Physiol Mol Biol Plants
https://doi.org/10.1007/s12298-022-01190-8

RESEARCH ARTICLE

Unravelling genetic architecture and development of core set from elite rice lines using yield-related candidate gene markers

Rameswar Prasad Sah[1] · Sasmita Behera[1] · Sushant Kumar Dash[1] · T. P. Muhammed Azharudheen[1] · Jitendriya Meher[1] · Awadhesh Kumar[2] · Bishnu Charan Marndi[1] · Meera Kumari Kar[1] · H. N. Subudhi[1] · C. Anilkumar[1]

Received: 23 July 2021 / Revised: 8 May 2022 / Accepted: 17 May 2022
© Prof. H.S. Srivastava Foundation for Science and Society 2022

Abstract Assessing genetic diversity and development of a core set of elite breeding lines is a prerequisite for selective hybridization programes intended to improve the yield potential in rice. In the present study, the genetic diversity of newly developed elite lines derived from *indicax tropical japonica* and *indicax indica* crosses were estimated by 38 reported molecular markers. The markers used in the study consist of 24 gene-based and 14 random markers related to grain yield-related QTLs distributed across the rice genome. Genotypic characterization was carried out to determine the genetic similarities between the elite lines. In total, 75 alleles were found using 38 polymorphic markers, with polymorphism information content ranging from 0.10 to 0.51 with an average of 0.35. The genotypes were divided into three groups based on cluster analysis, structure analysis and also dispersed throughout the quadrangle of PCA, but nitrogen responsive lines clustered in one quadrangle. Seven markers (GS3_RGS1, GS3_RGS2, GS5_Indel1, Ghd 7_05SNP, RM 12289, RM 23065 and RM 25457) exhibited PIC values $\geq$ 0.50 indicating that they were effective in detecting genetic relationships among elite rice. Additionally, a core set of 11 elite lines was made from 96 lines in order to downsize the diversity of the original population into a small set for parental selection. In general, the genetic information collected in this work will aid in the study of grain yield traits at molecular level for other sets of rice genotypes and for selecting diverse elite lines to develop a strong crossing programme in rice.

Keywords Diversity · Population structure · Rice · Grain yield

Rameswar Prasad Sah and Sasmita Behera have contributed equally to this work.

✉ T. P. Muhammed Azharudheen
tpazhar@gmail.com

Rameswar Prasad Sah
ramesh.pbg@gmail.com

Sasmita Behera
sasmitabehera7575@gmail.com

Sushant Kumar Dash
skdash139@gmail.com

Jitendriya Meher
jmehercrri@gmail.com

Awadhesh Kumar
awadh_iari@yahoo.com

Bishnu Charan Marndi
Bishnu.Marndi@icar.gov.in

Meera Kumari Kar
meerakkar@gmail.com

H. N. Subudhi
dr_hatanath_subudhi@yahoo.co.in

C. Anilkumar
anilcgpb@gmail.com

[1] Crop Improvement Division, ICAR-National Rice Research Institute, Cuttack, Odisha, India

[2] Crop Physiology and Biochemistry, ICAR-National Rice Research Institute, Cuttack, Odisha, India

Introduction

Rice is the world's most widely consumed staple food crop. To ensure global food security, it is critical to maintain a steady increase in rice productivity that keeps up with the growing human population (Bandumula 2018). The genetic enhancement in any crop, including rice, is a function of the genetic diversity available in the gene pool. However, during the last

Published online: 08 July 2022

five decades, the use of only a few parental genotypes, such as Jaya and IR8, in rice breeding (Subudhi et al. 2020) has resulted in genetic erosion and a narrow genetic base among today's cultivated high yielding rice varieties (Choudhary et al. 2013). The rice genetic base needs to be broadened by introgressing novel alleles of genes/QTLs associated with yield and yield attributing traits in rice. Gene-based markers can be used to investigate the allelic diversity of yield influencing genes/QTLs and the novel alleles could be introgressed in the elite genetic backgrounds by inter-specific and inter-sub-specific hybridization programes. The conventional breeding approaches to select parents for hybridization involve testing for the general combining ability of a line or selective mating for trait introgression. The modern breeding strategies or Transformative Rice Breeding (TRB) strategies suggest using only elite parents in crossing programes to accrue some major and thousands of minor yield-related alleles in the large segregating population (Collard et al. 2019). Speed breeding is one of the steps in such a strategy that involves elite × elite crosses followed by rapid generation advancement through the single seed descent (SSD) method (Akhtar et al. 2010; Shivakumar et al. 2018). Thus, recently, plant breeders are focused on developing new elite breeding lines for product development that can replace existing old mega varieties like Swarna, MTU-1010, Swarna Sub1, Pooja, etc. of Eastern India in terms of superiority, higher yield, good stability along with basic traits of the mega varieties. The elite lines chosen for hybridization should carry alleles for grain yield and yield contributing traits which upon hybridization (elite × elite) accumulate in derived progeny and express in a cumulative manner to produce better phenotype (Collard et al. 2019). Such populations also possess higher buffering against fluctuations in micro- and macro-environmental changes due to the presence of diverse allelic combinations.

Genotyping is always advantageous to review the strength of a parental line since it discerns the genetic information of the line. Molecular markers could be used to assess the genetic diversity in rice germplasm resources and to develop heterotic pools (Katara et al. 2021). However, the genetic diversity estimated based on non-genic random DNA markers are not useful in the selection of parental genotypes in the hybridization programe as it does not represent the diversity for yield and yield attributing traits in the germplasm. Hence to increase the accuracy of parental selection, trait-specific molecular markers need to be used (Molla et al. 2019) to capture maximum grain yield-related alleles and to introgress them in a single parental background. This would help to plan rapid and effective breeding strategies to increase genetic gain per breeding cycle, and improve the grain yield and climate resilience (Atlin et al. 2017; Gobu et al. 2020). Furthermore, using these markers to develop a fine core set from a large population has been proven to save resources while keeping the same population structure and variety as the original population. Hence the genotyping by grain yield attributing gene/QTL/pleiotropic gene-based markers for analyzing the strength of elite breeding materials is a prelude to the selection of parents with high genetic value. In the present study, we have used molecular markers designed based on the genes governing important yield genes/QTLs to assess the diversity and population structure of a large set of elite rice breeding lines with > 5.0 t/ha grain yield under Eastern Indian situations which are identified from the different breeding programe. Based on the allelic diversity and the phenotypic performance of the breeding lines, the number is then reduced to a manageable pool of parental genotypes. The information provided in the study would help rice breeders to select appropriate elite lines as parents for a breeding programe.

Materials and method

Plant materials

The study involved 314 rice breeding lines from different breeding programes and varieties of different ecologies. All the 314 lines were evaluated for yield *per se* in the wet season, 2017 and 2018 at the experimental plots, ICAR-National Rice Research Institute (20.45° N, 85.93° E), Cuttack, Odisha, India. The experiments were laid out in augmented block design to accommodate a large number of genetic materials at initial testing. Each line was sown in 4 rows and each row had 30 hills (plot size 3.6 m^2 per line). On the basis of grain yield results of all 314 lines, a set of 119 elite lines were selected on the basis of average grain yield greater than 5 t/ha across the years. The selected 119 lines were evaluated in two replications for yield trial during the wet season of 2019. Each line was sown in 6 rows and each row had 30 hills (plot size 5.6 m^2). Out of 119 elite lines, 96 elite lines were chosen on the basis of maturity duration between 125 and 140 days of maturity. As a result of the primary filters based on yield criteria and maturity duration, the total sample size was reduced from 314 to the best 96 after being evaluated over 3 years. All the 96 lines were used for molecular study and listed in supplementary table 1. These 96 elite rice lines included 60 new generation rice (NGR) (*indica* × *tropical japonica*), 11 new plant type lines (NPT), 14 advance breeding lines (ABL) (*indica* × *indica*), 4 high nitrogen responsive lines (NRL) (*indica* × *indica*) and 7 released varieties (*indica* × *indica*) constituted experimental material for the present study.

In all years of evaluation, 25 days old seedlings were transplanted in the first week of August. The crop was applied with a fertilizer dose of 80 N:40 P_2O_5:40K_2O as per recommended. The evaluation in all the years was in the wet season, where sufficient rainfall allowed the crop to raise successfully.

Marker assay

Genomic DNA extraction

The seedlings of 96 elite lines were raised in the seedling tray for leaf collection. Young leaves from 15 days old plants were collected from the seedling tray for isolation of genomic DNA. The DNA was extracted through CTAB method (Doyle and Doyle 1990) and dissolved in TE buffer (Tris EDTA pH-8). The concentration and quality of DNA were measured using Spectrophotometer Nanodrop (TM 1000/1000c).

Marker and PCR amplification

A total of 38 reported molecular markers for grain yield, distributed across the rice genome were used for the study (Supplementary table 2, Fig. 1). These markers included 24 gene-based and 14 random markers related to grain yield-related QTLs/genes in rice. The QTLs for yield attributing traits viz; grain size, grain shape, grain weight, grain width, grain number, grain filling, spikelet number, panicle branching, dense panicle, narrow leaf, etc. were considered while selecting markers for the study (Table 1). The primers were synthesized at Eurofins Scientific, (USA). Polymerase chain reactions (PCR) were carried out in an Agilent Thermal cycler (Sure Cycler 8000) with a 10 µl reaction mixture using 96-well plates. PCR master mix (Xceleris 2 × premix taq version 2.0, Xceleris Genomics, Ahmedabad, India) was used for better amplification. An amount of 1 µl genomic DNA (concentration 20 ng/µl) and 1 µl each of forward and reverse primers were added to the PCR mix for DNA amplification. The amplification in PCR was: initial denaturation at 94 °C for 4 min, denaturing at 94 °C for 40 s, annealing at 56 °C (annealing temperatures for all primer) for 40 s and extension at 72 °C for 3 min. The PCR cycle was repeated for 35 cycles and a final extension was set at 72 °C for 7 min was carried out. The reactions were then set up to hold at 4 °C until electrophoresis and visualization.

Amplicon visualization

Gel electrophoresis was done in Agarose gel (3.5%) using Genaxy Maxi horizontal tank, in 0.5% Tris Boric Acid

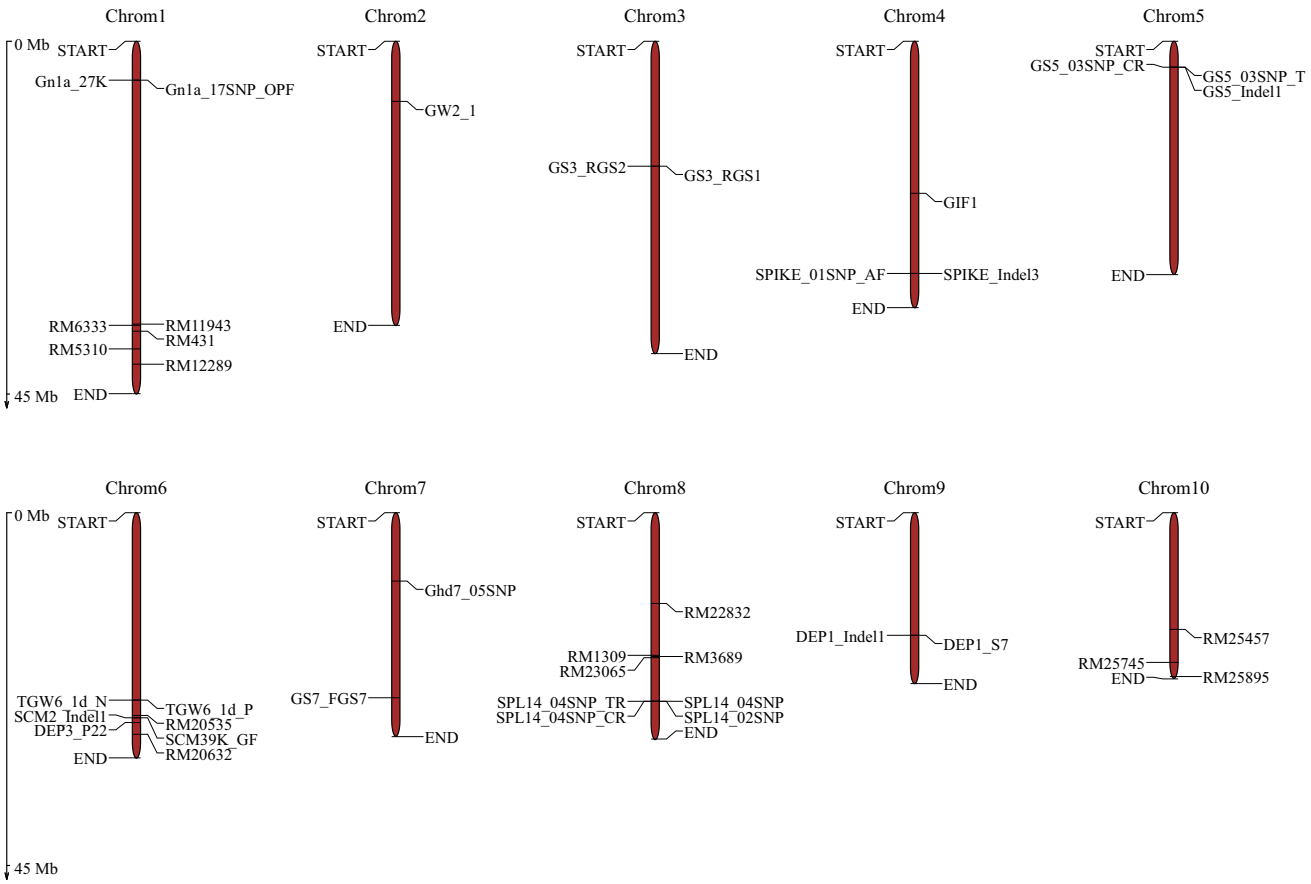

Fig. 1 Distribution of 38 markers across the rice chromosome. Chrom: Chromosome; START: Start point of the chromosome; END: End point of the chromosome, Mb: Million basepairs

Table 1 List of the marker, gene, associated trait and relation of this marker with traits

Gene/QTL	No. of Marker	Marker	Marker type	PIC	Polymorphism (%)	Heterozygosity	Observed Size Range (bp)	Chromosome no	Mechanism of Gene/relation of marker with traits	Trait associated	References
GS3	1	GS3_RGS1	SSR	0.50	48.96	0.36	900–1000	3	A series of molecular signals that proceeds with an activated receptor promoting the exchange of GDP (Guanosine diphosphate) for GTP (Guanosine Triphosphate) on the alpha-subunit of an associated heterotrimeric G-protein complex	Grain weight and grain length	Wang et al. (2011)
	2	GS3_RGS2	SSR	0.51	45.31	0.48	280–290				
GW2	3	GW2-1	SSR	0.47	62.50	0.43	180	2	Grain Width 2 (GW2), a RING-type E3 ubiquitin ligase, can control seed development by catalysing the ubiquitination of expansin-like 1 (EXPLA1), a cell wall-loosening protein that increases cell growth	Grain width and weight, a positive regulator of grain size by regulating grain width, filling and weight	Song et al. (2007)
GS5	4	GS5-03SNP-TF	SNP	0.29	17.50	0.22	250–300	5	This gene encodes a putative serine carboxypeptidase and function as a positive regulator of grain size, such that higher expression of GS5 is correlated with larger grain size	Grain size	Kim et al. (2016)
	5	GS5-03SNP-CR	SNP	0.36	23.96	0.30	50–70	5			
	6	GS5-indel1	InDel	0.50	47.92	0.19	50–60	5			
TGW6	7	TGW6-1d p	SNP	48.96	0.58	0.16	430	6	TGW6 encodes a novel protein with indole-3-acetic acid (IAA)-glucose hydrolase activity. The Kasalath allele contained six nucleotide substitutions. and a 1-bp deletion at nucleotide 313 compared to the *Nippon bare* allele. *Nipponbare* TGW6 protein hydrolysed IAA-glucose into IAA and glucose, but Kassala TGW6 did not	1000 Grain number	Kim et al. (2016)
	8	TGW6-1d-N	SNP	47.36	0.49	0.14	380				
DEP1	9	DEP1_S7	SSR	0.25	14.38	0.19	100–400	9	Dense and Erect panicle 1-this gene correspond to the erect panicle architecture shows a pleiotropic effect in increasing grain yield and nitrogen efficiency in rice	Dence and erect panicle1	Huang et al. (2009) and Kim et al. (2016)
	10	DEP1-indel1	InDel	0.44	32.64	0.11	200–320	9			

Table 1 continued

Gene/QTL	No. of Marker	Marker	Marker type	PIC	Polymorphism (%)	Heterozygosity	Observed Size Range (bp)	Chromosome no	Mechanism of Gene/relation of marker with traits	Trait associated	References
DEP3	11	DEP3_P22	SSR	0.10	90.00	0.00	180	6	The Gγ protein like domain strongly interacts with *RGB1*, and the VWFC (von Wilebrand factor typeC) domain at the C-terminus may participate in the association between Gγ and Gα. The *dep1* allele inhibits nitrogen responses through interactions with the Gβ and Gα subunits, and reduced Gα or enhanced Gβ activity inhibits nitrogen responses	Dense and erect panicle 3, DEP3, which confers high grain yield in rice	Qiao et al. (2011)
GS7	12	GS7_FGS7	SSR	0.47	38.54	0.23	210–360	7	GS7 acts differently from any of the other related QTLsGW2, GS3, qGW5 and GS5). Moreover, the alignment of the cv. D50 and cv. HB277 ORF2 sequences suggest that its nucleotide variation is responsible for variation in grain shape	Grain shape	Shao et al. (2012)
Gn1a	13	Gn1a_17SNP-OPF	SNP	0.49	42.19	0.00	50–650	1	Gn1a (OsCKX2), which encodes cytokinin oxidase/dehydrogenase, plays an important role in regulating rice grain yield	Grain number, plant height and heading date 7	Kim et al. (2016)
	14	Gn1a_27K	SSR	0.46	35.94	0.34	810–820	1			
OsSPL14/ WFP	15	SPL14-02SNP	SNP	0.16	91.67	0.18	500	8	Squamosa promoter binding protein-like 14, Higher expression of *OsSPL14* in the reproductive stage promotes panicle branching and higher grain yield in rice. *OsSPL14* controls shoot branching in the vegetative stage and is affected by microRNA excision. We also demonstrate the feasibility of using the *OsSLP14*.^{WFP} allele to increase rice crop yield	Promotes Panicle branching and higher grain productivity in rice. wealthy farmer's panicle	Kim et al. (2016)
	16	SPL14-04SNP	SNP	0.42	30.21	0.28	280	8			
	17	SPL14-04SNP-CR	SNP	0.10	94.79	0.10	280	8			
	18	SPL14-04SNP-TR	SNP	0.42	63.54	0.43	290	8			
SCM2/ APO1	19	SCM2-indel1	InDel	0.43	31.25	0.43	110–120	6	*SCM2* increased both the width and height of the meristem, Consequently, producing larger internode cells. *SCM2* had an additive effect on the number of parenchyma cells in culm internodes	Strong culm, aberrant panicle organization	Kim et al. (2016)
OsFbox321 Os_F0393	20	SCM 39 K GF	InDel	0.40	27.60	0.40	60–70	6			

Table 1 continued

Gene/QTL	No. of Marker	Marker	Marker type	PIC	Polymorphism (%)	Heterozygosity	Observed Size Range (bp)	Chromosome no	Mechanism of Gene/relation of marker with traits	Trait associated	References
Ghd7-05SNP	21	Ghd7-05SNP	SNP	0.50	47.92	0.22	200–210	7	The *Ghd7* encoding a CCT domain protein is involved in the regulation of heading date, plant height, and grain number per panicle. Fully functional alleles *Ghd7-1* delayed heading date and increased plant height and yield	Grain number, plant height and heading date	Kim et al. (2016)
GIF1	22	GIF1	SSR	0.43	30.21	0.14	400–700	4	The GIF1 gene shows a restricted expression pattern during grain-filling Ectopic expression of the cultivated GIF1 gene with the 35S or rice Waxy promoter resulted in smaller grains, whereas overexpression of GIF1 driven by its native promoter increased grain production	Grain incomplete filling	Wang et al. (2008)
NAL1/ SPIKE	23	SPIKE-01SNP-AF	SNP	0.20	11.46	0.17	220–320	4	Nal1 controls plant architecture through the regulation of genes involved in the photosynthetic apparatus, cell cycle, and GA and BR signaling pathways	Spikelet number per panicle	Kim et al. (2016)
	24	SPIKE-indel3	InDel	0.24	0.37	0.21		4			
RM12289	25	RM12289	SSR	0.50	45.31	0.35	100–110	1	Candidate gene annotation within QTLs suggested the role of transcription factor and genes involved in osmotic potential regulation through catalytic/ metabolic pathway in drought tolerance mechanism contributing to yield. These markers come under qGYDS1.1/DSR Grain yield QTLs	Grain Yield QTL under DSR	Sandhu et al. (2015)
RM11943	26	RM11943	SSR	0.49	44.27	0.28	60–100	1			
RM6333	27	RM6333	SSR	0.39	73.96	0.36	200	1			
RM431	28	RM431	SSR	0.26	15.63	0.25	490	1			
RM5310	29	RM5310	SSR	0.38	75.00	0.39	100	1			
RM20535	30	RM20535	SSR	0.37	24.22	0.29	120–170	6	These markers are linked to qGYDS6.1 QTL	Grain Yield QTL under DSR	Sandhu et al. (2015)
RM20632	31	RM20632	SSR	0.47	38.02	0.31	130–180	6			
RM22832	32	RM22832	SSR	0.12	93.75	0.16	90–93	8	These markers are linked to qGYDS8.1 QTL	Grain Yield QTL under DSR	Sandhu et al. (2015)
RM1309	33	RM1309	SSR	0.08	95.83	0.07	60–65	8			
RM3689	34	RM3689	SSR	0.15	8.33	0.14	300–310	8			
RM23065	35	RM23065	SSR	0.50	54.17	0.37	141–150	8			
RM25457	36	RM25457	SSR	0.50	53.13	0.43	110–115	10	These markers are linked to qGYDS10.1 QTL	Grain Yield QTL under DSR	Sandhu et al. (2015)
RM25895	37	RM25895	SSR	0.10	94.79	0.13	120–125	10			
RM25745	38	RM25745	SSR	0.23	13.02	0.19	200–210	10			

QTL quantitative trait loci, *DSR* direct seeded rice, *RM* random marker

Ethylene Diamine Tetra Acetic acid (TBE) running buffer. The agarose was stained with ethidium bromide (23 µl for 550 ml) dye. Electrophoresis was set at 110 V for 240 min using 10 µl of the amplified PCR products. Then molecular ladder of 100 bp obtained from MP Biomedical (Cambridge, UK) was used to compare and estimate the molecular size of the amplified products. The gel was visualized and photographed on Alpha Imager Gel Documentation System.

Data analysis

The DNA amplicons on the gel were scored based on the fragment length of each allele. Alleles were scored as present (1) or absent (0) depending upon their amplification. Amplicon sizes for each marker were scored as a/b where 'a' was homozygote of one kind and 'b' was homozygote of another kind among genotypes. GenAlEx6.502 software (Peakall and Smouse 2006) was used to analyze allele frequency, allele number per locus, gene diversity, heterozygosity for the marker and populations derived from STRUCTURE analysis. Analysis of molecular variance (AMOVA) was analyzed to distinguish the percentage of molecular genetic variance within and among populations classified STRUCTURE analysis using GenAlEx6.502 software. Polymorphic Information Content (PIC) was estimated in Excel 2016. For population structure analysis, data from 38 polymorphic markers were used, and the analysis was done based on the admixture model clustering method utilizing the software package STRUCTURE 2.3.4 (Pritchard et al. 2000). This model was run by assuming the populations (K) from 1 to 10 with at least 5 iterations for each K. A burn-in period of 100,000 and Markov Chain Monte Carlo (MCMC) and replications of 20,0000 after each burn-in were set for analysis. The optimum population (K) which best estimated the structure of the 96 elite lines was predicted using Evanno's method through web-based software STRUCTURE HARVESTER (http://taylor0.biology.ucla.edu/structureHarvester/). The lines were assigned to each subpopulation based on their probability of association of 70% to each of the groups, the lines with a probability of association $< 70\%$ were considered as admixtures. Cluster analysis was carried out in DARwin 6 (Perrier and Jacquemoud-Collet 2006) by neighbour-joining algorithm. To ensure the reliability of the results, 1000 bootstraps were performed in the construction of the dendrogram. The dendrogram was constructed based on the genetic distance calculated using allele frequencies between pairs of individuals in the population. The threshold bootstrap value accepted for the construction of the dendrogram was 60. Principal Component Analysis (PCA) and Principal Coordinate Analysis (PCoA) were done to classify the elite lines into different groups using PAST 4.0 software (https://www.nhm.uio.no/english/research/infrastructure/past/). The allele frequencies between individuals in the population were used for PCA and PCoA analysis. A combined analysis of genotypes was done with a common set of primers using graphical genotyping software GGT 2.0 (Berloo 2008) for graphical representation of variations in genotypes at each chromosome. Later, we intended to downsize the number of genotypes from 96, based on the concept of a core collection that would contain as much genetic diversity as the original set (Vaughan 1991) for harnessing the best out of the parental pool. The core collection using molecular data represented a full coverage of all the alleles existing in the entire elite pool and a heuristic approach of maximum allele retention and higher yield potential was followed to develop a fine core set from the population of 96 genotypes (Chung et al. 2009).

Results

Population screening

Initially, a total of 314 individuals were evaluated for yield trait over two seasons. Based on yield performance recorded over 2 years population size was reduced by 119 individuals by rejecting individuals with yielding ability less than 5 t/ha. The mean yield performance of selected and rejected lines from the original population was compared using a simple '*t*' statistic. The yield performance of the selected 119 lines was found significantly better than the average performance of 195 rejected lines. However, information on days to maturity was not considered at the initial level of phenotypic screening (Supplementary table 3). Further, from among the selected 119 lines, 23 lines were rejected due to extreme maturity duration while the selection was focused on selecting lines having maturity duration between 125 to 140 days. The yield performance and days to maturity of the initial phenotype-based core of 96 lines were compared with that of rejected 23 lines and found significant differences only for days to maturity (Supplementary table 3). The descriptive parameter estimated between the initial of 96 individuals and the original population of 314 individuals specifies the authenticity of initial core development (Supplementary table 4). Hence, the initial core with 96 individuals was carried forward for marker assay to estimate allele diversity.

Allelic diversity

A total of 38 reported molecular markers for yield from different chromosomes were used to evaluate the genetic

diversity of the population. All of the 38 markers were polymorphic across the 96 rice accessions and a total of 75 alleles were detected. Number of alleles ranged from 2 (in most of the markers) to 5 (GS7, Chr#7) with an average allele frequency of 1.42. Allele frequency ranged from 0.01 (DEP3_P22, Chr#-6) to 0.99 (SPIKE-01SNP, Chr#4), with a mean of 0.38. Gene diversity varied from 0.01 (DEP3_P22, Chr#6) to 0.67 (GS3_RGS2, Chr#12). The marker GS3_RGS2 (Chr#12) had the highest observed heterozygosity (0.48), while marker DEP3 (0.00) and Gn1a (0.00) had the lowest. Polymorphism Information Content (PIC) ranged from 0.10 to 0.51 with a mean of 0.35 and the marker GS3_RGS2 was found most polymorphic with a PIC value of 0.51 (Table 1). Seven markers (18.42%) were highly informative (PIC $\geq$ 0.5), 23 (60.49%) moderately informative (PIC $\geq$ 0.25 and $<$ 0.50), and 8 (31.04%) slightly informative (PIC $<$ 0.25). The presence of amplicons of DEP3-P22 was identified in most of the rice lines, while RM25895 was specifically present in a few elite lines. The amplicon of RM3689 (DSR QTL related) was unique to SR-4-2-1, N-301, N-333, M-1023, Jasmine-85, SR-28-1-1–1, S-594, and S-435 by its presence.

Population structure analysis

The model-based simulation of population structure using molecular markers demonstrated that as the model parameter K raises the log-likelihood, and thus the statistic K was used to calculate an appropriate value for K. The genetic relationship between 96 elite rice lines was determined using a Bayesian model-based programe called STRUCTURE v2.3.4. The estimated membership fractions (K) of 96 accessions varied between 2 and 10. The population structure analysis estimated that the optimum number of subpopulations K which best explained the structure of the accessions was 3 (K = 3, 64.67) using the Evanno method (Fig. 2; Supplementary Fig. 2). The majority of elite lines fell within 3 populations (K = 3) except three lines in population-1 (N-306, N-331 and C-774-7-1-1-1), ten lines in population-2 (N-305, N-333, SR-331-1, SR-27-8-1, M-476, SR-165-1, Sambha Mahasuri, Annada, GYD-279 and Moti) and five lines in population-3 (Padmini, SR-16-1-2-1, SR-57-1, SR-87-1 and SR-66-1) as admixtures at an association probability of 70%. The 21 admixtures consisted of 8 NGR, 5 NPT, 3 ABL, 1 NRL and 4 varieties. Population-1 consisted of 20 elite lines and was made up of NGR (17 lines), ABL (1 line) & NRL (2 lines). The population-2 had 51 elite lines which included NGR (30 lines), NPT (7 lines), ABL (6 lines), NRL (5 lines) & varieties (3 varieties). Population-3 had 25 elite lines which included NGR (13 lines), NRL (1lines), ABL (10 lines) &release varieties (1). The heterozygosity of 0.26 for the more heterogeneous was observed in populations-1 and 2, while it is 0.24 for population-3. The population-2 had a higher gene diversity (0.43) indicating a high genetic diversity compared to other the two populations (Table 2). The population-2 had the highest proportion of the genotypes (0.53), highest allele number per locus (1.84), gene diversity (0.43) polymorphism loci (89.61) and heterozygosity also (0.26).

Within and between populations diversity

The analysis of molecular variance (AMOVA) revealed that within-group variation accounted for 87% of the total variation, while variation between groups accounted for only 13% of total variation (Table 3). In comparison, population-2 had the highest proportion of polymorphic loci (89.61%) and was also found to be the most diverse, with the highest observed heterozygosity (0.26), allele number per locus (1.84), and gene diversity (0.43) (Table 2). Population-1 had 1.58 alleles per locus, population-2 had 1.84 alleles per locus, and population-3 had 1.67 alleles per locus. These findings on allelic variation amongst subpopulations demonstrated the allelic richness that exists to demonstrate genotype diversity. Similarly,

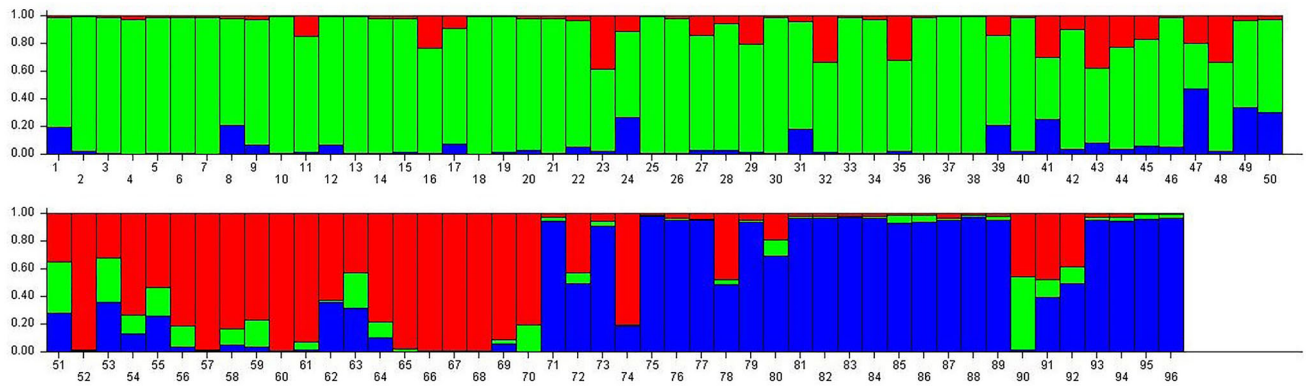

Fig. 2 Population structure (STRUCTURE) at K = 3 of 96 rice genotypes based on genotypic data using gene based and random markers related to grain yield of rice

Table 2 Analysis of molecular variance (AMOVA) between populations predicted by STRUCTURE analysis of 96 rice elite lines

Source of variation	DF	SS	MS	EV	% Variation	Stat	Value	Probability[a]
Among pops	2	113.427	56.714	1.596	13			
Within pops	93	953.021	10.248	10.248	87			
Total	95	1066.448		11.844	100	PhiPT	0.13	0.01

DF degrees of freedom, *SS* sum of squares, *MS* mean squares, *EV* estimated variance

*,***Significant at 5% and 1% probability respectively

[a]The probability is based on permutation across the full data set. PhiPT is a statistic measure for comparison between co-dominant data sets

Table 3 Genetic diversity parameters among three populations of 96 rice elite lines predicted by STRUCTURE analysis

Population	Sample size	Proportion of genotypes	Allele number per locus	Gene diversity	%Polymorphic loci	Expected heterozygosity	Mean grain yield (t/ha)
Population 1	20	0.21	1.58	0.36	76.62	0.24	5.65 ± 0.11
Population 2	51	0.53	1.84	0.43	89.61	0.26	7.35 ± 0.18
Population 3	25	0.26	1.67	0.41	83.12	0.26	6.59 ± 0.17

Ho observed heterozygosity, *PIC* polymorphism information content

population-1 gene diversity was 0.36, population-2 gene diversity was 0.43, and population-3 gene diversity was 0.41 (Table 2).

Clustering of the genotypes

The amplicons observed were used to create clusters across 96 elite lines using the neighbor-end joining (NJ) process. Three major clusters emerged from the un-rooted tree when 96 rice elite lines were classified (Fig. 3), cluster one with 37 lines, cluster two with 8 lines, and cluster three with 51 lines. The cluster-1 further classified to three sub-clustered viz., sub-cluster-a (17 lines), sub-cluster-b (13 lines) and sub-cluster-c (7 lines). Similarly, cluster-2 sub-divided to two sub-clusters viz; sub-clusters-d (5 lines) and sub-clusters-e (3 lines). The cluster three sub-divided to three sub-clusters viz; sub-clusters-f (26 lines), sub-clusters-g (13 lines) and sub-clusters-h (12 lines). Cluster one is mainly dominated by NGR lines followed by NPT lines. Cluster two included only eight genotypes with a small proportion of different categories of lines. Cluster three included mainly NGR lines followed by ABLs. The genotypes were categorized on the basis of groups and presented in Table 4.

Principal component analysis and graphical visualization

PCA was used to confirm the genetic variability of 96 elite rice lines using a variant-covariant matrix. The first and second components accounted for 8.80 and 7.15% of the variance, respectively, with eigenvalues of 0.91 and 0.71. Although the PCA plot revealed three major clusters, lines were distributed across all four coordinates. The NRL genotypes are clustered together at the upper left coordinate, forming a distinct category and distinguishable from cultivars in their two-dimensional dispersion. The lines in the PCA figure were labeled with five different colors and five different types of symbols to represent the various types of lines (Fig. 4). Two-dimensional scaling determined by PCA analysis revealed the same pattern of grouping as the population structure and classified the majority of cultivars into three major clusters.

Comparison between grouping approaches

The results obtained from different approaches used for grouping genotypes to assess their variability were found to follow a similar trend. When PCA and dendrogram groupings were compared, genotypes were grouped into three major clusters both in PCA and dendrogram, but the composition of clusters varied. However, a maximum variation that exists among genotypes was successfully captured with two components in PCA and the nearest variation was captured in a dendrogram to classify genotypes into three major groups. The first two principal components were found to be most indicative of variation patterns between accessions. A similar trend was also found with a bar graph obtained from STRUCTURE analysis, where 96 genotypes were subdivided into three

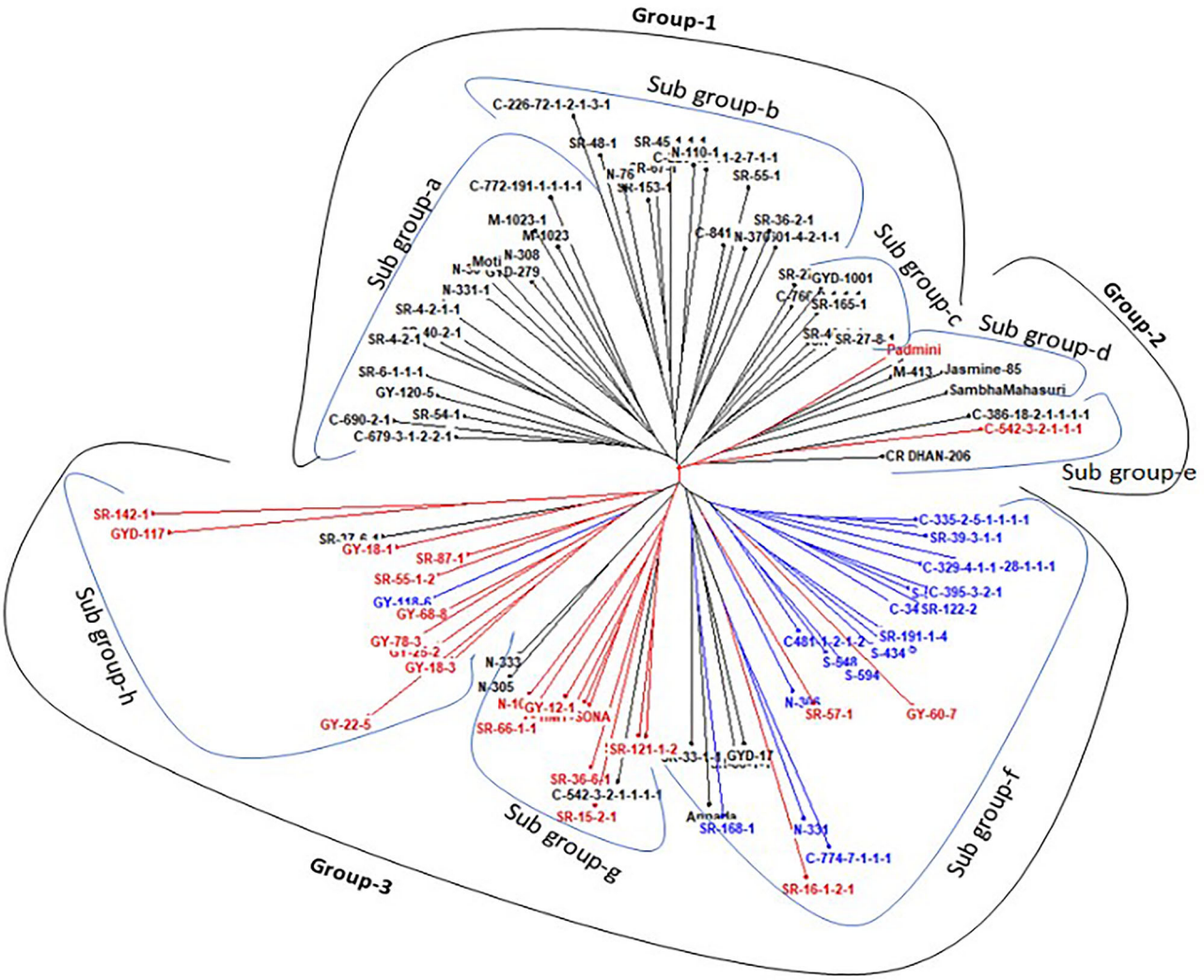

Fig. 3 Neighbor joining tree illustrating the genetic relationships of the 96 genotypes

subpopulations based on allelic variations that exist among genotypes. This confirms the pattern variation that exists in the population was consistent enough to carry out further condensation in order to develop a fine core set of genotypes. Thus, suggesting markers used to classify genotypes may be recommended for the purpose of differentiating population subgroups and could be used as decision support tools in developing a fine core set for the breeding application. Further, graphical genotyping (GGT) was used to re-emphasize the results of grouping patterns, where GGT helps to visualize the distinctness of genotypes-based marker alleles on each chromosome. From the results, it could be possible to find that no two genotypes completely shared similar genetic profiles (Fig. 5). However, patterns of colors indicate partial sharing of alleles between different genotypes in the population indicating the relatedness among each other. These results demonstrate the ability of gene-based markers to discriminate genotypes and also serve as a unique identity of elite rice lines in the core set.

Core set

From an elite population size of 96, a fine and potential core with only 11 (11.45% of 96 elite lines) genotypes maintaining the same genetic diversity and allelic richness as that of the original population, which can serve as a potential pool of parental lines to consider for rice improvement breeding programs were developed. This core collection was composed of 7.29% (7 lines) of NGR, 3.12% (3 lines) of NPT, and 1.04% (1 line) of NRL of the entire 96 elite lines (Table 4). The fine core set was extracted considering the maximum allelic retention explained by Shannon's allele diversity coefficient and yield potential. The selected 11 genotypes of the fine core set represent different populations depicted by

Table 4 Cluster analysis of 96 elite rice lines based on NJ approach and 11 identified core genotypes

SL no.	Cluster	Sub cluster	Lines	Core set genotypes	Mean grain yield for clusters (t/ha)
1	Cluster-1	Subcluster-1	C-679-3-1-22-1, C-690-2-1, SR-54-1, GY-120-5, SR-6-1-1, SR-4-2-1, SR-54-1, SR-40-2-1, SR-4-2-1, N-33-1, N-301, Moti, GYD-379, N-308, M-1023-1, M-10230, C-772-191-1-1-1-1	N-301, SR-48-1, N-370, SR-36-2-1	6.79 ± 0.21
		Subcluster-2	C-226-72-1-2-1-3-1, SR-48-1, N-76, SR-153-1, SR-67-1, SR-45-1-1-1, C-226-45-11-27-1-1, SR-36-2-1, C-601-4-2-1-1, N-110-1, SR-55-1, N-370, C-841-1-1-2-1		
		Subcluster-3	SR-27-4-1-2, C-766-9-1-1-1, SR-165-1, SR-40-1, SR-14-5, SR-27-8-1, GYD-1001		
2	Cluster-2	Subcluster-1	Padmini, M-476, M-413, Jasmini-85, Sambha Mahsuri	M-476, C-386-18-2-1-1-1-1	5.89 ± 0.13
		Subcluster-2	C-386-18-2-1-1-1-1, C-542-3-2-1-1, CRDHAN-206		
3	Cluster-3	Subcluster-1	C-325-2-5-1-1-1-1, SR-45-3-1, SR-39-3-1, SR-28-1-1-1, C-329-4-1-1, C-395-3-2-1, S-5-7-1, SR-122-2, C-345-1, SR191-1-4, S-345, S-434, GY-60-7, S-594, C-481-1-2-1-2, SR-57-1, N-306, C-774-7-1-1-1, N-331, SR-16-1-2-1, GYD-17, SR-80-1-1, SR-168-1, Annada-1, SR-33-1-1	S-5-7-1, C-774-7-1-1-1, N-305, SR-36-6-1	7.10 ± 0.18
		Subcluster-2	SR-121-2, SR-97-1, C-542-3-2-1-1-1-1, SR-15-2-1, SR-36-6-1, HMT-Sona, SR-14-3-1-1, SR-66-1-2, GY-2-1, SR-66-1-1, N-102, N-305, N-333		
		Subcluster-3	GY-225, GY-183, GY-25-2, GY-68-8, GY-118-6, SR-55-1-2, SR-87-1, GY-18-1, SR-37-6-1, GYD-117, SR-142-1, GY-78-3	SR-142-1	

N-New plant type, SR and C-New generation rice, M-Nitrogen responsive line, GY and GYD-Advance breeding material, other name is release variety

STRUCTURE and PCA analysis, thus maintaining similar genetic diversity as that of the original population. The composition of the fine core set was similar to the original core set of 96 genotypes in terms of allelic diversity and could represent the original population of 314 genotypes with other plant characters.

Discussion

A panel of 96 elite rice lines was selected for the present study, which was bred for high grain yield for different ecologies. Genetic diversity in such lines is expected to be unique for the development of high yield segregants by crossing elite × elite lines. This could be possible if genetic relatedness is less and the presence of different alleles for high grain yield in the population. Thus, understanding population structure and genomic variations are crucial for the selection of lines from a population. Breeding through the selection of superior progenies for different ecological situations results in yield enhancement due to the accumulation of major and minor alleles in the breeding lines (Collard et al. 2019). In this regard, our study provides useful information on elite lines distinct at the molecular level also. This enabled us to reconstruct the genetic relationships and genetic diversity within several elite lines that exhibit enormous genetic variability.

The extent of available variation dictates the use of elite lines. The elite lines used in this study exhibited a high level of genetic diversity both in terms of yield and at the molecular level. While morphological trait analysis is a valuable tool for examining genetic differences reflected in phenotypic expression, its results may not always accurately reflect true genetic variation due to genotype-environment interaction and unknown genetic control of polygenic traits (Sanghamitra et al. 2018). Using molecular markers to characterize genotypes based on polymorphisms at the DNA level provides a powerful tool for estimating genetic divergence (Hashimoto et al, 2004). The number of alleles per locus ranged between 2 and 5, which is considered adequate for gene-specific and QTL-based markers, as demonstrated in this study. Allele number of 1–6 alleles per locus with an average of 3.24 was reported in colored upland rice germplasm (Ahmad et al. 2015). The number of alleles in a population indicates the genetic richness of the population.

The PIC values of markers across genotypes are yet other criteria to assess genetic diversity at the molecular level by using these markers. Even though all the lines analyzed in the present study were high-yielding, elite, and also diverse, but still, it is important to choose appropriate

parents for hybridization very judiciously. The PIC value ranged from 0.00 to 0.51 with an average of 0.35. Such a medium-level PIC value of 0.35 is expected for gene specific or QTL based markers because of the lower level of polymorphism. A medium level of PIC of 0.38 was also reported from 40 Pakistan rice accessions. However, the use of grain yield linked and trait-based marker shows a lower rate of polymorphism (Saha et al. 2004).

Among the 38 polymorphic markers, 7 markers (18.42%) exhibited PIC values ≥ 0.50 thereby suggesting their suitability for genetic relationship estimation and diversity study in relation to grain weight, grain length (GS3_RGS1 and GS3_RGS2), grain size (GS5_Indel1), grain number, plant height, heading date (Ghd 7_05SNP) and yield under osmotic stress (RM 12289, RM 23065 and RM 25457). The PIC and polymorphism percentage revealed that markers used in this study were polymorphic to explore the genetic relationship among elite rice lines from the different breeding programes. Thus, the grouping of rice genotypes based on polymorphic markers indicated high genetic diversity among the elite lines in relation to these traits. The maximum diversity among the elite lines indicated distinct allelic differences in their genetic constitution (Surapaneni et al. 2016).

The heterozygosity for populations-2 (0.26) and 3 (0.26) shows heterogeneous than the population-1 (0.24). The population-2 had the highest proportion of genotypes (0.53), highest allele number per locus (1.84), gene diversity (0.43), polymorphism loci (89.61), and heterozygosity of < 0.50 (Table 2). Thus, population-2 would serve as the best group for utilization in breeding with regard to the use of parents for crossing. Inter crossing of individuals from the populations-2 with other populations also would be beneficial to accumulate the maximum yield genes/QTLs in the segregating populations and later bring into single background (Table 2). Similar results for heterozygosity for populations were also observed by Aljumaili et al. (2018). The proportion of genotypes observed in each subpopulation reflected the 96 genotypes' diversity (Fig. 3). In light of these findings, it can be concluded that

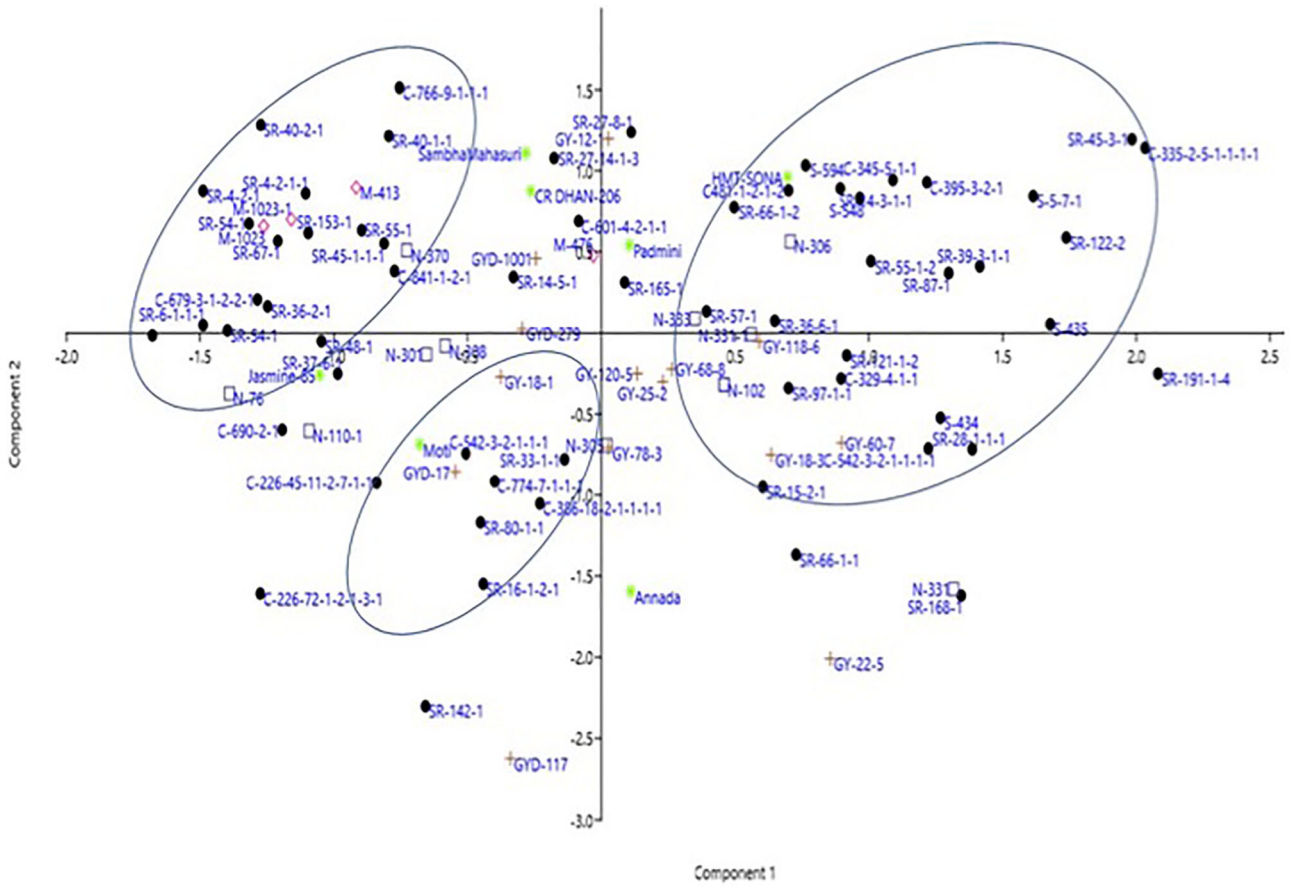

Fig. 4 PCA scattered polar diagram of 96 genotypes and its distribution in quadrangles (genotypes are labeled as Black color dot symbol denoted NGR genotype, blue color square symbol denoted NPT genotype, Chartreus green color star symbol denoted released variety genotype, deep pink diamond color symbol denoted NRL genotype, Grey color plus symbol denoted ABL genotype (Color figure online)

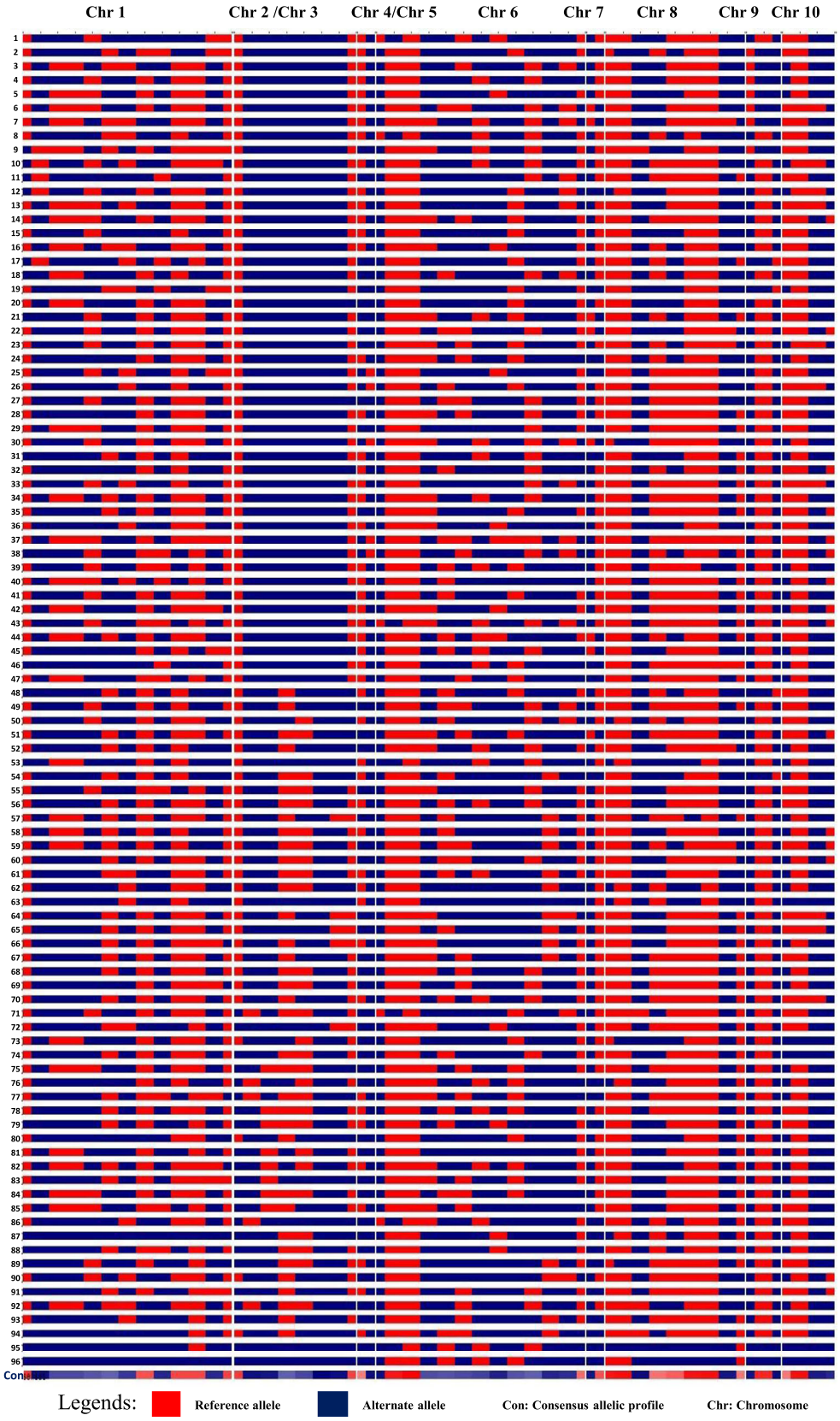

Fig. 5 GGT image for 96 elite rice genotypes using 38 markers

the elite genotypes under study possessed sufficient potential for successful application in rice improvement.

Grouping of these elite lines based on marker information showcased the presence of genetic diversity within the elite pool. This classification made the prevision to select elite lines as parents for hybridization from contrasting clusters. PCA is a dimension reduction technique that can be used to condense a large set of variables into a smaller set that retains the majority of the information contained in the larger set (Price et al. 2006). For convenience and better discussion, only two PCs that possessed the eigen values more than one were considered for analysis. However, the first two PCs considered were explaining very less of total variation suggesting elite lines considered and molecular markers employed exhibiting multidimensional genetic diversity. Further, the NRL lines considered in the study were concentrated in a single quadrant indicating their uniqueness from other genotypes. Although PCA considers Eigen values for plotting genotypes, it does not consider multi-dimension scaling and similarity matrix of genotypes while reducing the dimensionality of the marker information. Another multivariate tool principal coordinate analysis (PCoA) used to extract the variation in marker information using the Euclidian similarity index. The variation extracted by the first two PCoA was also negligible confirming the results of PCA. PCA described the appropriate amount of genetic diversity required to derive segregates with different gene combinations upon crossing these elite lines.

Population structure is a well-studied but little understood feature of population genetics. Structure applies to any deviation from random mating and involves inbreeding, associative mating (in which reproduction is stratified by genotype), and geographical sub-division. Unequal distribution of alleles among sub-populations from different ancestries defines the population structure of the pool. When these sub-populations are considered for parental selection advantage of allelic combinations can be rewarded in terms of promising progenies. Before understanding the population structure of the pool, these elite lines were grouped into population 1 (NGR), population 2 (ABL), and population 3 (NRL). These populations were tested for their significant differences by performing an analysis of molecular variation (AMOVA). The AMOVA results suggested higher variation between populations and lower within population variation suggesting these groups significantly different. The population structure of 96 elite lines with marker information delivered the existence of three sub-populations as defined by *Ad-hoc* statistic ΔK with MCMC model. Graphical representation of population structure clearly indicated the existence of a relatively good population structure. This implies that allelic information captured from the population is sufficiently explaining the genetic diversity of elite lines considered in this study. These results are in confirmation with results obtained from AMOVA, to promise the existence of preferably high diversity among sub-populations. The proportions of genotypes shared by each sub-population were also relatively comparable, suggesting each sub-population is equally potent to carry diverse and elite parental lines for choosing parental lines. In support of the results above Graphical GenoTyping (GGT) also displayed the discriminating ability of markers thereby assessing the genetic diversity existing among elite lines considered for this study. This also implies providing a unique identity to each line by displaying chromosome-wise distribution of DNA markers used to genotype these elite lines.

The two-dimensional data generated by PCA analysis resembled the results of the radial dendrogram (Fig. 4) and population structure (Fig. 2) analyses, in which the majority of lines were distributed and grouped in similar ways. However, there is some specific distribution of genotypes in the PCA, where NRL lines were fell into a single upper right quadrant (Fig. 4), in PCoA it was in a lower right quarter (figure not presented) and in group 1 and group 2 of the dendrogram drawn using DARwin software. When the radial dendrogram was created, the genotypes belonging to group-1 have uniformly distributed in the PCA right upper and lower quadrants (subgroup a, b, c, d) and left upper and lower quadrants of PCoA diagram (subgroup a, b, c, d). The elite genotypes in subgroup-e (group-2) were distributed in the right lower quadrant of the PCA and left upper quadrant of the PCoA. The elite genotypes of subgroup-f and g (group-3) of radial dendrogram were distributed in the left lower quadrant of PCA and right upper quadrant of PCoA. Similarly, the elite genotypes of subgroup-h (group-3) were distributed in the left upper quadrant of PCA and the right lower quadrant of PCoA. However, the PCA identifies the best representatives of the population's differences in elite genotypes; it does not provide information about population's size or subgroups. Additionally, it can be advantageous to distinguish between subgroups of populations involved in classifying genotypes. Besides that, when the PCA and dendrogram were compared, they revealed similar tendencies and similar group compositions. Hence, the information generated from PCA, PCoA and dendrogram can be effectively used for genetically differentiating lines in the pool during the process of selection of parental lines while generating elite $\times$ elite cross combinations.

A set of core formation from 96 elite lines resulted in a shrinking the total number of genotypes to 11 with the same level of genetic variation and diversity. The diversity value of the new core set was found on par with the original set of 96 lines as a principle of the core collection.

The fine core developed was constituted with very a smaller number of genotypes represented total genetic variation explained by candidate gene-based markers along with similar allelic richness. Hence, it assures the diversity of the original population in the new core collection of 11 lines. The 11 genotypes and the genotypes that fall in population-2 (as per structure analysis) could be used for crossing programs for developing high-yielding rice cultivars. This has implication in strengthening the breeding material to be chosen as parents from a small set of genotypes.

Conclusion

Since the gene sequences are conserved in many regions, finding variation in all the genes and genomic regions is not possible. However, a small variation in genetic material at genic regions may be captured using gene-specific markers which is very useful to identify allelic variation in parental genotypes selected for hybridization. A breeding population with 314 individuals was downsized by considering yield as a major criterion through critical evaluation of a core set of 96 genotypes with similar genetic diversity. The size of the core set was further reduced to a fine core set with only 11 genotypes by considering maximum allele frequency retention considering genetic differences at the genic level and yield expression. The fine development represents the original core set of 96 genotypes for allelic diversity by representing the population and sub-populations of the original core set. Since these 11 genotypes of a fine core set are from a distant clusters and elite in terms of yielding potential, these genotypes may be utilized for developing elite × elite crosses with distant relationship with the expectation of obtaining more promising progenies in segregating generations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12298-022-01190-8.

Acknowledgements This research was carried out as part of ICAR Plan Project at ICAR-National Rice Research Institute, Cuttack, India. The authors are highly grateful to the Director, ICAR–NRRI for providing all the necessary facilities.

Authors' contributions RPS and SB performed the experiments, prepared the manuscript and contributed equally to this work; SB and AKC analyzed the data; SKD, JM, MK and BCM shared the elite lines for experiments; AZTP collected the marker information used in the experiment and interpretation of data; AK and AZTP coordinated the experiments and improved the manuscript.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Ahmad F, Hanafi MM, Hakim MA, Rafii MY, Arolu IW, Abdullah SNA (2015) Genetic divergence and heritability of 42 coloured upland rice genotypes (*Oryzasativa*) as revealed by microsatellites marker and agro-morphological traits. PLoS ONE 10(9):e0138246. https://doi.org/10.1371/journal.pone.0138246

Akhtar S, Bhat MA, Wani SA, Bhat KA, Chalkoo S, Mir MR, Wani SA (2010) Marker assisted selection in rice. J Phytol 2(10):66–81

Aljumaili AJ, Rafii MY, Latif MA, Sakimin SZ, Arolu IW, Miah G (2018) Genetic diversity of aromatic rice germplasm revealed by SSR markers. BioMed Res Int 7658032:1–11. https://doi.org/10.1155/2018/7658032

Atlin GN, Cairns JE, Das B (2017) Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. Glob Food Secur 12:31–37. https://doi.org/10.1016/j.gfs.2017.01.008

Bandumula N (2018) Rice production in Asia: key to global food security. Proc Natl Acad Sci India Sect B Biol Sci 88(4):1323–1328

Berloo RV (2008) GGT 2.0: versatile software for visualization and analysis of genetic data. J Hered 99(2):232–236. https://doi.org/10.1093/jhered/esm109

Choudhary G, Ranjitkumar N, Surapaneni M, Deborah DA, Vipparla A, Anuradha G, Siddiq EA, Vemireddy LR (2013) Molecular genetic diversity of major Indian rice cultivars over decadal periods. PLoS ONE 8(6):e66197. https://doi.org/10.1371/journal.pone.0066197

Chung HK, Kim KW, Chung JW, Lee JR, Lee SY, Dixit A, Kang HK, Zhao W, McNally KL, Hamilton RS, Gwag JG (2009) Development of a core set from a large rice collection using a modified heuristic algorithm to retain maximum diversity. J Integr Plant Biol 51(12):1116–1125

Collard BC, Gregorio GB, Thomson MJ, Islam MR, Vergara GV, Laborte AG, Nissila E, Kretzschmar T, Cobb JN (2019) Transforming rice breeding: re-designing the irrigated breeding pipeline at the International Rice Research Institute (IRRI). Crop Breed Genet Genom 1:8. https://doi.org/10.20900/cbgg20190008

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Gobu R, Shiv A, Anilkumar C, Basavaraj PS, Harish D, Adhikari S, Vinita R, Umesh H, Sujatha M (2020) Accelerated crop breeding

towards development of climate resilient varieties. Climate change and Indian agriculture: challenges and adaptation strategies

Hashimoto Z, Mori N, Kawamura M, Ishii T, Yoshida S, Ikegami M, Takumi S, Nakamura C (2004) Genetic diversity and phylogeny of Japanese sake-brewing rice as revealed by AFLP and nuclear and chloroplast SSR markers. Theor Appl Genet 109(8):1586–1596

Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009) Natural variation at the DEP1 locus enhances grain yield in rice. Nat Genet 41(4):494–497

Katara JL, Parameswaran C, Devanna BN, Verma RL, Anilkumar C, Patra BC, Samantaray S (2021) Genomics assisted breeding: The need and current perspective for rice improvement in India. ORYZA Int J Rice 58(1 Spl):61–68

Kim SR, Ramos J, Ashikari M, Virk PS, Torres EA, Nissila E, Hechanova SL, Mauleon, R, Jena KK (2016) Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-genome sequencing strategy to increase yield potential of rice, Oryza sativa L. Rice 9(1):1−17

Molla KA, Azharudheen TPM, Ray S, Sarkar S, Swain A, Chakraborti M, Vijayan J, Singh ON, Baig MJ, Mukherjee AK (2019) Novel biotic stress responsive candidate gene based SSR (cgSSR) markers from rice. Euphytica 215(2):17

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Perrier X, Jacquemoud-Collet JP (2006) DARwin software: dissimilarity analysis and representation for windows. http://darwin.cirad.fr/darwin

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

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

Qiao Y, Piao R, Shi J, Lee SI, Jiang W, Kim BK, Lee J, Han L, Ma W, Koh HJ (2011) Fine mapping and candidate gene analysis of dense and erect panicle 3, DEP3, which confers high grain yield in rice (Oryza sativa L.). Theor Appl Genet 122(7):1439–1449

Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang LJ, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

Sanghamitra P, Sah RP, Bagchi TB, Sharma SG, Kumar A, Munda S, Sahu RK (2018) Evaluation of variability and environmental stability of grain quality and agronomic parameters of pigmented rice (*O. sativa* L.). J Food Sci Technol 55(3):879–890. https://doi.org/10.1007/s13197-017-2978-9

Sandhu N, Torres RO, Sta Cruz MT, Maturan PC, Jain R, Kumar A, Henry A (2015) Traits and QTLs for development of dry direct-seeded rainfed rice varieties. J Exp Bot 66(1):225–244

Shao G, Wei X, Chen M, Tang S, Luo J, Jiao G, Xie L, Hu P (2012) Allelic variation for a candidate gene for GS7, responsible for grain shape in rice. Theor Appl Genet 125(6):1303–1312

Shivakumar M, Vangala R, Kumawat G (2018) Speed breeding for Indian agriculture: a rapid method for development of new crop varieties. Curr Sci 115:1241

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39(5):623–630

Subudhi HN, Prasad KVSV, Ramakrishna Ch, Sah RP, Pathak H, Ravi D, Khan AA, Padmakumar V, Blümmel M (2020) Genetic variation for grain yield, straw yield and straw quality traits in 132 diverse rice varieties released for different ecologies such as upland, lowland, irrigated and salinity prone areas in India. Field Crops Res 245:107626. https://doi.org/10.1016/j.fcr.2019.107626

Surapaneni M, Balakrishnan D, Mesapogu S, Raju AK, Rao YV, Neelamraju S (2016) Genetic characterization and population structure of Indian rice cultivars and wild genotypes using core set markers. 3 Biotech 6:95. https://doi.org/10.1007/s13205-016-0409-7

Vaughan DA (1991) Choosing rice germplasm for evaluation. Euphytica 54:147–154

Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, Zhang L, He W, Lu B, Lin H, Ma H (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat Genet 40(11):1370–1374

Wang C, Sheng C, Sibin Y (2011) Functional markers developed from multiple loci in GS3 for fine marker-assisted selection of grain length in rice. Theor Appl Genet 122(5):905–913

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.