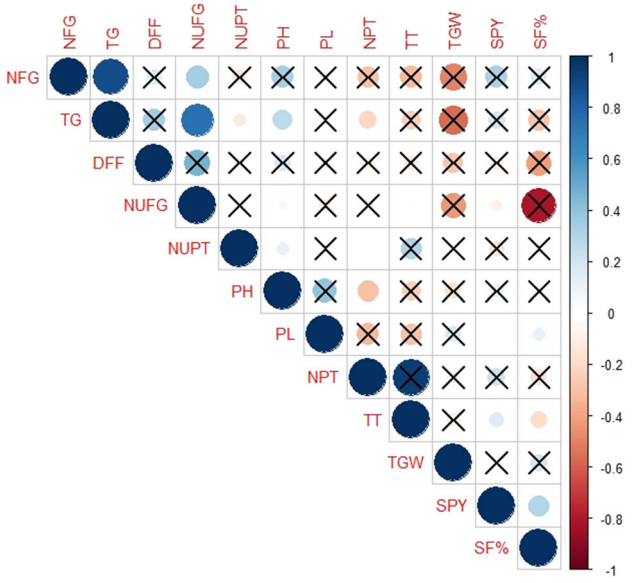

Fig. 2 Correlogram for 12 phenotypic traits (Correlations with p value > 0.05 are considered as non-significant and shown by cross mark)

$p < 0.05$), NFG (r = 0.32, $p < 0.01$), TG (r = 0.191, $p < 0.05$) and with SF% (r = 0.299, $p < 0.01$).

Yield and yield related traits are having complex inter-relationships, which results in multicolinearity. Multicolinearity can mask the real importance of a variable and bias the estimation. Hence PCA was used to resolve the issue of multi-colinearity associated with yield components. Principal component analysis (PCA) is one of the multivariate techniques used to identify the traits which classify the genotypes into separate groups (Ariyo 1987; Nair et al. 1998). It measures the independent impact of each trait towards the total variance. And high Eigen values correspond to high variances. The first four principal components captured 72.02% of cumulative variance with eigen value of above 1.0 (Table 3). The first two principal components PC1 and PC2 explain 48.02% of the total variation (Fig. 3). Degree of contribution of each variable towards the each principal component is given in Table 3. The traits DFF, NFG, NUFG and TG have contributed 30% of total variance towards the first principal component and remaining 18% of total variance explained by PH, PL and PT for second principal component. Similarly SPY alone captured 15% of total variance for third principal component. NUPT and TGW have explained 9% of variance for fourth principal component.

Morphological genetic divergence by Mahalanobis D^2-statistics

Morphological diversity has been estimated among 110 lines including 100 parental lines of Hybrid rice and 10

Table 3 Variables contributions to major 4 principal components and Eigen values and percent of variance explained by each principal component

	PC1	PC2	PC3	PC4
DFF	− **0.308**	0.141	− 0.238	0.288
PH	− 0.189	− **0.411**	− 0.221	0.082
PL	0.00748	− **0.38**	− 0.3781	0.378
SPY	− 0.0582	− 0.159	**0.5466**	0.309
PT	0.0967	− **0.4083**	0.3828	0.042
NUPT	0.0477	− 0.0088	− 0.33	− **0.644**
TGW	0.376	− 0.061	− 0.127	**0.42**
NFG	− **0.408**	− 0.351	0.264	− 0.15
NUFG	− **0.4616**	0.287	− 0.114	0.112
TG	− **0.5208**	− 0.1077	0.131	− 0.051
SF%	0.2463	− **0.5026**	0.282	− 0.196
Eigen values	3.26	2.03	1.61	1.03
Standard deviation	1.80	1.42	1.27	1.02
Proportion of variance	0.30	0.18	0.15	0.09
Cumulative proportion	0.30	0.48	0.63	0.72

Bold values represent the traits that contribute majorly to total variance under each principal component

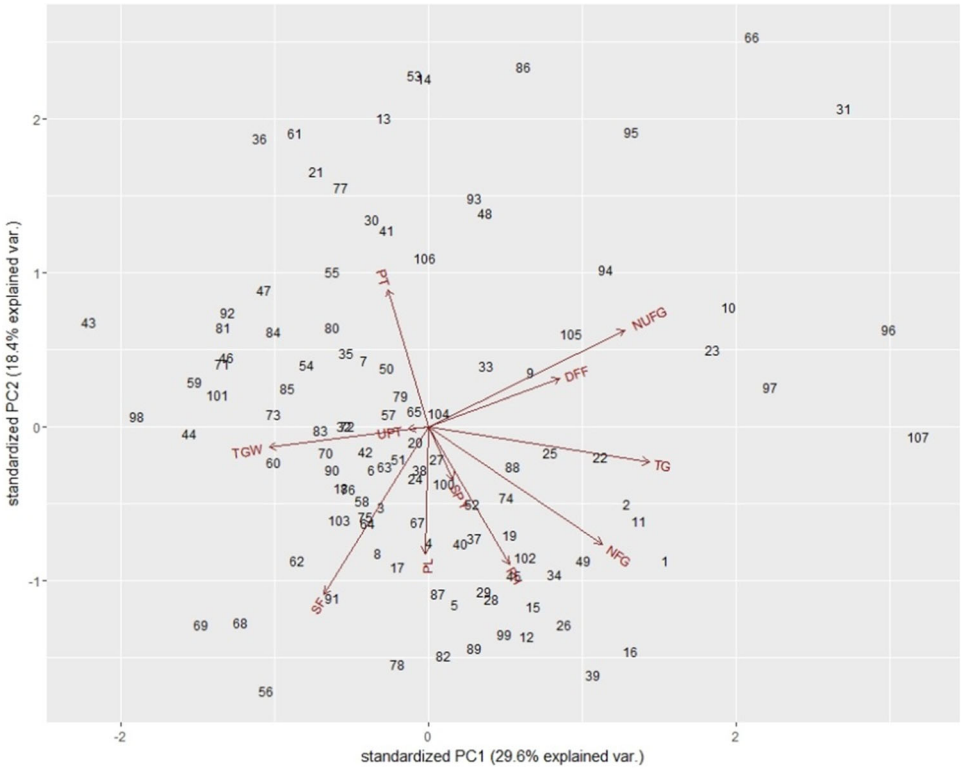

Fig. 3 Principal component analysis

checks using Mahalanobis D^2-statistics in terms of generalized group distance. ANOVA for 12 yield and yield component traits revealed that significant difference was found among all varieties for majority of the traits except number of unproductive tillers which justified for further calculations to D^2-Statistics. Based on Mahalanobis D^2-Statistics total 110 lines grouped into 18 clusters after removing outliers. Cluster-I was the largest which constitutes 38 genotypes followed by Cluster III (23 genotypes), Cluster V (12 genotypes), Cluster XII and XIII (8 genotypes) and the remaining 13 clusters [BK35-155, EPLT-109, DR 714-1-2R, IR10198, TCP349, TCP798, US312, TCP766, TCP765, TCP771, Anjali, BK-116 and Improved Samba mahsuri (ISM)] were monogenotypic clusters with single genotype in each cluster. Lower intra cluster distances were observed than inter cluster distances (Table 4). Intra cluster distances ranged from 0 (Cluster II, IV, VI, VII, VIII, IX, X, XI, XIV, XV, XVI, XVII and XVIII) to

Table 4 Intra and inter- cluster divergence in 100 parental lines of hybrid rice along with 10 checks

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15	Cluster 16	Cluster 17	Cluster 18
Cluster 1	8.42	14.85	15.79	14.13	15.64	13.84	15.98	19.84	12.73	19.82	14.29	18.83	24.54	29.57	31.98	31.77	22.19	18.61
Cluster 2		0	5.15	29.25	26.15	32.13	35.02	33.45	28.3	42.23	20.43	19.8	24.54	10.12	34.71	10.73	17.29	31.03
Cluster 3			7.98	28.25	27.81	30.53	32.36	32.19	24.47	41.94	19.95	20.7	22.46	15.82	34.00	19.47	18.92	34.08
Cluster 4				0	29.18	16.55	7.99	3.98	7.04	18.9	15.77	17.38	24.59	56.53	16.32	41.71	27.88	23.32
Cluster 5					11.17	25.8	23.15	40.36	25.23	33.1	25.27	35.07	38.23	32.95	63.14	52.18	44.48	29.56
Cluster 6						0	22.2	19.11	15.01	6.2	24.57	23.71	39.8	49.95	30.22	44.45	22.01	12.41
Cluster 7							0	12.17	5.98	21.3	14.9	26.19	29.15	58.79	29.75	57.95	43.26	35.1
Cluster 8								0	9.4	18.89	20.91	22.52	28.57	64.56	7.03	40.24	23.12	24.25
Cluster 9									0	20.24	8.93	22.15	19.48	50.36	22.05	52.36	26.8	27.97
Cluster 10										0	29.88	31.43	51.25	66.26	30.71	48.89	32.81	12.58
Cluster 11											0	27.33	19.38	38.42	30.01	46.56	22.47	38.46
Cluster 12												18.54	29.46	37.37	26.95	27.78	28.57	30.53
Cluster 13													17.88	36.91	35.48	46.55	29.7	46.52
Cluster 14														0	67.24	24.87	38.22	52.17
Cluster 15															0	32.12	19.46	38.5
Cluster 16																0	23.41	34.69
Cluster 17																	0	24.5
Cluster 18																		0

Diagonal values are intra-cluster divergence and off-diagonal values inter-cluster divergence

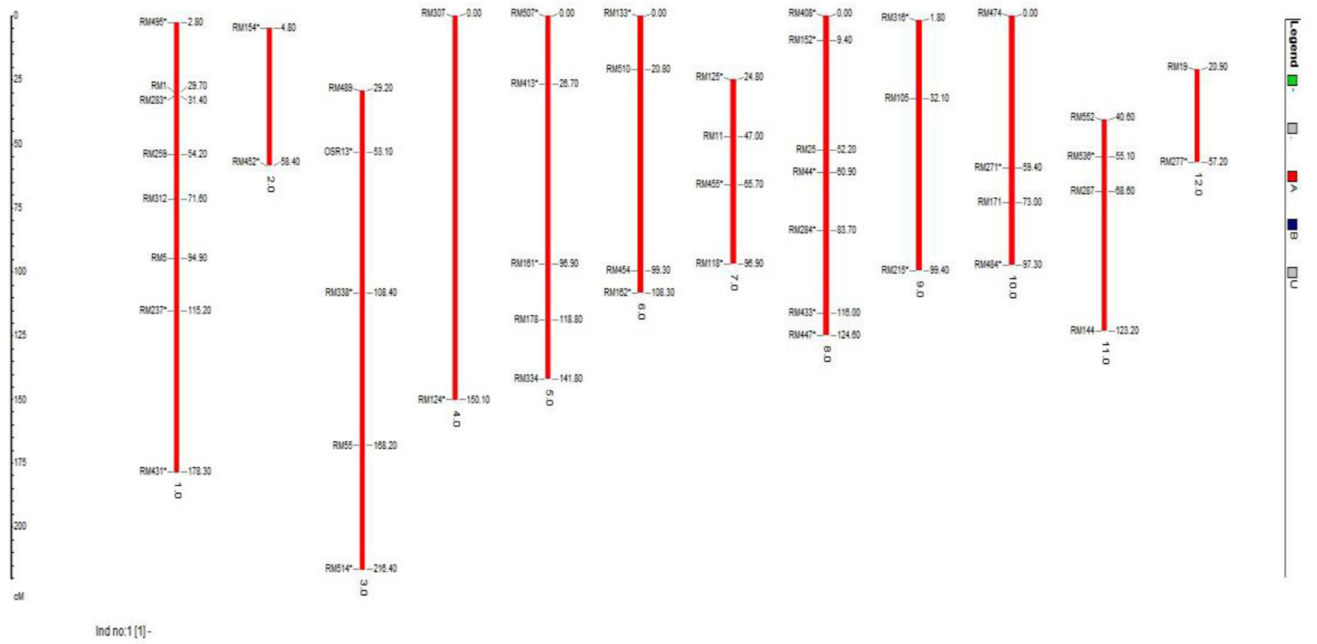

Fig. 4 Graphical representation of GCP markers used for molecular diversity analysis

18.54 (Cluster XII). Inter cluster distances ranged from 3.98 (between Cluster IV & VIII) to 66.26 (between Cluster X & XIV). Cluster means specifies average performance of all varieties grouped in a single cluster. Unique performance of cluster members was reflected in ESM_2. Cluster 13 recorded highest mean values for DFF, NFG and NUFG. Cluster 14 contained genotypes with highest mean values for plant height. Cluster 11 contained long panicle genotypes (mean = 23 cm) and cluster 4 recorded highest values for NPT, TT and SPY. Cluster 18 recorded highest values for NUPT and cluster 17 recorded highest spikelet fertility genotypes. Cluster 5 contained highest TGW genotypes. At inter cluster level, NUFG (CV = 64.79%), NUPT (CV = 56.21%), NFG (CV = 33.53%) and SPY (25.76%) are potential contributors to differentiation in lines and at genotypic level NUPT (CV = 84.67%) and NUFG (CV = 59.55%) are major contributors for variation in genotypes (ESM_2).

Morphological clustering by DARwin software

Morphological clustering was also done based on distance (boot strapped dissimilarities) matrices generated using Darwin Software ver 6.0.010 which grouped the 110 parental lines into total 7 clusters (5 major clusters and 2 minor clusters) (Fig. 5a), 2 minor clusters which consists of single and two genotypes/lines separately, 5 major clusters manifested for late duration genotypes. In minor clusters, cluster VI single genotype grouped for mid early duration. In cluster I, the genotypes were showing highest mean value for PH, PL and SF%. Cluster II genotypes showing highest mean values for DFF, NFG, NUFG and TG. Custer IV was showing high mean values for PT and TGW. Cluster V includes highest mean value for SPY and NUPT. The genotype IR10198 formed a separate cluster with mid early duration (96 days) and two more genotypes, TCP-1145 and TCP-343 formed a separate cluster with late duration genotypes. Out of 7 clusters, 6 grouped for both maintainers and restorers except one minor cluster with single genotype.

Molecular marker analysis

A random set of 50 SSR (Fig. 4) markers recommended by IRRI under Generation challenge programme for diversity studies was used to assess the genetic diversity among the parental lines. Out of 50 SSR markers, 39 were found to be polymorphic, 1 monomorphic and 10 did not amplify. The results for major allele frequency, allele number per locus, gene diversity, heterozygosity and polymorphic information content (PIC) for polymorphic markers have been presented in Table 5. A total of 173 alleles were amplified by 39 polymorphic markers. Allele number per locus generated by each marker ranged from 2 (OSR13, RM338, RM55, RM161) to 8 (RM152) with an average of 4.435 alleles per locus. Major allele frequency represents frequency of key allele for each marker. Major allele frequency is defined as most common allele at each locus and it ranged from 30 to 89%. The value of major allele frequency ranged from 0.30 (RM1) to 0.89 (RM162) with a

Table 5 Details of polymorphic markers along with molecular diversity parameters

S. No.	Marker	Chromosome No.	Allele No.	Major allele frequency	Gene diversity	Heterozygosity	PIC
1	RM495	1	4	0.48	0.61	0.15	0.53
2	RM1	1	7	0.30	0.76	0.11	0.72
3	RM283	1	4	0.58	0.60	0.00	0.55
4	RM312	1	3	0.48	0.62	0.00	0.55
5	RM5	1	3	0.58	0.50	0.08	0.40
6	RM452	2	3	0.51	0.51	0.04	0.39
7	OSR13	3	2	0.71	0.42	0.02	0.33
8	RM338	3	2	0.52	0.50	0.00	0.37
9	RM55	3	2	0.53	0.50	0.00	0.37
10	RM514	3	3	0.66	0.51	0.06	0.46
11	RM307	4	7	0.51	0.68	0.23	0.64
12	RM124	4	3	0.64	0.51	0.00	0.44
13	RM507	5	6	0.35	0.74	0.01	0.69
14	RM413	5	4	0.38	0.71	0.11	0.66
15	RM161	5	2	0.50	0.50	0.00	0.38
16	RM178	5	4	0.51	0.57	0.00	0.48
17	RM334	5	5	0.38	0.71	0.06	0.66
18	RM510	6	4	0.46	0.65	0.08	0.59
19	RM454	6	3	0.50	0.62	0.00	0.55
20	RM162	6	3	0.89	0.20	0.05	0.19
21	RM125	7	4	0.48	0.66	0.05	0.60
22	RM11	7	4	0.49	0.66	0.08	0.61
23	RM455	7	5	0.42	0.71	0.00	0.67
24	RM118	7	4	0.51	0.63	0.01	0.56
25	RM408	8	4	0.47	0.59	0.03	0.50
26	RM152	8	8	0.48	0.70	0.08	0.66
27	RM25	8	7	0.41	0.73	0.09	0.69
28	RM44	8	4	0.50	0.55	0.01	0.45
29	RM284	8	6	0.44	0.69	0.02	0.64
30	RM433	8	3	0.55	0.54	0.00	0.44
31	RM447	8	7	0.39	0.70	0.12	0.65
32	RM316	9	6	0.30	0.77	0.13	0.73
33	RM105	9	6	0.35	0.73	0.09	0.68
34	RM215	9	5	0.41	0.68	0.04	0.62
35	RM474	10	6	0.37	0.73	0.11	0.68
36	RM271	10	6	0.55	0.60	0.04	0.54
37	RM171	10	7	0.29	0.79	0.09	0.76
38	RM536	11	3	0.52	0.54	0.00	0.44
39	RM277	12	4	0.55	0.60	0.01	0.53
	Mean		4.44	0.49	0.62	0.05	0.55

mean value of 0.49. On an average 49% of the 96 parental lines shared a common major allele at any given loci. Gene diversity ranged from 0.20 (RM162) to 0.79 (RM171) with an average value of 0.62. Gene diversity defined as probability that two randomly chosen alleles are different in the sample (Choukan and Warburton 2005). Heterozygosity shows the proportion of heterozygous individuals in a population and it was ranged from 0.01(RM507, RM118, RM44, RM277) to 0.23 (RM307) with a mean 0.05. Eleven markers (RM283, RM312, RM338, RM55, RM124, RM161, RM178, RM454, RM455, RM433 and RM536) showed no heterozygosity. In the present study PIC values

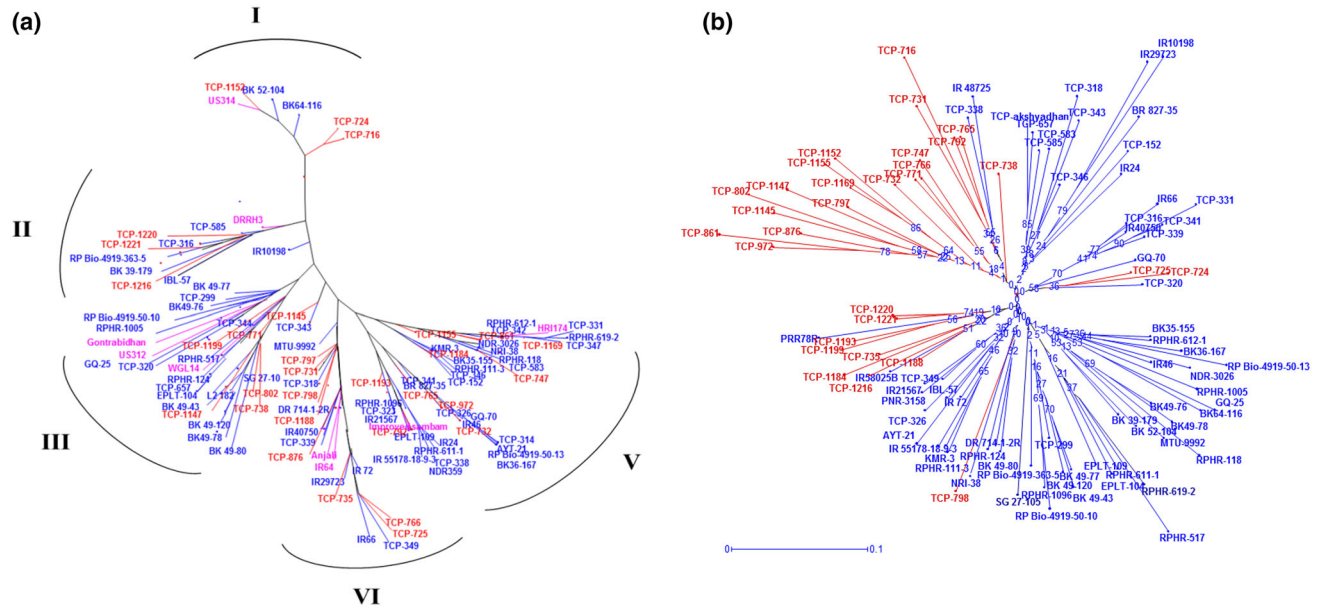

Fig. 5 Morphological and molecular clustering pattern: **a** Radial representation of phenotypic cluster of 110 parental lines (70 restorers, 30 maintainers 10 checks) based on Euclidean distances and **b** Radial representation of Neighbour-joining tree of 96 parental lines (31 maintainers and 65 restorers)

ranged from 0.19 (RM162) to 0.76 (RM171) with an average PIC value of 0.55. The SSR loci RM162 on chromosome 6 and RM171 on chromosome 10 showed lowest (0.19) and highest (0.76) PIC values respectively.

Molecular genetic diversity pattern by cluster analysis

Cluster analysis was performed using UPGMA based Neighbour joining algorithm in DARwin software. It resolved 96 parental lines (comprising 31 maintainer and 65 restorer lines) into 2 major clusters (Fig. 5b). Total parental lines grouped for maintainers and restorers. All maintainers grouped under first cluster with few restorers i.e., NRI-38, RPHR-124, GQ-70, IR10198, IR40750 and TCP-585 while second cluster consists of mostly the restorers with one or two maintainers. Few admixtures were also observed, possibly due to shared ancestry during the breeding process.

Genetic structure of a population

Model based clustering was performed by using STRUCTURE software ver 2.3.4. K value was set from 1 to 10 and the run was repeated 10 times. Run length was set as 10,000 burning period followed by 5000 Markov Chain Monte Carlo (MCMC) replication. The optimum K value estimated by plotting the mean estimate of the log posterior probability of the data [L (K)] against the K value using STRUCTURE HARVESTER (Earl 2012). The log likelihood displayed by structure revealed the optimum K value as 2 ($K = 2$), which indicated that the entire population can be grouped into two subgroups (Fig. 6a, b). Out of 96 parental lines, 30 grouped into cluster I and 66 grouped into cluster II. Out of 30 from cluster I, 24 lines were considered as pure and six as admixtures (ESM_3). In second cluster, out of 66, 54 were pure and 12 were admixtures (Fig. 6b). Allele frequency between two clusters was 0.1262 and average distances between individuals within two clusters were 0.5282 (Cluster I) and 0.5808 (Cluster II). Mean alpha value was 0.1395. Mean fixation index (Fst) values of cluster I and cluster II were 0.1719 and 0.1710, respectively.

Analysis of molecular variance (AMOVA)

AMOVA is a statistical method which was used to partition the genetic variation into successive levels and helps to study the level of genetic difference among different populations. Analysis of molecular variance for parental lines of hybrid rice showed 11% variation among populations, whereas, 82% variation was present among individuals and 7% variation within individuals (ESM_4).

Comparison between molecular and morphological data

The agreement between the genetic distances calculated from molecular and morphological data was tested using mantel's test. Negative non-significant correlation

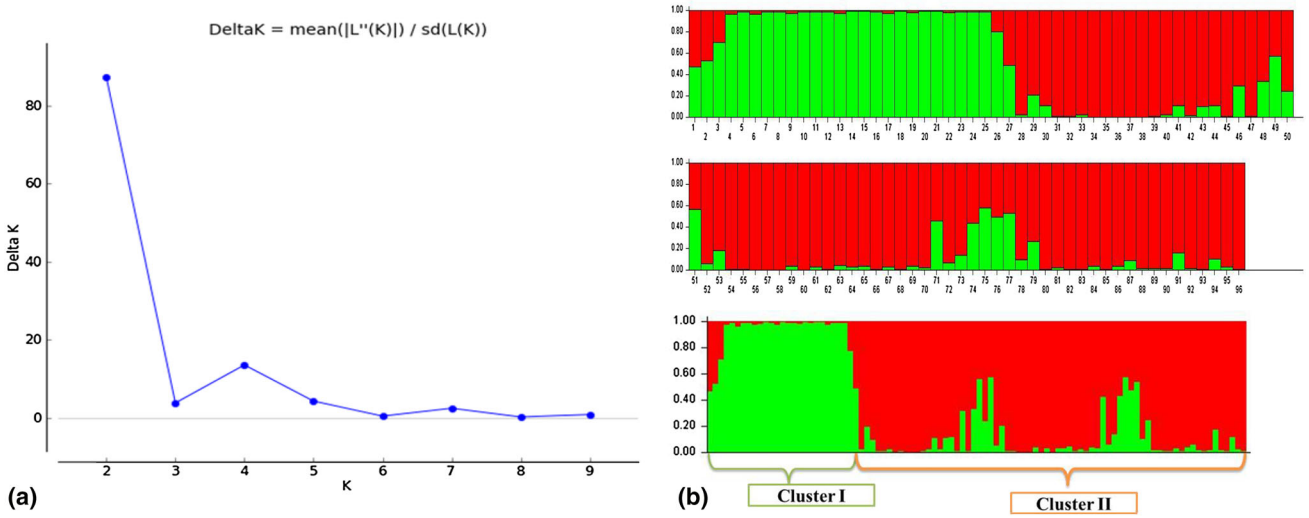

Fig. 6 Population structure: **a** Estimation of population using LnP(D) derived delta K for determining optimum number of subpopulations. The maximum of adhoc measure ΔK determined by structure harvester was found to be K = 2, which indicated that the entire population can be grouped into 2 subgroups and **b** Population structure of 96 parental lines with K = 2

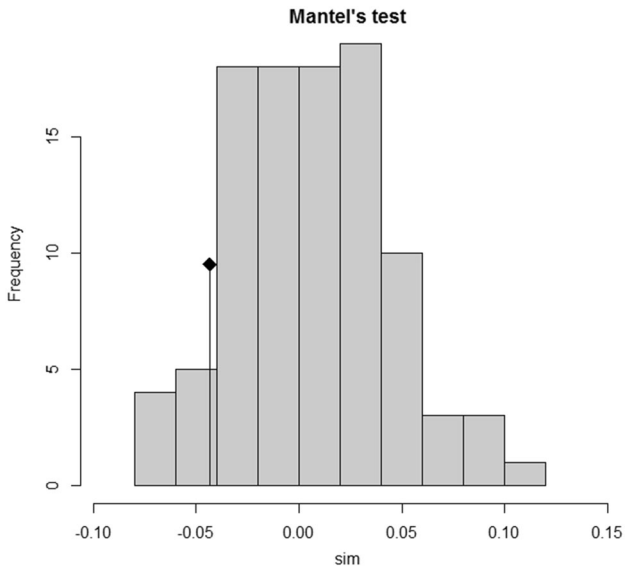

Fig. 7 Mantel plot showing the negative correlation value of -0.04

(r = − 0.04) was found between two distance matrices (Fig. 7).

Discussion

To begin any breeding program, knowledge of genetic diversity of parental lines is essential for judicious use of germplasm and to select best heterotic parents for hybridization. To breed heterotic rice hybrids, parental lines should be genetically diverse with high yielding ability and they must be good combiners. Narrow genetic base of hybrid rice parental lines is the major constraint to develop heterotic hybrids in the tropics (Xiao et al. 1996a, chapter 10). Careful selection of maintainer and restorer lines on the basis of their genetic diversity may lead to the development of hybrids with higher yield potential than parents and standard check varieties (Julfiquar et al. 1985). Nowadays, using molecular marker technology, genetic diversity of parental lines can be determined accurately. According to Zhang et al. (1994), molecular markers can also help in tagging heterotic gene blocks and developing heterotic combinations. In the present study, genetic diversity of parental lines has been studied in terms of both molecular and morphological means. 110 genotypes which include 100 parental lines along with 10 checks were screened for morphological diversity and for PCR convenience, only selected 96 lines were genotyped for molecular diversity. Molecular diversity parameters such as major allele frequency ranged from 0.30 to 0.89 with an average of 0.49. Number of alleles per locus ranged from 2 to 8. The value is comparable to the study of Siwach et al. (2004) where they observed 1–8 allele per SSR locus with an average number of alleles of 4.58 per locus for various classes of microsatellite and lower than the value reported by Wang et al. (2016) which varied from 2 to 13 with an average of 5.60. Higher number of alleles identified in the present study can be attributed to the larger number of parental lines studied and the use of more number of markers as compared to the previous studies. High heterozygosity detected at any SSR locus is potentially meaningful because increase in heterozygosity levels would indicate that the plant population likely has a substantial amount of adaptive genetic variation to escape the effects of a control agent that limits the development and maintenance of plants (Allendorf and Luikart 2007;

Maranho et al. 2014), compared to plant populations showing a lower level of heterozygosity. Different mutational properties of markers result in variation in heterozygosities and allele frequencies.

Polymorphic Information Content (PIC) value represents the allelic diversity in a population which accentuate the evolutionary pressure on the allele at a locus. To estimate the discriminatory power of a marker, the polymorphic information content for each SSR marker was calculated. In the present study PIC values ranged from 0.19 (RM162) to 0.76 (RM171) with an average PIC value of 0.55. The SSR loci RM162 on chromosome 6 and RM171 on chromosome 10 showed lowest (0.19) and highest (0.76) PIC values respectively. In earlier studies PIC values ranging from 0.02 to 0.97 (Singh et al. 2011; Sripathy et al. 2012; Yadav et al. 2013; Tiwari et al. 2015; Shrivastava et al. 2015; Surapaneni et al. 2016; Mahalingam et al. 2016) have been reported. In the present study, the average PIC value of 0.55 was comparatively higher than the earlier observations reported by Yadav et al. (2013) and Shrivastava et al. (2015), with an average PIC value of 0.41 and 0.4289 respectively. It was less than the observation reported by Tiwari et al. (2015) where they reported an average PIC value of 0.671. Markers with PIC values of 0.5 or above are considered highly useful in distinguishing the genotypes (DeWoody et al. 1995; Akkaya and Buyukunal Bal 2004). Hence in the present study, a set of markers with an average PIC value of more than 0.5 could be considered as highly informative. Out of 39 markers, 26 markers recorded more than PIC value of 0.5, which can be considered as highly informative and useful to study molecular genetic diversity.

Genetic diversity of popular hybrid rice parental lines has been studied using phenotypic data (yield and component traits) and genotypic data (SSR markers). As per molecular diversity analysis, the 96 parental lines considered in the present study clustered into two major groups (restorer group and maintainer group). A similar kind of molecular grouping for hybrid rice parental lines has been noticed by many research studies (Xu et al. 2002; Wang et al. 2016; He et al. 2012; Rajendran et al. 2012). Wang and Lu (2006) studied genetic diversity present between Chinese hybrid rice parental lines using coefficient of parentage and reported high degree of diversity among the parental lines. Another study carried out by Saidaiah et al. (2010) revealed existence of sufficient genetic variation at DNA level among the parental lines analyzed.

Different types of markers have been used for genetic diversity studies. Some studies employed random SSRs, within random SSRs, EST-SSRs (Jaikishan et al. 2006; Sripathy et al. 2012; Pavani et al. 2018) and (GATA) n motif specific SSRs (Saidaiah et al. 2010) and trait linked markers (Zhang et al. 2010; Yadav et al. 2013). When compared to random markers, trait linked markers showed high PIC values (Yadav et al. 2013). Mahalingam et al. (2016) determined the genetic divergence among 51 restorer and five maintainer lines using 55 SSR markers. Out of them, 37 SSR markers were found to be polymorphic and the number of amplified fragments ranging from one to five. The highest polymorphic information content (PIC) value (more than 0.60) was observed for eight primers and average PIC value was 0.444. Shrivastava et al. (2015) studied genetic diversity of maintainers and restorers using 29 polymorphic SSR markers and grouped them into three clusters. Sripathy et al. (2012) assessed genetic diversity of parental lines of rice hybrids (KRH-2, PSD-1, DRRH-2, CORH-3 and DRRH-3) using genomic and EST-SSR markers and revealed that EST-SSRs are better predictors of genetic diversity with higher PIC values. Yadav et al. (2013) determined the pattern of genetic diversity in terms of both phenotypic and genotypic variability and assessed the efficiency of random QTL linked/gene based simple sequence repeat markers and observed that slightly higher average PIC values for the trait-linked SSRs (0.48) than the random SSRs (0.41).

Based on Mahalanobis D^2-Statistics, total parental lines grouped into 18 clusters, in that five are multigenotypic clusters and 13 are monogenotypic clusters. Morphological clustering was also performed by DARwin software ver 6.0.010 grouped into 7 clusters where 5 major clusters and 2 minor clusters. In both the kind of groupings, IR10198 grouped under monogenotypic cluster. 5 major clusters were almost similar with slight variations. Morphological genetic divergence of hybrid rice parental lines has been studied by many researchers (Nghia et al. 1999; Kulsum et al. 2011; Kiani 2012; Hasan et al. 2012) using Mahalanobis D^2-Statistics.

In the present study SPY exhibited significant positive correlation with NPT, NFG, TG and SF%. These results are in accordance with Ekka et al. (2011) for NFG, Babu et al. (2012) for NPT, Yadav et al. (2013) for NFG, TG and SF%, Vinoth et al. (2016) for NFG, Krishna Naik et al. (2005) and Akhtar et al. (2011) for TG. However, SPY showed positive non-significant correlation with PH, PL, TT, TGW and negative non-significant correlation with DFF (Balakrishnan et al. 2016), NUPT and NUFG. Negative correlations may arise due to repulsion phase linkages between desirable and undesirable traits (Sharma 1988). Selection for high single plant yield results into lower number of unproductive tillers and lower number of unfilled grains. Useful negative correlations may be broken by repeated hybridization between selected parents.

Based on morphological trait data analysis, total parental lines grouped randomly with no particular grouping pattern for fertility restoration, since it wouldn't have been possible to differentiate the restorers and maintainers by its

phenotype. The reason for this kind of association between restorer and maintainer grouping at molecular level might be due to the close or shared ancestry of parental lines or could be attributed to the kind of markers which were used in the present study, which might be linked with fertility restoration trait which can differentiate restorers from maintainers. When we look back to study the background of markers which were present on chromosome 1 and 10, since two major fertility genes *Rf3* and *Rf4* have been mapped on chromosome 1 and 10 respectively (Yao et al. 1997; Zhang et al. 1997). The markers RM171 (OSR 33) is the flanking marker for *Rf4* gene at a distance of 3.7 cM on the long arm of chromosome 10 (Jing et al. 2001; Nematzadeh and Kiani 2010). Namaky et al. (2016)and Xalxo et al. (2017) used RM171 for marker assisted screening of lines for fertility restoration genes and validated by Kiani (2015) and Bhati et al. (2018). In the same way, studies have employed RM474 marker for validation of fertility restoration trait (Sheeba et al. 2009). It has been reported that RM1 marker on chromosome 1 is linked with *Rf3* fertility restoration gene (He et al. 2002). Cai et al. (2014) and Cai and Zhang (2014) employed RM1 marker for detecting *Rf3* and *Rf4* genes in SSSLs (Single segment substitution lines). As per the report of Alavi et al. (2009) *Rf3* gene is flanked by two SSR markers RM1 and RM3873 at distances of 5.6 and 14 cM, respectively. Ahmadikhaha et al. (2007) used RM1 to genotype rice lines at the fertility restoration loci (*Rf*). Li et al. (2008) reported that RM283 was linked to one of the Rf genes on chromosome 1 at a distance of 6.7 cM.

Kinship between pedigree and molecular marker and morphological based grouping

When we observed the correspondence between pedigree and molecular grouping, interestingly we found that RPHR1005 and GQ25 which were derived from the same pedigree, were clustered together at molecular level. However, IBL57 from same cluster, which is also having same pedigree as RPHR1005 and GQ25 was observed to be genetically distant from these two lines. All BK lines have been derived from single cross combination of Improved samba mahsuri/KMR3R//KMR3R. As per molecular grouping all BK lines clustered under a single cluster, however in morphological clustering they dispersed under different clusters. According to morphological trait based clustering, all IR lines grouped under same cluster and majority of the BK lines were grouped into single cluster along with KMR3 which is the common parent of all BK lines and the other parent is Improved Samba Mahsuri which formed separate cluster, as the BK lines were derived through a single round of backcrossing with KMR3R (Hari et al. 2011). When we see the comparability between pedigree and morphological grouping, the trend was similar to that observed molecular grouping where RPHR1005 and GQ25 placed very close to each other with least morphological distance. Restorer lines RPHR1096, RPHR619-2, RPHR612-1, RPHR118, which grouped under same cluster have a common parent BR-827-35.

Correspondence between molecular and morphological clustering

Even though, there is no statistical significance or similarity between molecular and morphological clustering, the pair of lines viz., TCP1221 and TCP1220, RPHR1005 and G25, RPHR-111-3 and RPHR-124 and BK52-104 and BK64-116 grouped together. Similar kind of results were observed by Kaladhar et al. (2004) and Xu et al. (2002) where no correlation was observed between morphological and molecular clustering pattern of maintainer and restorers. However, Yadav et al. (2013) observed correlation between two distance matrices of trait linked marker data and phenotypic data.

Structure analysis of parental lines

Maximum of the parental lines were clustered into respective B and R line groups when we performed model based clustering for structure analysis. Structure analysis showed entire population can be grouped into two subgroups which were in agreement with the distance based clustering. We observed sharp peak of Evanno's delta K at K = 2. Some B lines grouped with R line cluster and vice versa. Structure analysis of hybrid rice parental lines is of much use for heterotic grouping and to broaden the genetic diversity of parental lines (Wang et al. 2016). Alpha value indicates degree of admixture and in this study mean alpha value was 0.1395. Since the value was less than 1 shows an acceptable rate of admixtures. Lower value (less than 1) of alpha shows less number of admixtures, while, an alpha value greater than one indicates that most of the individuals are admixed. When alpha value approaches zero indicates that most of the individuals are from different populations. Wright described F_{st} (Fixation index Sub population within Total population) as one of the parameters to explain genetic structure of a population. F_{st} is used to quantify the genetic divergence between subpopulations. F_{st} is defined as the correlation of gametes within subpopulations relative to gametes drawn at random from the entire population (Wright 1951). The population with F_{st} values that range from 0.1 to 0.3 have been observed as most divergent populations (Cavalli-Sforza et al. 1994). Jakobsson et al. 2013 examined the relationship between F_{st} and major allele frequency and demonstrated that F_{st} values can be restricted by allele frequency distribution. According to

Rosenberg et al. (2002) and Li et al. (2008), F_{st} values are low for SSRs than SNPs. Singh et al. (2016) studied genetic diversity trend in Indian rice varieties and their structure analysis have shown three subpopulations with F_{st} values 0.0118, 0.3240 and 0.2667 for three subgroups respectively. Based on F_{st} values of present study, it was evident that two sub groups (B and R sub groups) were more divergent due to the uniqueness of ancestral genetic material used while breeding B and R lines.

Many studies have been published on understanding the genetic diversity present between parental lines of hybrid rice based on molecular profiling (Jaikishan et al. 2006; Sundaram et al. 2008; Saidaiah et al. 2010; He et al. 2012; Rajendran et al. 2012; Sripathy et al. 2012; Srivasthava et al. 2015; Mahalingam et al. 2016), based on only morphological trait (Kulsum et al. 2011; Hasan et al. 2012; Kiani 2012) and based on both molecular and morphological data (Xu et al. 2002; Kaladhar et al. 2004; Wang et al. 2016). These studies consistently revealed that there are two groups at molecular level B and R groups and this kind of differentiation was constrained at morphological level due to convergent selection for some common traits like high productivity and maturity (Xu et al. 2002). Several studies have been suggesting the immediate need for widening the genetic base of parental lines in order to realize the greater heterosis (Rajendran et al. 2012 and Brar and Khush 2017). From the present study, it was confirmed that at molecular level there are two groups B and R and at morphological level there is no strong factor to differentiate B and R groups and the real genetic relationship between lines might be masked by genotype by environment interaction.

Heterotic grouping in rice

Prediction of hybrid performance without going for development and testing of number of different cross combinations has been the major goal of any hybrid breeding programme using molecular marker data and in combination of molecular marker and morphological data especially in rice where manual emasculation and pollination is very difficult and time consuming. In many cases genetic distances are positively correlated with heterosis. Genetic diversity between parental lines is the major criteria for development of highly heterotic hybrids. Establishment of heterotic groups has been one of the key steps in hybrid crop breeding. In Rice very few studies have been reported (Zheng et al. 1994, 1995; Mackill et al. 1996; Xiao et al. 1996b; Xie et al. 2014; Wang et al. 2015; Beukert et al. 2017) on heterotic grouping concept. Most of the public and private sector breeding is centred on the existing heterotic groups. Although there were many studies on genetic diversity of hybrid rice parental lines and on combining ability studies, the research is not related to heterotic groups concept. The success of rice hybrids solely relay on grouping of parental lines into divergent pools.

Most of the parental lines used in developing rice hybrids so far have the narrow genetic base which resulted in lower magnitude of heterosis. Therefore it is important and urgent need to broaden the genetic base of parental lines by exploiting the novel methods in breeding, such as development of parental lines from inter sub specific combinations (indica x tropical japonica), breeding intermediate parental lines by transferring japonica genome segments into indica back ground by preserving restorer and wide compatibility genes. Identification and introgression of genes for fertility restoration (*Rf*), wide compatibility (WC), biotic and abiotic stress tolerance genes into parental lines, identification of yield QTLs and heterotic loci for greater heterosis and their introgression into parental lines were also essential to accelerate hybrid rice breeding programme. Formation of heterotic groups and patterns for developing high yielding hybrids can also be established and exploited as in maize. A hybrid with high yielding capacity along with potential to yield more F_1 seed, resistance to major pest and diseases and consumer preferred grain quality is most desirable for improving yield potential in rice and meet the global food production demands.

Acknowledgements We are grateful to Director, ICAR-IIRR, Hyderabad for guidance, providing research facilities and funding to conduct the experiment.

Author's contribution AKS and ASH conceptualised the experiment; KS conducted the lab and field experiment; KS and PK collected and compiled data; KS, BD, RKE, VJS, PKB, KKV and GKS molecular and field data analysis; KS, BD, ASH, RMS, PS, PR and KBK, RKE wrote and corrected the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare they have no conflicts of interest.

Human and animal rights This article does not contain any studies with human participants or animals performed by any of the authors.

References

Ahmadikhah A, Karlov G, Nematzadeh GH, Ghasemi Bezdi K (2007) Inheritance of the fertility restoration and genotyping of rice lines at the restoring fertility (Rf) loci using molecular markers. Int J Plant Prod 1:13–21

Akhtar N, Nazir MF, Rabnawaz A, Mahmood T, Safdar ME, Asif M, Rehman A (2011) Estimation of heritability, correlation and path coefficient analysis in fine grain rice (*Oryza sativa* L.). J Anim Plant Sci 21(4):660–664

Akkaya MS, Buyukunal Bal EB (2004) Assessment of genetic variation of bread wheat varieties using microsatellite markers. Euphytica 135:179–185

Alavi M, Ahmadikhah A, Kamkar B, Kalateh M (2009) Mapping Rf3 locus in rice by SSR and CAPS markers. Int J Genet Mol Biol 1(7):121–126

Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell Publishing, Maden

Ariyo OJ (1987) Multivariate analysis and the choice of parents for hybridization in okra (*Abelmoschus esculentus* (L.) Moench). Theor Appl Genet 74:361–363

Babu VR, Shreya K, Dangi KS, Usharani G, Shankar AS (2012) Correlation and path analysis studies in popular rice hybrids of India. Int J Sci Res Publ 2(3):1–5

Balakrishnan D, Subrahmanyam D, Badri J, Raju AK, Rao YV, Beerelli K, Mesapogu S, Surapaneni M, Ponnuswamy R, Padmavathi G, Babu VR, Neelamraju S (2016) Genotype × environment interactions of yield traits in backcross introgression lines derived from *Oryza sativa* cv.Swarna/Oryzanivara. Front Plant Sci 7:1530. https://doi.org/10.3389/fpls.2016.01530

Beukert U, Li Z, Liu G, Zhao Y, Ramachandra N, Mirdita V, Pita F, Pillen K, Rief JC (2017) Genome-based identification of heterotic patterns in rice. Rice 10:22. https://doi.org/10.1186/s12284-017-0163-4

Bhati PK, Singh SK, Kumar U (2018) Screening and validation of fertility restoration genes (Rf) in wild abortive CMS system of rice (*Oryza sativa* L.) using microsatellite markers. Indian J Genet Pl Br 78(2):270–274. https://doi.org/10.5958/0975-6906.2018.00035.4

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

Brar DS, Khush GS (2017) Rice breeding in the genomics era: perspectives. Agric Res J 54(4):612–619

Cai J, Zhang G (2014) The novel methods of development of the maintainer and cytoplasmic male sterile lines with different cytoplasms based on chromosome single-segment substitution lines in rice. Turk J Agric For 38:441–446

Cai J, Dai JZ, Zhu HT, Zeng RZ, Zhang Z, Ma TF, Zhang GQ (2014) Comparative analysis of fertility restoration genes for WA, Y, and DA cytoplasmic male sterility in rice. Biol Plantarum. https://doi.org/10.1007/s10535-013-0385-2

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton

Choukan R, Warburton MI (2005) Use of SSR data to determine relationships among early maturing Iranian maize inbred lines. Maydica 50:163–170

DeWoody JA, Honeycutt RL, Skow LC (1995) Microsatellite markers in white tailed deer. J Hered 86:317–319

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. J Stat Softw 22(4):1–20

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. Conserv Genet Resour 4(2):359–361

Ekka RE, Sarawgi AK, Kanwar RR (2011) Correlation and path analysis in traditional rice accessions of Chhattisgarh. J Rice Res 4(1&2):11–18

Hari Prasad AS, Senguttuvel P, Revathi P, Kemparaju KB, Sruthi K, Sundaram RM, Seshu Madhav M, Prasad MS, Laha GS (2018) Breeding strategies for hybrid rice parental line improvement. Oryza 55(spl):38–41

Hari Y, Srinivasarao K, Viraktamath BC, Hariprasad AS, Laha GS, Ahmed MI, Natarajkumar P, Ramesha MS, Neeraja CN, Balachandran SM, Rani NS (2011) Marker-assisted improvement of a stable restorer line, KMR-3R and its derived hybrid KRH2 for bacterial blight resistance and grain quality. Plant Breed 130(6):608–616

Hasan MJ, Kulsum U, Rahman MMH, Chowdhury MMH, Chowdhury AZMKA (2012) Genetic diversity analysis of parental lines for hybrid development in rice (*Oryza sativa* L.). Bangladesh J Agric Res 37(4):617–624

He GH, Wang WM, Liu GQ, Hou L, Xiao YH, Tang M, Yang ZL, Pei Y (2002) Mapping of two fertility-restoring gene for WA cytoplasmic male sterility in minghui63 using SSR markers. Acta Genet Sin 29(9):798–802

He ZZ, Xie FM, Chen LY, Dela Paz MA (2012) Genetic diversity of tropical hybrid rice germplasm measured by molecular markers. Rice Sci 19(3):193–201. https://doi.org/10.1016/S1672-6308(12)60040-7

IRRI (2013) Standard evaluation system (SES) for rice, 5th edn. IRRI, Philippines

Jaikishan I, Ramesha MS, Rajendrakumar P, Rao S, Neeraja CN, Balachandran SM, Viraktamath BC, Sujatha K, Sundaram RM (2006) Characterization of genetic diversity in hybrid rice parental lines using EST-derived and non-EST SSR markers. Rice Genet Newsl 23:24–28

Jakobsson M, Edge MD, Rosenberg NA (2013) The relationship between FST and the frequency of the most frequent allele. Genetics 193(2):515–528. https://doi.org/10.1534/genetics.112.144758

Jing R, Li X, Yi P, Zhu Y (2001) Mapping fertility-restoring genes of rice WA cytoplasmic male sterility using SSLP markers. Bot Bull Acad Sinica 42:167–171

Julfiquar AW, Virmani SS, Carpena AL (1985) Genetic divergence among some maintainer and restorer lines in relation to hybrid breeding in rice (*Oryza sativa* L.). Theor Appl Genet 70(6):671–678

Kaladhar K, Ramesha MS, Ramakrishna S, Ahmed MI, Viraktamath BC, Sarla N (2004) Clustering of maintainer and restorer lines of rice based on morphological and molecular diversity. Rice Genet Newsl 21:27–28

Kiani G (2012) Diversity assessment among some restorer lines using agronomic traits in rice (*Oryza sativa* L.). Biharean Biologist 6(1):1–4

Kiani G (2015) Validation of SSR markers linked to restoring fertility (Rf) genes and genotyping of rice lines at Rf loci. J Agr Sci Tech 17(20):1931–1938

Krishna Naik R, Sreenivasulu Reddy P, Ramana JV, Srinivasa Rao V (2005) Correlation and path co-efficient analysis in rice (*Oryza sativa* L.). Andhra Agric J 52(1–2):52–55

Kulsum U, Hasan MdJ, Begum H, Billah MM, Rahman H (2011) Genetic diversity of some restorer lines for hybrid rice development. Bangladesh J Agric Res 36:21–28. https://doi.org/10.3329/bjar.v36i1.9226

Li LJ, Zhou HP, Zhan XD, Cheng SH, Cao LY (2008) Mapping of rice fertility-restoring genes for ID-type cytoplasmic male sterility in a restorer line R68. Rice Sci 15:157–160. https://doi.org/10.1016/S1672-6308(08)60036-0

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21(9):2128–2129

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

Mahalingam A, Sarawathi R, Ramalingam J (2016) SSR marker diversity analysis for WA cytoplasm based fertility restorer genes of hybrid rice (*Oryza sativa* L.). J Innov Agric 3(1):19–27

Maranho RC, Augusto R, Mangolin CA, Machado MFPS (2014) Use of differential levels of mean observed heterozygosity in microsatellite loci of commercial varieties of sugarcane (Saccharum spp.). Genet Mol Res 13(4):10130–10141

Nair NV, Ballakrishnan R, Screenivasan TV (1998) Variability for quantitative traits in exotic hybrid germplasm of sugarcane. Genet Resour Crop Evol 45:459–464

Namaky RE, Sedeek S, Moukoumbi YD, Ortiz R, Manneh B (2016) Microsatellite-aided screening for fertility restoration genes (Rf) facilitates hybrid improvement. Rice Sci 23(2):160–164. https://doi.org/10.1016/j.rsci.2016.04.003

Nematzadeh A, Kiani G (2010) Genetic analysis of fertility restoration genes for WA type cytoplasmic male sterility in Iranian restorer rice line DN-33-18. Afr J Biotechnol 9:6273–6277

Nghia PK, Malik JPS, Pandey MP, Singh NK (1999) Genetic distance analysis of hybrid rice parental lines based on morphological traits and RAPD markers. Omonrice 7:57–69

Pavani M, Sundaram RM, Ramesha MS, Kavi Kishor PB, Kemparaju KB (2018) Prediction of heterosis in rice based on divergence of morphological and molecular markers. J Genet 97(5):1263–1279

Peakall R, Smouse P (2012) GenAlEx 65: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28(19):2537–2539

Perrier X, Jacquemoud-Collet JP. DARwin software (2006) http://darwin.cirad.fr/. Accessed 30 Sept 2015

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

Rajendran N, Mukherjee L, Kamalnath KR, Shashidhar HE (2012) DNA fingerprinting and estimation of genetic diversity among hybrid rice parental lines (*Oryza sativa* L.) using simple sequence repeats (SSR) markers. J Plant Breed Crop Sci 4(11):169–174

Rao CR (1952) Advanced statistical methods in biometric research. Wiley, New York (NY)

R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN: 3-900051-07-0

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphism in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8019

Saidaiah P, Ramesha MS, Sundaram RM, Hari Y (2010) Molecular profiling of parents of some released rice hybrids using SSR markers. Indian J Plant Genet Resour 23(3):269–275

Sharma JR (1988) Statistical and biometrical techniques in plant breeding. New Age International (P) limited, Publishers, New Delhi

Sheeba KN, Viraktamath B, Siva Sivaramakrishnan, Gangashetti MG, Pawan Khera, Sundaram RM (2009) Validation of molecular markers linked to fertility restorer gene (s) for WA-CMS lines of rice. Euphytica 167:217–227. https://doi.org/10.1007/s10681-008-9865-4

Shrivastava A, Mishra D, Koutu G, Singh S, Prakash V, Sohgaura N (2015) Assessment of genetic diversity of restorer and maintainer lines of rice using SSR markers. The Ecascan. 9(1–2):265–269

Singh VK, Upadhyay P, Sinha P, Mall AK, Jaiswal SK, Singh A, Ellur RK, Biradar S, Sundaram RM, Singh S, Ahmed I, Mishra B, Singh AK, Kole C (2011) Determination of genetic relationships among elite thermosensitive genic male sterile lines (TGMS) of rice (*Oryza sativa* L.) employing morphological and simple sequence repeat (SSR) markers. J Genet 90(1):1–19

Singh N, Choudhury DR, Tiwari G, Singh AK, Kumar S, Srinivasan K, Tyagi RK, Sharma AD, Singh NK, Singh R (2016) Genetic diversity trend in Indian rice varieties: an analysis using SSR markers. BMC Genet 17(1):127

Siwach P, Jain S, Saini N, Chowdhury VK, Jain RK (2004) Allelic diversity among Basmati and non-Basmati long grain indica rice varieties using microsatellite markers. J Plant Biochem Biot 13:25–32

Sripathy KV, Hosamani J, Bellundagi A (2012) Assessment of genetic diversity in hybrid rice parental lines using genomic and EST-SSR markers. Bioinfolet 9(1):23–28

Sundaram RM, Naveenkumar B, Biradar SK, Balachandran SM, Mishra B, IlyasAhmed M, Viraktamath BC, Ramesha MS, Sarma NP (2008) Identification of informative SSR markers capable of distinguishing hybrid rice parental lines and their utilization in seed purity assessment. Euphytica 163(2):215–224

Surapaneni M, Balakrishnan D, Mesapogu S, Krishnam Raju A, Rao YV, Neelamraju S (2016) Genetic characterization and population structure of Indian rice cultivars and wild genotypes using core set markers. Biotech 6(1):95. https://doi.org/10.1007/s13205-016-0409-7

Thioulouse J, Chessel D, Dolédec S, Olivier JM (1997) ADE-4: a multivariate analysis and graphical display software. Stat Comput 7:75–83

Tiwari KK, Singh A, Pattnaik S, Sandhu M, Kaur S, Jain S, Tiwari S, Mehrotra S, Anumalla M, Samal R, Bhardwaj J (2015) Identification of a diverse mini-core panel of Indian rice germplasm based on genotyping using microsatellite markers. Plant Breed 134(2):164–171

Vinoth R, Shivramakrishnan R, Sivaji M, Tamilkumar P, Kumar B, Shailesh M (2016) Genetic analysis and correlation studies for grain yield in rice (*Oryza sativa* L.) under the Allahabad agro climatic region. Int J For Crop Improv 7(1):93–100

Virmani SS (1996) Hybrid rice. Adv Agron 57:328–462

Wang S, Lu ZM (2006) 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 K, Qiu F, Larazo W, Angelita MDP, Xie F (2015) Heterotic groups of tropical indica rice germplasm. Theor Appl Genet 128:421–430. https://doi.org/10.1007/s00122-014-2441-5

Wang I, Zheng Y, Cai Q, Liao C, Mao X, Xie H, Zhu Y, Lian L, Luo X, Xie Zhang J (2016) Population structure and association analysis of yield and grain quality traits in hybrid rice primal parental lines. Euphytica 212:261. https://doi.org/10.1007/s10681-016-1766-3

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354

Xalxo A, Chaudhari PR, Sharma D, Saxena RR, Singh S, Tiwari A (2017) Identification of rice hybrids and restorer line using microsatellite markers. Int J Current Microbiol Appl Sci 6(11):3405–3411

Xiao CL, Virmani SS, Viraktamath BC (1996a) Improving parental lines to increase efficiency of hybrid rice breeding: some new approaches-Chapter 10 in Advances in hybrid rice technology. In: Virmani SS, Siddiq, EA, Muralidharan K (eds) Proceedings of the 3rd international symposium on Hybrid Rice. 14–16 November, Hyderabad, India

Xiao J, Li J, Yuan L, McCouch SR, Tanksley SD (1996b) Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor Appl Genet 92:637–643

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

Xie F, He Z, Esguerra MQ, Qiu F, Ramanathan V (2014) Determination of heterotic groups for tropical indica hybrid rice germplasms. Theor Appl Genet 127:407

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

Yadav S, Singh AK, Singh MR, Goel N, Vinod KK, Mohapatra T (2013) Assessment of genetic diversity in Indian rice germplasm (*Oryza sativa* L.): use of random versus trait-linked microsatellite markers. J Genet 92:545–557

Yao FY, Xu CG, Yu SB, Li JX, Gao YJ, Li XH, Zhang QF (1997) Mapping and genetic analysis of two fertility restorer loci in the wild abortive cytoplasmic male sterility system of rice (*Oryza sativa* L.). Euphytica 98:183–187

Zhang Q, Gao YT, Yang SH, Ragab RA, Saghai Maroof MA, Liz B (1994) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor Appl Genet 89:185–192

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

Zhang G, Bharaj TS, Lu Y, Virmani SS, Huang N (1997) Mapping of the Rf-3 nuclear fertility-restoring gene for WA cytoplasmic male sterility in rice using RAPD and RFLP markers. Theor Appl Genet 94(1):27–33

Zhang T, Ni XL, Jiang KF, Yang QH, Li Yang, Wan XQ, Cao YJ, Zheng JK (2010) Correlation between heterosis and genetic distance based on molecular markers of functional genes in rice. Rice Sci 17(4):288

Zheng K, Qian H, Shen B, Zhuang J, Liu H, Lu J (1994) RFLP-based phylogenetic analysis of wide compatibility varieties in *Oryza sativa* L. Theor Appl Genet 88:65–69

Affiliations

K. Sruthi[1] · B. Divya[3] · P. Senguttuvel[1] · P. Revathi[1] · K. B. Kemparaju[1] · P. Koteswararao[1] · R. M. Sundaram[4] · Vikram Jeet Singh[2] · E. Ranjith Kumar[2] · Prolay Kumar Bhowmick[2] · K. K. Vinod[2] · S. Gopala Krishnan[2] · A. K. Singh[2] · A. S. Hari Prasad[1]

✉ A. S. Hari Prasad
hariprasad34@gmail.com

K. Sruthi
kanneboinasruthi30@gmail.com

B. Divya
divyabalakrishnan05@gmail.com

P. Senguttuvel
senguttuvel@gmail.com

P. Revathi
revathi.ponnusamy@gmail.com

K. B. Kemparaju
kbkemparaju@gmail.com

P. Koteswararao
Pkrao13@gmail.com

R. M. Sundaram
rms_28@rediffmail.com

Vikram Jeet Singh
Jeet2012vikram@gmail.com

E. Ranjith Kumar
ranjithellur@gmail.com

Prolay Kumar Bhowmick
prolaybhowmick@iari.res.in

K. K. Vinod
kkvinod@iari.res.in

S. Gopala Krishnan
gopalkrish@iari.res.in

A. K. Singh
aksingh@iari.res.in

[1] Crop Improvement Section, ICAR-Indian Institute of Rice Research, Rajendranagar, Hyderabad 500030, India

[2] Genetics Division, ICAR-Indian Agricultural Research Institute, New Delhi, India

[3] ICAR- National Professor Project, ICAR- Indian Institute of Rice Research, Hyderabad, India

[4] Biotechnology, Crop Improvement Section, ICAR-IIRR, Rajendranagar, Hyderabad 500030, India