

RESEARCH ARTICLE

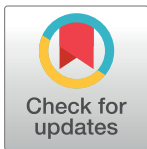
Identification of QTLs for high grain yield and component traits in new plant types of rice

Ravindra Donde¹, Shibani Mohapatra¹, S. K. Yasin Baksh¹, Barada Padhy¹, Mitadru Mukherjee¹, Somnath Roy², Krishnendu Chattopadhyay¹, A. Anandan¹, Padmini Swain¹, Khirod Kumar Sahoo³, Onkar Nath Singh¹, Lambodar Behera¹✉, Sushanta Kumar Dash¹✉*

1 ICAR-National Rice Research Institute (NRRI), Cuttack, Odisha, India, **2** ICAR-NRRI, Regional Research Station (CRURRS), Hazaribagh, Jharkhand, **3** Department of Botany, Ravenshaw University, Cuttack, Odisha

✉ These authors contributed equally to this work.

* skdash139@gmail.com



OPEN ACCESS

Citation: Donde R, Mohapatra S, Baksh SKY, Padhy B, Mukherjee M, Roy S, et al. (2020) Identification of QTLs for high grain yield and component traits in new plant types of rice. PLoS ONE 15(7): e0227785. <https://doi.org/10.1371/journal.pone.0227785>

Editor: Tapan Kumar Mondal, ICAR - National Research Center on Plant Biotechnology, INDIA

Received: December 25, 2019

Accepted: June 11, 2020

Published: July 16, 2020

Copyright: © 2020 Donde et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The author(s) received no specific funding for this work. This is a student work and contributions.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: NPT, New plant type; GLM, General Linear Model; MLM, Mixed Linear Model; MS,

Abstract

A panel of 60 genotypes comprising New Plant Types (NPTs) along with *indica*, *tropical* and *temperate japonica* genotypes was phenotypically evaluated for four seasons in irrigated situation for grain yield *per se* and component traits. Twenty NPT genotypes were found promising with an average grain yield varying from 5.45 to 8.8 t/ha. A total of 85 SSR markers were used in the study to identify QTLs associated with grain yield *per se* and related traits. Sixty-six (77.65%) markers were found to be polymorphic. The PIC values varied from 0.516 to 0.92 with an average of 0.704. A moderate level of genetic diversity (0.39) was detected among genotypes. Variation to the tune of 8% within genotypes, 68% among the genotypes within the population and 24% among the populations were observed (AMOVA). This information may help in identification of potential parents for development of transgressive segregants with very high yield. The association analysis using GLM and MLM models led to the identification of 30 and 10 SSR markers associated with 70 and 16 QTLs, respectively. Thirty novel QTLs linked with 16 SSRs were identified to be associated with eleven traits, namely tiller number (*qTL-6.1*, *qTL-11.1*, *qTL-4.1*), panicle length (*qPL-1.1*, *qPL-5.1*, *qPL-7.1*, *qPL-8.1*), flag leaf length (*qFLL-8.1*, *qFLL-9.1*), flag leaf width (*qFLW-6.2*, *qFLW-5.1*, *qFLW-8.1*, *qFLW-7.1*), total no. of grains (*qTG-2.2*, *qTG-a7.1*), thousand-grain weight (*qTGW-a1.1*, *qTGW-a9.2*, *qTGW-5.1*, *qTGW-8.1*), fertile grains (*qFG-7.1*), seed length-breadth ratio (*qSib-3.1*), plant height (*qPHT-6.1*, *qPHT-9.1*), days to 50% flowering (*qFD-1.1*) and grain yield *per se* (*qYLD-5.1*, *qYLD-6.1a*, *qYLD-11.1*). Some of the SSRs were co-localized with more than two traits. The highest co-localization was identified with RM5709 linked to nine traits, followed by RM297 with five traits. Similarly, RM5575, RM204, RM168, RM112, RM26499 and RM22899 were also recorded to be co-localized with more than one trait and could be rated as important for marker-assisted backcross breeding programs, for pyramiding of these QTLs for important yield traits, to produce new-generation rice for prospective increment in yield potentiality and breaking yield ceiling.

Mean square; Fst, F-statistics, subpopulations within the total population; Fis, F-statistics individuals within subpopulations; DFF, Days to fifty percent flowering; PH, Plant height; TL, Tiller number; PL, Panicle length; FLL, Flag leaf length; FLW, Flag leaf width; FG, Fertile grains; TG, Total number of grains; TGW, 1000 grains weight; YLD, grain yield t/ha; SLBR, Seed length-breadth ratio; FDR, False Discovery Rate.

Introduction

Rice (*Oryza sativa* L.) is a staple food crop sustaining more than 3.5 billion people in the globe. In current scenario, rice productivity is increasing at a rate of 1% per year which is less than the required rate of 2.4% per year to double the global production by 2050 [1]. Considering a glimpse of the history, a quantum jump in productivity was achieved due to the green revolution in mid-sixties, which drastically enhanced the rice production of the world. However, a ceiling of productivity potential is reported by and large in semi-dwarf inbred *indica* genotypes since release of IR-8 [2], in spite of substantial improvement in yield stability, per day productivity and grain quality [3]. A breakthrough in productivity barrier is necessitated because of increasing competition for natural resources such as, land, water and others given population explosion coupled with expanding industrialization, urbanization and diversion of agricultural land [1,4,5]. This is further aggravated with the abnormal change in weather and climate with significant influence on crop productivity and quality [6,7].

Rice scientists are facing many challenges for doubling rice production by 2050. Irrigated rice has a share of 75% of total rice production in the world, although it has a share of about 55% of the total rice area [2]. Therefore, improvements and modification of rice genotypes for this ecology are supposed to have significant impact on rice productivity in future. During the past decade, there has been a significant slowdown in the production potential of modern rice cultivars. In this context, physiologists and breeders hypothesised that this stagnation could be overcome by improving the plant type. The existing plant type bears several unproductive tillers in high tillering type and limited sink size i.e., small panicles. The excessive leaf area causes mutual shading, low light and a reduction in canopy photosynthesis [8,9]. Apart from that, there are several bottlenecks viz., spikelet sterility, short panicle length, limited grain numbers, lodging susceptibility, etc. Moreover, there is also the loss of genetic diversity in improved varieties for which breeders are facing difficulties in finding divergent gene pools. Modern high yielding rice varieties have been associated with some unfavourable traits/alleles, which may be sensitive to biotic and abiotic stresses and may be responsible for lowering grain yield [9].

In this context, IRRI scientists developed “New Plant Type” (NPT-2nd generation) by recombining some suitable features of *tropical japonicas* with *indica* [9]. The main idea behind the 2nd generation NPT was development of high yielding super rice varieties, which could be able to produce high yield lines endowed with stability. Some of the NPTs performed exceedingly well and produced even more than 10t ha⁻¹ in the Philippines [2]. During the process of development, some of advanced generation segregating materials were shared with NRRI, India. The materials were subjected to further selection at NRRI under irrigated ecology, as appreciable variability was still available, with an objective of developing promising NPTs suitable for the climate specific to eastern region in particular and country in general. Trait specific selections were done for few generations to establish fixed lines, i.e., NPT selections (NPTs). In this context, NPTs were evaluated systematically under observational yield trial (OYT) for one season and the number was narrowed down subsequently. Advanced Yield Trial (AYT) followed it for four wet seasons at NRRI. Some of the highly promising NPTs were identified with high-quality agronomic traits like higher grain number per panicle, panicle length, panicle weight, grain size, ear bearing tiller number along with ideal plant height. Some of NPTs performed exceptionally well and showed the productivity of more than 10.0 t ha⁻¹ during dry season 2011 [10].

With this backdrop, we wish to proceed for development of still higher yielding genotypes or super rice kind of crop ideotype utilizing the existing set of highly promising NPTs, which should have the productivity potential, at least 20% higher than the popular rice and check varieties. The target was utilization of one of the most promising gene pools through

conventional as well as molecular approach. The focus is to accumulate the thousands of minor QTLs with additive genetic variance along with major ones. Here, the extent of genetic variation and relationships between genotypes are more important for designing effective breeding strategy [11].

The association mapping (AM) is an useful tools in identifying QTLs/genes associated with different traits in plant species. It utilizes natural variation [12], and is supposed to have great potential to evaluate and characterize a wide range of alleles. Several researchers have reported the utility of association analysis in the identification of QTLs for different traits in rice, viz., grain yield [13], grain yield under water deficit [14], grain quality traits [15], agronomic traits [16,17], grain yield under reproductive phase drought stress [18], panicle architecture and spikelet's/ panicle [19], plant height and grain yield [20].

However, there are meagre reports on QTLs association for grain yield and yield-related complex traits particularly on NPT. The genes/ QTLs related to high grain yield would be of great help in breaking yield ceiling. Moreover, it would be beneficial in identifying traits specific donors for designing effective breeding strategy for the development of super rice. The present study was undertaken to identify QTLs associated with grain yield and yield-related agronomic traits using diverse genotypes.

Materials and methods

Plant materials

A panel comprising sixty rice genotypes, including 48 NPTs, six highly popular released *indica* varieties, three temperate *japonica* and three tropical *japonica*, most of them significantly diverse and distinct, were used for identification of QTLs for 11 yield-related traits through association studies (S1 Table). The NPTs were from 41 NPT populations collected from IRRI at the advance segregating stage. From those populations, conscious trait specific single plant selections (SPS) were made basically for yield-related traits for few generations and finally ~500 promising fixed SPS were identified and evaluated in OYT at ICAR-National Rice Research Institute (NRRI) (coordinates 20.4539° N, 85.9349° E). Subsequently, the number was drastically narrowed down to 48 strictly based on yield and important agronomic traits. Forty-eight best performing NPTs were evaluated under four environments along with 12 checks (6 *indica*, 3 *tropical japonica* and 3 *temperate japonica*) and these were further studied for molecular diversity and QTL association.

Phenotyping

All the 60 genotypes were grown in two replications, each genotype covering 5.04 m² area (800 m² total plot size), following Randomized Complete Block Design (RCBD) during wet seasons of 2011, 2012, 2013 and 2014 (S2 Table). The phenotypic data of yield *per se* and yield-related traits were recorded at different phenological stages. Normal management practices and plant protection measures were taken during crop growth. The genotypes were harvested at maturity, i.e., after 30 to 35 days of flowering. The post-harvest data were recorded after the crops were harvested, threshed and dried. This study primarily focused on 11 yield and yield-related traits, namely, days to 50% flowering (DFF), plant height (PH), tiller number (TL), panicle length (PL), flag leaf length (FLL), flag leaf width (FLW), no of fertile grains (FG), total no of spikelets per panicle (TG), 1000-grain weight (TGW), seed length-breadth ratio (SLBR) and grain yield t/ha (YLD). The yield *per se* was measured by weighing the plot yield (4m² each) at 13% moisture level and converted it to tons/ha. Other yield contributing traits were measured using standard procedure. The seed length-breadth ratio was measured using Anndarpan machine and software developed by CDAC, Govt. of India [21]. The phenotypic

data were used for statistical analyses viz. SD, CV, ANOVA, correlation, regression, and principal component analysis (PCA), Bi-plots using XLSTAT software version 2019.1 (Addinsoft, Paris, France). The ClustVis, an online web tool (<http://biit.cs.ut.ee/clustvis/>) was used for analysis of phenotypic traits. The visualizing clustering of multivariate data of yield *per se* and yield-related traits were analyzed by Heat-map and PCA [22]. The ClustVis has been written in the Shiny web application framework by using R package version 0.10.2.1 for R statistics software [15,23].

Genotyping

The genomic DNA was isolated from 3–4 g of fresh leaf tissues of each rice genotype following Cetyl Trimethyl Ammonium Bromide (CTAB) method (Murray and Thompson, 1980). The extracted genomic DNA samples were dissolved in TE buffer (10 mM Tris-base, 1 mM EDTA, pH-8.0). The quality and quantity of DNA of each sample were measured by agarose gel electrophoresis and spectrophotometer. The SSR markers were selected on the basis of the previous report associated with different yield QTLs [24–30] and polymorphic contents (<http://www.gramene.org>). The polymerase chain reaction (PCR) was performed in a 20 μ l reaction mixture containing 5 pM (pico-mole) of forward and reverse primers of each SSR locus, 200 mM of each dNTP, 0.5 U of *Taq* DNA polymerase, 10 mM Tris-HCl (pH = 8.3), 50 mM KCl and 1.5 mM MgCl₂. The PCR amplification was carried out in a thermal cycler (Veriti 96, Applied Biosystems, USA) as per the following cycling parameters: initial denaturation at 94^o C for 3 min, followed by 35 cycles of denaturation at 94^o C for 1 min, annealing at 55–67^o C (depending upon primer) for 1 min and extension at 72^o C for 1.5 min and final extension at 72^o C for 5 min. The amplified products were separated on 2.5% - 3% agarose gels using 1X TBE buffer and stained with ethidium bromide (0.5 μ g/ μ l). The gels were observed under UV radiation and were photographed using a gel documentation system (G-Box, Syngene, USA) to detect amplified fragments. The size of amplified bands was determined based on the migration relative to molecular weight size markers (50 bp DNA ladder, MBI Fermentas, Lithuania).

Genetic diversity

The amplified bands were scored as present (1) or absent (0) for each genotype and micro-satellite marker combination. Each band was considered as an allele. The data were entered into a binary matrix as discrete variables and subsequently used for assessing allelic and molecular diversity such as number of total alleles (TA), unique alleles (UA), rare alleles (RA), expected alleles (Ne), polymorphism information content (PIC), gene diversity, homozygosity (Ho) and heterozygosity (He) by using Power-Marker Ver 3.25 [31]. The polymorphism information content (PIC) was calculated using the formula, “ $PIC_i = 1 - \sum_{j=1}^n P_{ij}^2$ ” where P_{ij} is the frequency of j^{th} allele for the i^{th} marker and summation extend over k alleles [32].

The genotypic data of 66 polymorphic markers were utilized for genetic diversity analysis. Jaccard’s similarity coefficient was calculated by using the NTSYS-PC software package [33,34]. Cluster analysis was performed using UPGMA and sequential agglomerative hierarchical nested (SAHN) module of NTSYS-PC. The Nei’s pairwise genetic distance neighbour-joining [35] and Shannon’s diversity index (*I*) was calculated using POPGENE v 1.32 (<http://www.ualberta.ca/fyeh>) and MEGA 6 software. The Power-Marker was used for better visualization and understanding the clustering pattern of genotypes. The estimation of population differentiation among and within the genotypes was analyzed by Principal coordinates analysis (PCoA) and AMOVA by using software GeneAEx 6 version 6.501 [36]. AMOVA was used to assess molecular variance within and between populations at 999 permutations.

Structure analysis

Bayesian model-based clustering analysis available in STRUCTURE software 2.3.4 was used for data analysis to obtain possible population structure [37,38]. The software provides the likelihood, classifies according to their population types, and assumes as K. The highest likelihood was interpreted by the corresponding estimate of the basic number of clusters [38]. Each genotype was burned 10,000 and 150,000 steps, followed by 100,000 and 150,000, respectively, using Monte Carlo Markov Chain replicates (MCMC). The K-value was run for 10 times with a K-value ranging from 1 to 10. The optimum K-value was determined by plotting the log posterior probability data to the given K-value. The ΔK value was estimated using the parameters described by Evanno et al. (2005) [39] using online software program Structure Harvester v6.0 (<http://btismysore.in/strplot>). In structure, the value of K is not constant because the distribution of L (K) does not show a clear cutoff point for the true K. An *ad hoc* measure of ΔK was used to detect the numbers of the subgroups. Some independent replicates, the admixture model and allele frequency model (length of burn-in + length of an MCMC repetitions x, number of independent replicates) were also calculated [13,38,40].

QTLs association

The GLM, MLM, Quantile-Quantile (Q-Q) plot and Manhattan plot were used for association analysis of 11 yield related traits, by incorporating Q+K matrices using TASSEL version 5.2.9 [41]. The p-values at <0.005 level of significance were used to determine the significant association of SSR markers. In GLM and MLM, association analysis was performed at 1000 permutations for the correction of multiple testing [42,43]. The False Discovery Rate (FDR) was calculated using SPSS statistical v20. (<http://www-01.ibm.com/support/docview.wss?uid=swg21476447>) at the 5% threshold level for multiple testing to standardise p-value [44]. The false-positive markers-traits association was controlled by applying models Q, K, and Q+K that were compared with each other using quantile-quantile (Q-Q) plot [45].

In-silico study

The *in-silico* study was carried out for analysis of previously reported QTLs and genes associated with respective traits using computer and web-based servers. For this study, several web-servers were used i.e. <http://www.gramene.org/>, <https://www.ncbi.nlm.nih.gov/> and <https://rapdb.dna.affrc.go.jp/> etc. This study helps in searching association between QTLs and genes in rice population.

Results

Phenotypic variation

The grain yield in rice is considered to be one of the most important traits in crop improvement which is influenced by several complex traits. The present study focused on ten yield-related traits that directly or indirectly control the grain yield. The set of 60 genotypes was phenotypically evaluated for grain yield and associated traits under irrigated situation for four consecutive wet seasons. A wide range of phenotypic variations was observed in all the grain yield and 10 yield-related traits (Table 1).

Grain yield varied from 1.19 t/ha (Curinga, 2013) to 9.89 t/ha (N-129, 2014) with mean yield varying from 1.82 t/ha to 8.8 t/ha. Similarly, a wide range of phenotypic variation was observed for all the traits as could be observed from the value of range and mean as depicted in Table 1. Other statistical parameters viz., CV, P value and correlation with grain yield were also calculated to show their importance along the validity.

Table 1. The performance of 60 genotypes in four seasons based on grain yield and yield-related traits.

Traits	Wet Seasons	Mean \pm SD	CV (%)	Range	Correlations (r) with yield	P-value
DFF	2011	100.99 \pm 8.61	8.53	72.5–126.0	0.34	0.006
	2012	101.78 \pm 8.02	7.88	75.78–123.74	0.49	0
	2013	104.44 \pm 6.09	5.83	80.5–116.5	0.35	0.005
	2014	102.03 \pm 7.72	7.56	78.25–124.65	0.5	0.0001
	Mean	102.31 \pm 7.13	6.97	76.76–121.47	0.48	0
PH	2011	109.38 \pm 18.24	16.67	60.65–181.9	0.38	0.001
	2012	115.40 \pm 16.66	14.43	58.34–186.41	0.51	< 0.0001
	2013	109.04 \pm 14.00	12.84	66.96–178.76	0.37	0.004
	2014	107.92 \pm 13.84	12.82	76.55–186.93	0.18	0.176
	Mean	110.43 \pm 14.25	12.91	65.63–183.5	0.42	0.001
TL	2011	8.03 \pm 0.85	10.55	6.2–10.5	0.36	0.001
	2012	8.71 \pm 1.40	16.13	5.83–12	0.31	0.015
	2013	8.70 \pm 1.49	17.08	5–12.6	0.34	0.008
	2014	7.43 \pm 0.85	11.41	5.625–9.55	0.44	0.0005
	Mean	8.22 \pm 0.80	9.75	5.92–9.73	0.56	< 0.0001
PL	2011	28.21 \pm 3.12	11.08	18.25–33.15	0.68	< 0.0001
	2012	28.24 \pm 3.27	11.57	17.19–34.8	0.67	< 0.0001
	2013	27.60 \pm 2.74	9.94	21.27–32.65	0.51	< 0.0001
	2014	27.12 \pm 2.00	7.37	21.26–31.6	0.41	0.0013
	Mean	27.79 \pm 2.34	8.42	20.17–31.45	0.73	< 0.0001
FLL	2011	35.98 \pm 6.49	18.05	17.92–50.55	0.54	< 0.0001
	2012	35.84 \pm 5.40	15.07	17.72–51.84	0.49	< 0.0001
	2013	34.39 \pm 4.35	12.64	23.63–44.30	0.37	0.003
	2014	34.64 \pm 3.72	10.75	26.51–46.12	0.46	0.0002
	Mean	35.22 \pm 4.30	12.21	21.52–48.08	0.17	0.195
FLW	2011	1.43 \pm 0.23	16.35	0.65–1.81	0.53	< 0.0001
	2012	1.55 \pm 0.28	17.86	0.74–2.1	0.49	< 0.0001
	2013	1.46 \pm 0.23	15.88	0.92–1.9	0.28	0.029
	2014	1.37 \pm 0.24	17.66	0.86–2.38	0.35	0.0054
	Mean	1.46 \pm 0.20	13.82	0.78–1.82	0.55	< 0.0001
FG	2011	114.49 \pm 38.69	33.8	63.50–284.2	0.14	0.43
	2012	98.10 \pm 34.95	35.63	59.22–282.22	0.04	0.776
	2013	141.56 \pm 33.66	23.78	70.07–220.71	0.32	0.014
	2014	218.85 \pm 71.19	32.53	75.88–417.5	0.44	0.0005
	Mean	143.25 \pm 30.83	21.52	67.75–246.14	0.38	< 0.0001
TG	2011	136.92 \pm 37.02	27.04	83.28–270.2	0.14	0.431
	2012	115.26 \pm 34.96	30.33	82.30–288.95	-0.02	0.87
	2013	201.91 \pm 53.49	26.49	81.79–329.80	0.33	0.011
	2014	253.05 \pm 87.34	34.51	82.57–517.7	0.37	0.0041
	Mean	176.78 \pm 38.00	21.5	83.29–275.73	0.53	0.002
TGW	2011	24.05 \pm 3.27	13.59	14.77–32.0	0.17	0.239
	2012	24.20 \pm 3.50	14.45	16.64–34.90	0.3	0.021
	2013	22.85 \pm 3.01	13.18	15.24–29.02	-0.27	0.035
	2014	26.25 \pm 4.61	17.57	17.08–38.58	0.02	0.8765
	Mean	24.34 \pm 2.49	10.24	17.50–29.17	0.36	0.005
SLBR	2014	3.93 \pm 0.41	10.5	2.4–4.51	0.15	0.2481
	Mean	3.57 \pm 0.41	11.55	2.44–4.52	0.17	0.566

(Continued)

Table 1. (Continued)

Traits	Wet Seasons	Mean ± SD	CV (%)	Range	Correlations (r) with yield	P-value
YLD	2011	5.54±1.60	28.92	1.22–9.12	-	-
	2012	5.63±1.61	28.56	1.73–9.89	-	-
	2013	4.55±1.35	29.67	1.19–6.59	-	-
	2014	5.90±1.68	28.41	1.95–9.6	-	-
	Mean	5.41±1.48	27.33	1.82–8.80	-	-

Significant level of alpha at 0.05% significant and $p < 0.0001$; the correlation and p-values in bold are different from 0 with a significance level $\alpha = 0.05\%$.

DFW-Days to 50% flowering, PH-Plant height in cm, TL-Tiller number, PL-Panicle length in cm, FLL-Flag leaf length in cm, FLW-flag leaf width in cm, FG-No of fertile grains, TG-Total no of spikelets per panicle, TGW-1000-grain weight in gm, SLBR-Seed length-breadth ratio (SLBR), YLD-Grain yield t/ha.

<https://doi.org/10.1371/journal.pone.0227785.t001>

Some genotypes produced appreciably higher grain yield in respective seasons. Among them, N-129 produced highest grain yield in all the four seasons (9.12 t/ha in 2011; 9.89 t/ha in 2012; 6.59 t/ha in 2013 (productivity affected due to Cyclonic Storm Phailin) and 9.60 t/ha in 2014). It was followed by R-261 (8.01 t/ha in 2011), N-370 (8.78 t/ha in 2012), N-8 (6.34 t/ha in 2013) and N-8 (8.86 t/ha in 2014), respectively. The mean of grain yield in four seasons varied from 1.82 to 8.8 t/ha. Twenty NPTs performed very well consistently in all the four seasons (S2 Table). The average grain yield of four seasons of these 20 genotypes varied from 5.45 to 8.8 t/ha, while popular varieties such as IR64 and MTU1010 produced the maximum yield of 4.80 to 4.99 t/ha (Fig 1A, 1B and 1C, S2 Table).

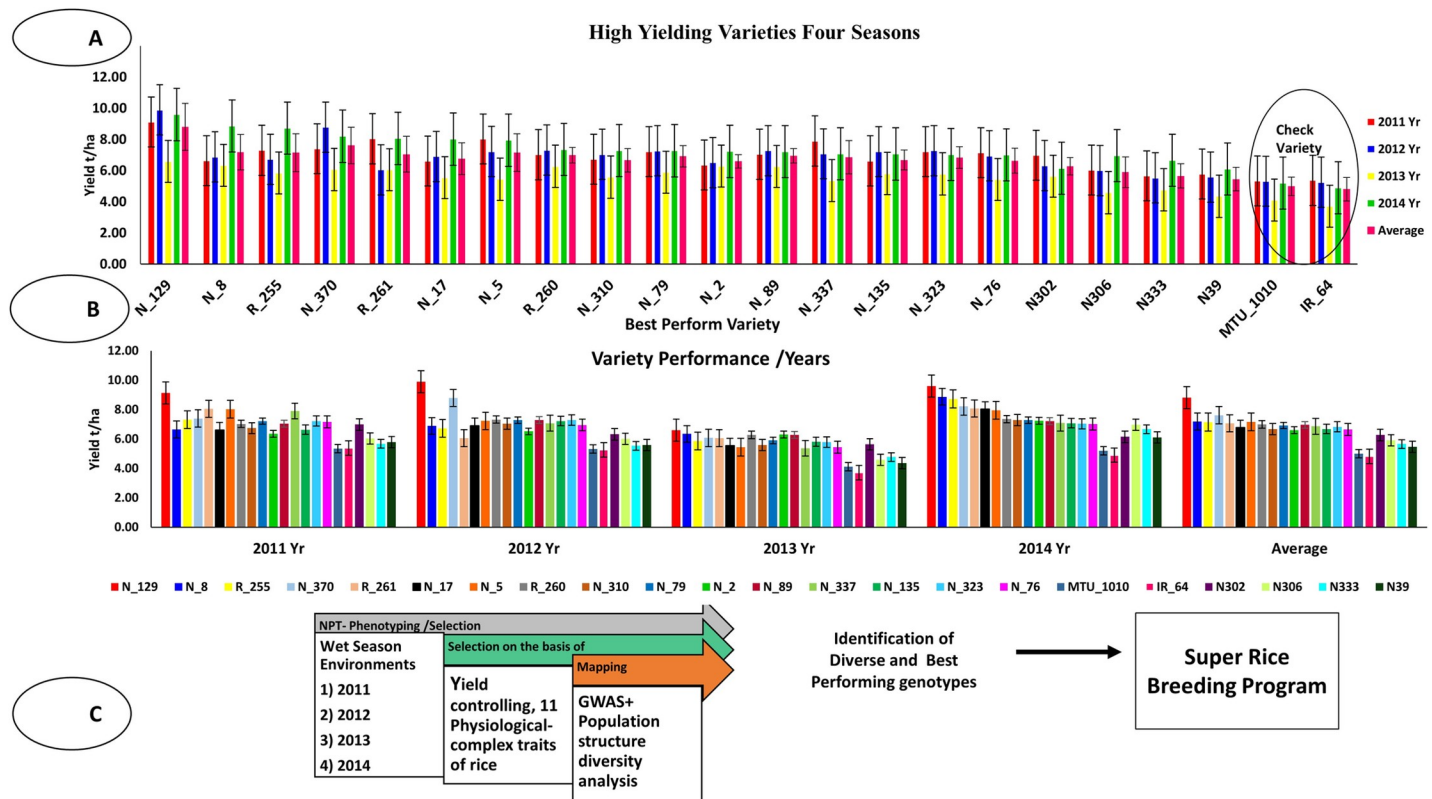


Fig 1. (A,B) Performance of New Plant Types (NPTs) with reference to check varieties based on high grain yield in different seasons. (C) The working hypothesis fulfil the objective of the study. The circle of (A) indicates the performance of standard check varieties i.e. IR-64 and MTU1010.

<https://doi.org/10.1371/journal.pone.0227785.g001>

The CV%, correlation and linear regression analysis of all traits were calculated at 5% level of significance (Table 1, S3 Table). Eight traits viz. DFF, PH, PL, TL, FLW, FG, TG and TGW were positively correlated with grain yield (mean of four-season data) (S3 Table). Similarly, linear regression showed a positive association of six yield contributing traits (PL, DFF, TL, FG, FLL, FLW) with grain yield while four traits (PH, TG, TGW, and SLBR) showed a negative association with grain yield (S4 Table, Fig 2A). The standardized coefficient plots showed that PH, SLBR and TG, were negatively associated grain yield, whereas DFF, TL, PL, FLL, FLW and FG were positively associated with grain yield (S4 Table, Fig 2A). The standardized plot also showed that PH, SLBR, TG and TGW were not directly involved in controlling grain yield. Their significance level might be influenced by environmental factors. The standardized coefficient plot was shown with positive bars for genotypes, which showed grain yield of more than 6.0 t/ha. Similarly, a negative bar of traits having less grain yield is not associated with the traits (S4 Table, Fig 2A).

The PCA Biplot analysis was carried out for focusing on dominant phenotypic traits (Fig 2B). The first PC1 explained 45.67%, while PC2 explained an additional 12.18% of the phenotypic variance. The analysis indicated that the traits viz., TGW, SLBR, TL, FLL, FLW, PL and YLD were predominant for the genotypes situated in the green circle in quadrant I (S5 Table). Similarly, the genotypes belonging to the pink circle were having a preponderance of traits viz., DFF, TG, FG and PH. However, these traits were dominant in quadrant IV (lower right side) (Fig 2B).

The circle represents genotypes are close to each other and have many similarities between them. The genotype belonging to the green and pink circles produced approximately 6–10 t/ha grain yield. The Red and Blue circles were represented by genotypes having less grain yield, i.e. 3–4 t/ha and it was distinguished from corresponding right-side circles. The present study reported a broad range of grain yield in different years, (Table 1, Fig 1). The current study identified sixteen NPT genotypes, which performed better than standard check variety IR64 and MTU1010 (Fig 1). The three best NPT genotypes were identified and the highest grain yield was recorded in N-129 i.e. 9.12 t/ha (WS 2011), 9.59 t/ha (WS 2012), 6.59 t/ha (WS 2013) and 9.6 t/ha (WS 2014), respectively. N-8 showed second highest grain yield in four seasons i.e. 6.63 t/ha (2011), 6.88 t/ha (2012), 6.34 t/ha (2013) and 8.86 t/ha (2014). Similarly, third-highest grain yield was reported for R-255 i.e., 7.30 t/ha (2011), 6.72 t/ha (2012), 5.85 t/ha (2013) and 8.72 t/ha (2014) in four consecutive years. Therefore, these genotypes could be used as a donor for yield-related specific traits.

Heat map helps to understand the specific diversity and dominance pattern between genotypes and traits. Heat map showed that dominant traits were grouped into two significant clusters of genotypes based on certain similarities. Among the traits, the 1st cluster included FG, TG, YLD, TL, FLW, and SLBR, whereas the 2nd cluster comprised TGW, PH, FLL, DFF and PL, which were found very important for specific genotypes (Fig 3). In map red colour traits, i.e., FG, TG, PH and DFF in their particular clusters were showed dominance for respective genotypes. The trait, thousand grain weight (TGW) was dominant in most of the NPT genotypes which belonged to 2nd cluster (Fig 3).

Allelic diversity

Sixty-six (77.65%) out of 85 SSRs were found to be polymorphic. A total of 154 alleles were amplified by 66 polymorphic microsatellite markers with an average of 2.33 alleles per locus. Five markers viz., RM154, RM5709, RM204, RM70 and RM1132 amplified the highest number of alleles (i.e., 4). Two unique alleles (amplified by RM6266 and RM489) and eight rare alleles (5.19%) were identified (Table 2). The marker RM5709 amplified two rare alleles while markers RM168, RM6266, RM489, RM3276, RM528, and RM70 amplified one rare allele each (S6

Fig 2. (A) The standardized coefficient plot has shown the correlation between yield and yield-related traits and genotypes in a regression analysis. The positive bar indicates having a positive correlation with the traits and vice versa for the negative bar. (B) In PCA Biplot analysis, the genotypes were scattered along with their similarity and performance.

<https://doi.org/10.1371/journal.pone.0227785.g002>

Table). The major allele frequency (MAF) varied from 0.33 (RM1132) to 0.98 (RM6266) with an average frequency of 0.71. The genetic diversity varied from 0.033 (RM6266) to 0.732 (RM1132) with an average of 0.39 per locus. The number of effective alleles (N_e) varied from 1.034 (RM489) to 3.733 (RM1132) with an average of 1.78 per locus. The homozygosity (H_o) ranged from 0.262 (RM1132) to 0.967 (RM6266) with an average of 0.60, while genetic heterozygosity (H_e) varied from 0.033 to 0.732. Shannon’s information/diversity index (I) ranged from 0.085 (RM6266) to 1.347 (RM1132) with an average of 0.39 and 0.63 per locus for both the parameters, respectively. The polymorphism information content (PIC) indicates the allelic diversity and frequency of a marker locus with respective genotypes. It varied from 0.516 (RM6266 and RM489) to 0.92 (RM204) with an average of 0.70 (S6 Table). Positive correlations were observed between the total number of alleles (TA), low-frequency alleles (LFA), high-frequency alleles (LFA) and PIC (S6 Table, S7 Table).

Genetic diversity

The Nie’s pairwise genetics analysis by Neighbour-joining tree grouped all the 60 genotypes into three clusters/populations (POP1, POP2 and admixture) (Fig 4A). The first cluster included four sub-clusters containing *Indica*, *Trop. Japonica*, *Tem. Japonica* and one NPT

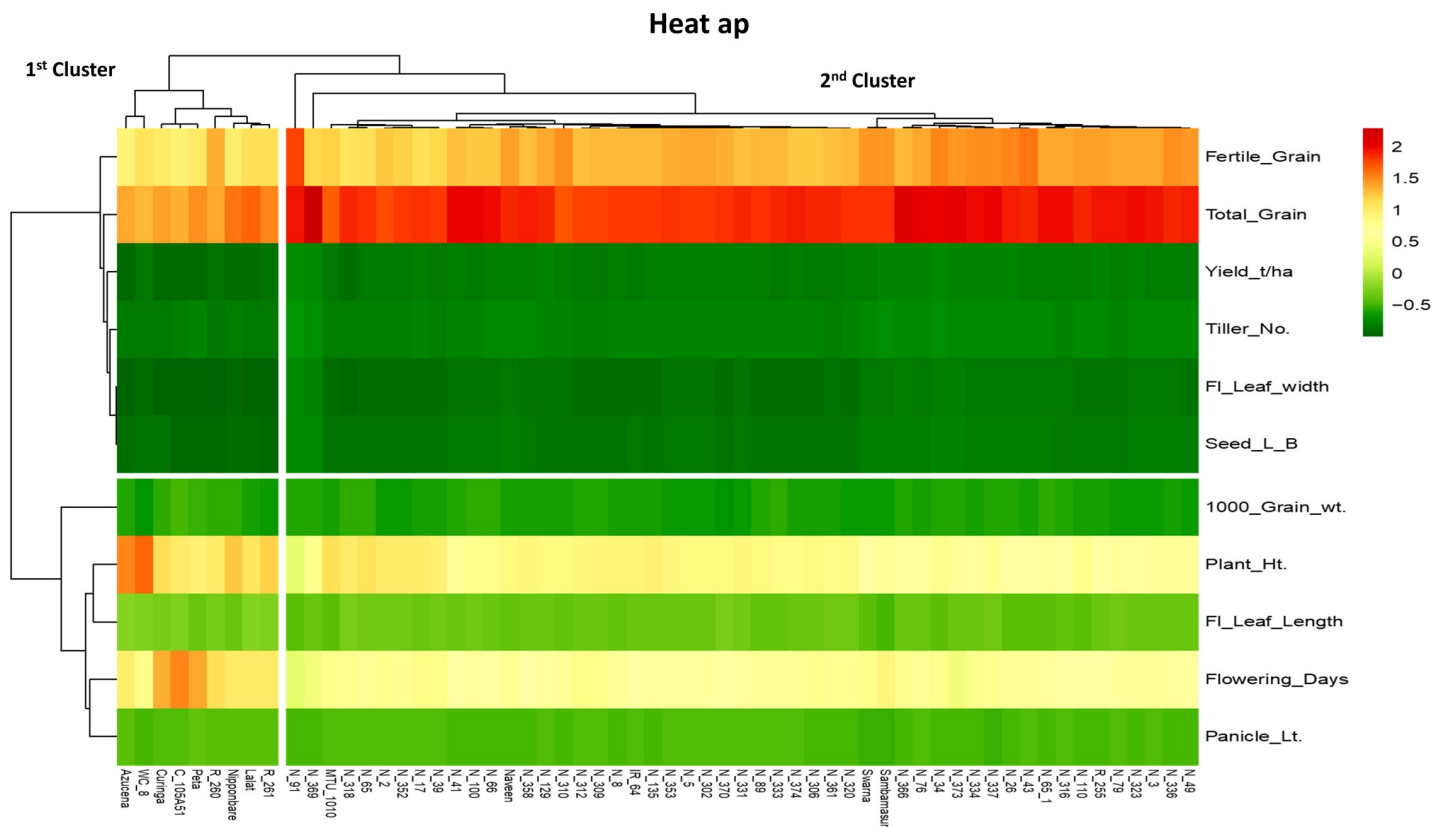


Fig 3. Heat Map showing that both genotypes and traits are grouped into rows and columns by using correlation distance and average matrix.

<https://doi.org/10.1371/journal.pone.0227785.g003>

Table 2. Unique and rare alleles amplified by microsatellite loci in different rice genotypes.

Chromosome No.	Primers	bp	Genotypes
a)Unique allele			
3	RM 6266	160	N-361
3	RM 489	300	N-34
b)Rare alleles			
3	RM168	80	Lalat, Samba Mahsuri
3	RM 6266	160	N-361
3	RM 489	300	N-34
4	RM 3276	190	N-334, N-110, N-318
4	RM 5709	200	C105A-51, AC41009, Nipponbare, N-48
4	RM 5709	220	N-2, N-333, N-320
6	RM 528	260	MTU1010, Samba Mahsuri, R-260
7	RM 70	200	R-266, N-306

<https://doi.org/10.1371/journal.pone.0227785.t002>

genotypes. However, first and second clusters included all of the NPT genotypes, while some of NPTs were found as an admixture in the second cluster (Fig 4A). The NPT genotypes have been derived from *indica*, *tropical* and *temperate japonica*. Hence, two types of populations were observed along with admixture cluster. Both the populations in the trees were observed to be distinctly different (Fig 4A, Table 3, S6 Table).

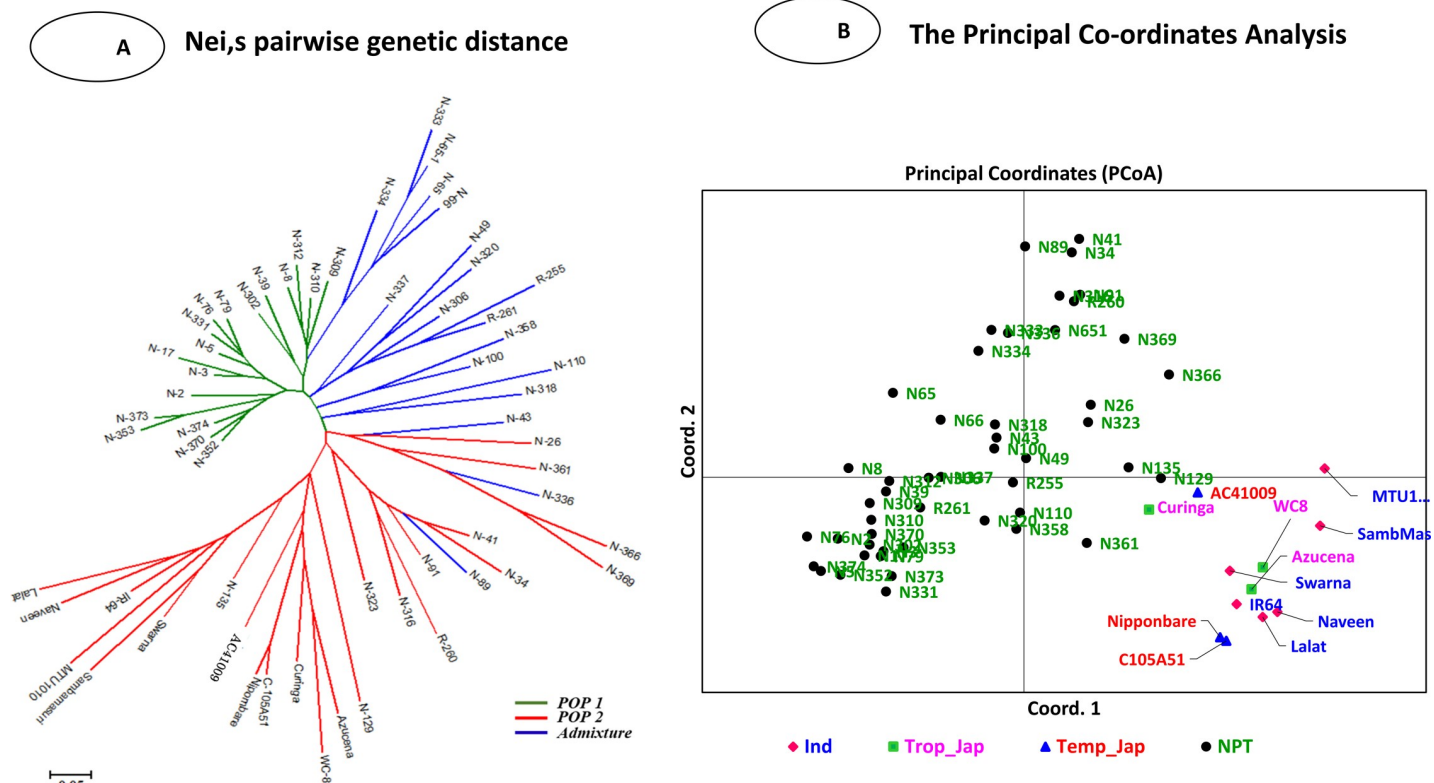


Fig 4. (A) The pairwise genetic distance (Nei, 1973) was calculated by POPGENE v 1.32, and it shows genotypes distributed according to their archetypes. **(B)** The Principal Coordinate analysis (PCoA) of 60 genotypes based on 66 SSR markers. The graph shows the position, and the distribution pattern of each genotype in population space spanned by coordinate 1st versus coordinate 2nd.

<https://doi.org/10.1371/journal.pone.0227785.g004>

Table 3. Grouping of 60 genotypes based on the UPGMA analysis.

Major Clusters	Sub Clusters	Sub-sub clusters	No. of genotypes	Genotypes	Types/ecology
I	1st	A	6	Naveen, IR-64, Lalat, Swarna, Samba Mahsuri, MTU 1010,	Irrigated
	2nd	B	3	Azucena, Curinga, WC-8	Tropical Japonica
	3rd	C	3	C105A51, Nipponbare, AC41009 (Peta)	Temperate japonica
	4th	D	1	N-110	Irrigated
II	5th	E	5	N-334, N-333, N-65-1, N-65, N-66	Irrigated
		F	37	N-323, N-89, N-41, N-91, N-316, N-34, R-260, N-26, N-43, N-49, N-320, R-255, N-306, N-358, N-135, R-261, N-100, N-337, N-2, N-76, N-331, N-79, N-352, N-370, N-374, N-3, N-5, N-17, N-353, N-373, N-8, N-312, N-309, N-310, N-302, N-39, N-318	Irrigated
		G	4	N-366, N-369, N-336, N-361	Irrigated
		H	1	N-129	Irrigated

<https://doi.org/10.1371/journal.pone.0227785.t003>

The Principal Coordinate Analysis (PCoA) differentiated all the 60 genotypes and separated NPTs from *indica* and *japonica* genotypes (Fig 4B). *Japonica* and *indica* genotypes were grouped separately and slightly different from each other. However, many NPT genotypes were grouped into the separate cluster as per neighbour-joining cluster and structure analysis. The PCoA percentage of molecular variance explained by three axes was found to be 12.43%, 7.93%, and 7.45%. In PCoA, the 4th quadrant group showed that some NPTs having an admixture of *indica* and *japonica* populations were intermixing (Fig 4B).

The UPGMA cluster analysis grouped all the 60 genotypes into two major clusters at 54% genetic similarity. The first major cluster (I) was sub-grouped into four sub-clusters, i.e., A, B, C, and D with similarity index varying from 0.54 to 0.92. These sub-clusters contained all the six *indica*, three tropical *japonica*, and three temperate *japonica* and one NPT, respectively (Fig 5, Table 3). Second major cluster (II) contained 47 NPTs with similarity index varied from 0.56 to 0.91. Further, it was sub-grouped into four sub-clusters i.e., E, F, G, and H containing 5, 37, 4 and 1 genotypes, respectively. The sub-cluster D and H contained only one genotype each, i.e., N-110 and N-129, respectively (Fig 5, Table 3).

Analysis of molecular variance (AMOVA)

The two populations obtained through STRUCTURE analysis were used to know the genetic variation between and within clusters using AMOVA. The analysis indicated that there was 8% variation within individuals (genotypes), 68% among individuals within a population, whereas 24% among populations (Fig 6, Table 4). The F-statistics on all three groups was found to be highly significant ($P < 0.001$). The overall F_{st} ($F_{st} = 0.240$) had significant ($P < 0.001$) genetic variation among the three populations (Fig 6, Table 4).

Population structure

The true value of K was identified according to the maximum value of LnP (D) (Pritchard et al., 2000). Structure harvester of Evano table (<http://taylor0.biology.ucla.edu>) analysis showed that at K = 2, the $\Delta K = 179.57$, where value was the highest in both independent burns [39]. The ΔK values were decreased from K = 2 to 10 in general but had a moderate value of 56.29 at K = 4. At K = 4, all the 60 genotypes were divided into four subpopulations, POP1, POP2, POP3, and POP4, which contained 6 *indica*, 3 temperate *japonica*, 3 tropical *japonica*, and 48 NPT genotypes, respectively (Fig 7). The populations POP2, POP3, and POP4 showed

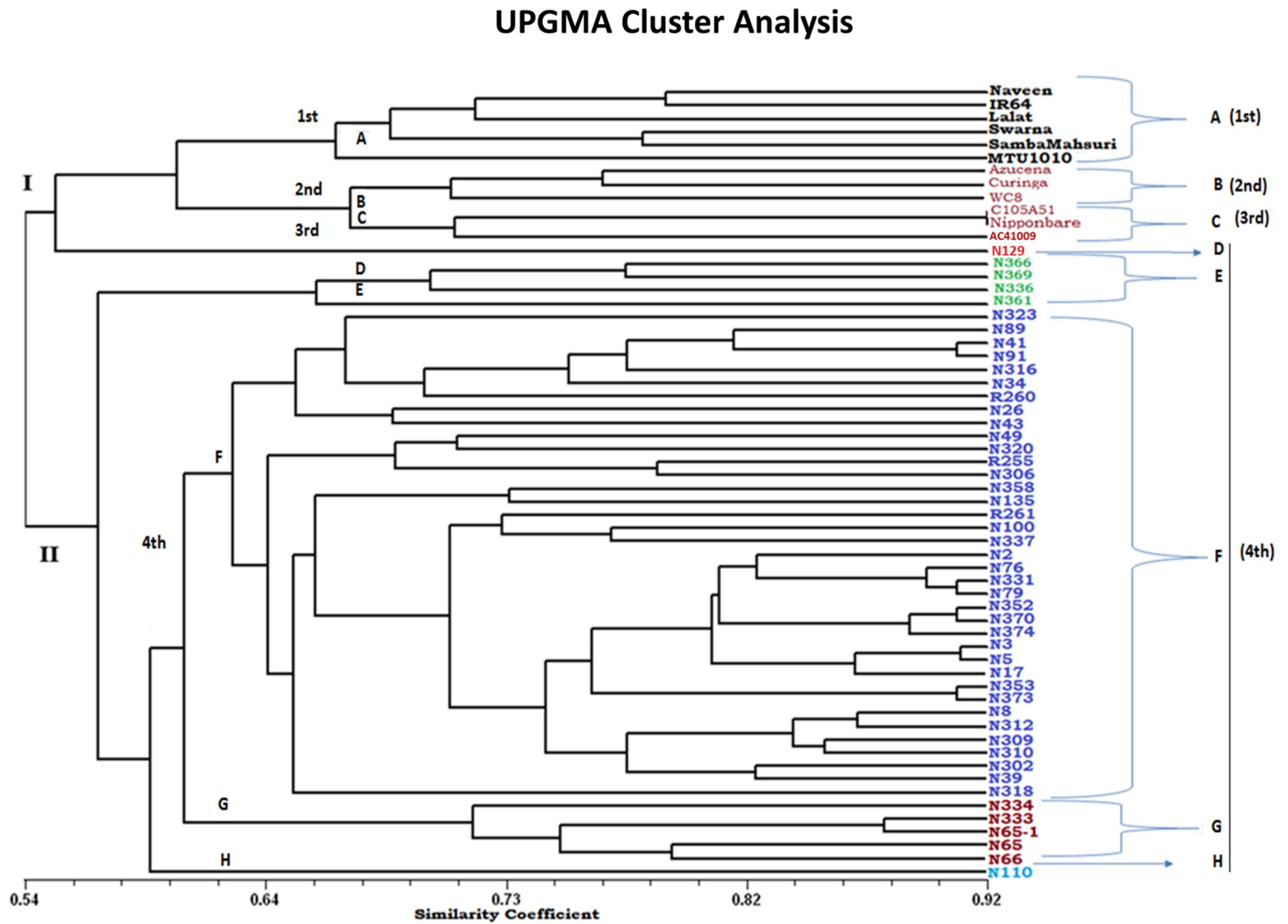


Fig 5. The genotypes clustering based on UPGMA analysis. The 60 genotypes were assigned into four groups (A-1st, B-2nd, C-3rd, and D, E, F, G, H-4th) and similarity varied between 0.54 to 0.92%.

<https://doi.org/10.1371/journal.pone.0227785.g005>

admixture types. Both Pritchard’s and Evonne’s methods confirmed the K-value as 2. Furthermore, analysis of POP gene software showed that 60 genotypes were grouped into two major populations, followed by one admixture group (Fig 7, S1 Fig). The STRUCTURE analysis grouped the genotypes into two types of populations at K = 2, while at K = 3 and K = 4, 60 genotypes were grouped into three and four types of populations, respectively.

QTL association

Marker-trait in general. The GLM and MLM models were used for association analysis at p-value <0.005 and FDR at 5% level. A total of 31 SSR markers were found to be associated with grain yield and other 10 related traits based on individual seasons and their mean data using GLM and MLM models (Table 5, Fig 8). A total 30 and 10 SSRs identified by GLM and MLM models were associated with 70 and 16 QTLs, respectively. Fifteen SSRs (RM6100, RM1132, RM222, RM297, RM154, RM168, RM551, RM5709, RM5575, RM20285, RM5711, RM234, RM26499, RM19 and RM204) were identified to be associated with grain yield *per se*

Percentages of Molecular Variance

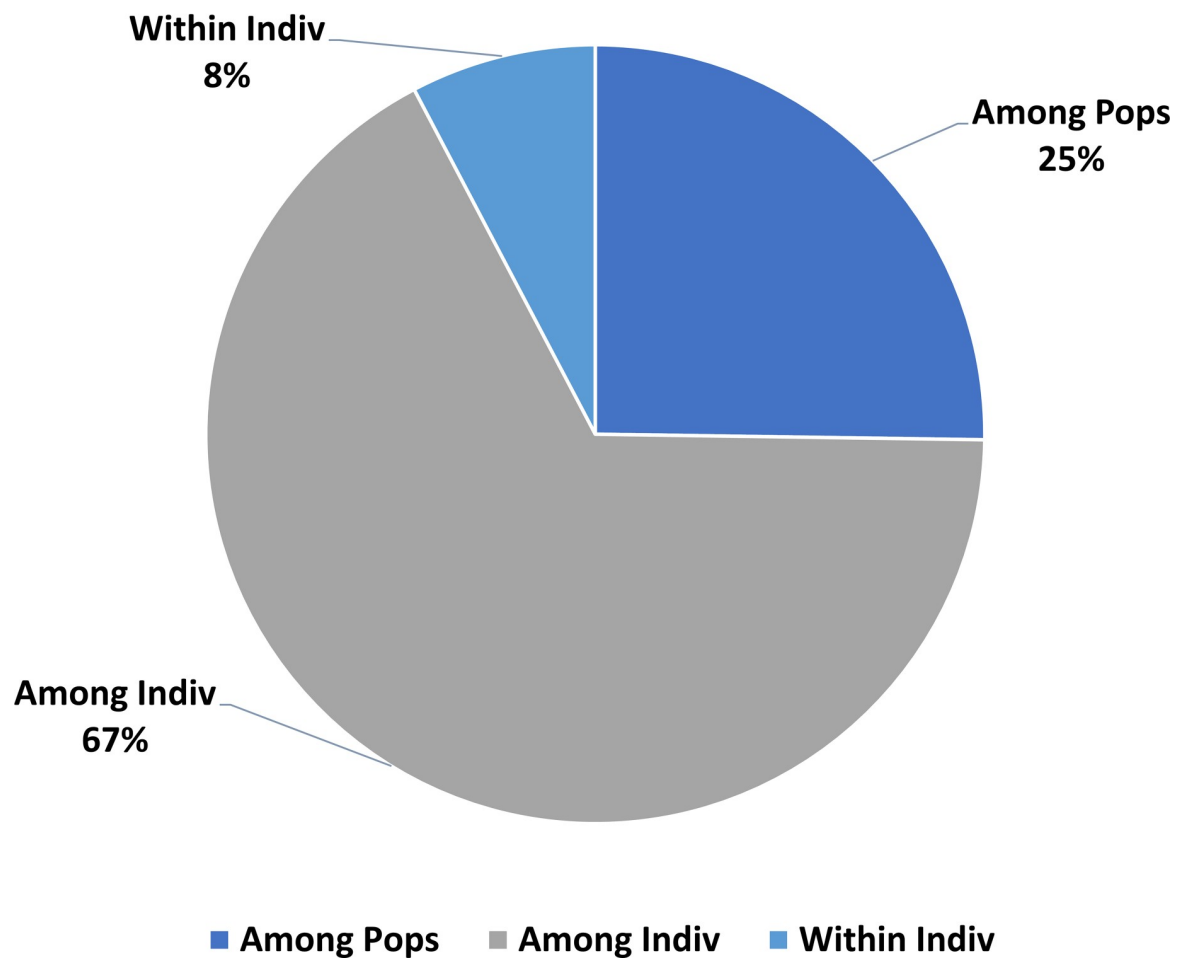


Fig 6. AMOVA-showing molecular distribution pattern within and among populations.

<https://doi.org/10.1371/journal.pone.0227785.g006>

(YLD). Similarly, association analysis led to the identification of five SSRs with DFF, three each with PH, FG and TG, four each with TL, FLL and SLBR, seven SSRs with TGW, 13 SSRs with PL and 14 SSRs with FLW (Table 5, Fig 8).

Table 4. Analysis of molecular variance (AMOVA) among and between populations.

Source of variation	df	SS	MS	Estimated Variation	% of variation	F statistics	Value	P value
Among Populations	2	198.9	99.5	3.8	24%	Fst	0.240	<0.001
Among Individuals within populations	57	1288.8	22.6	10.7	68%	Fis	0.898	<0.001
Within Individuals (genotypes)	59	73.0	1.3	1.3	8%	Fit	0.922	<0.001
Total	118	1560.7		15.7	100%			

MS-Mean sum of square

Fst- F-statistics; subpopulations within the total population

Fis- F-statistics; individuals within subpopulations

<https://doi.org/10.1371/journal.pone.0227785.t004>

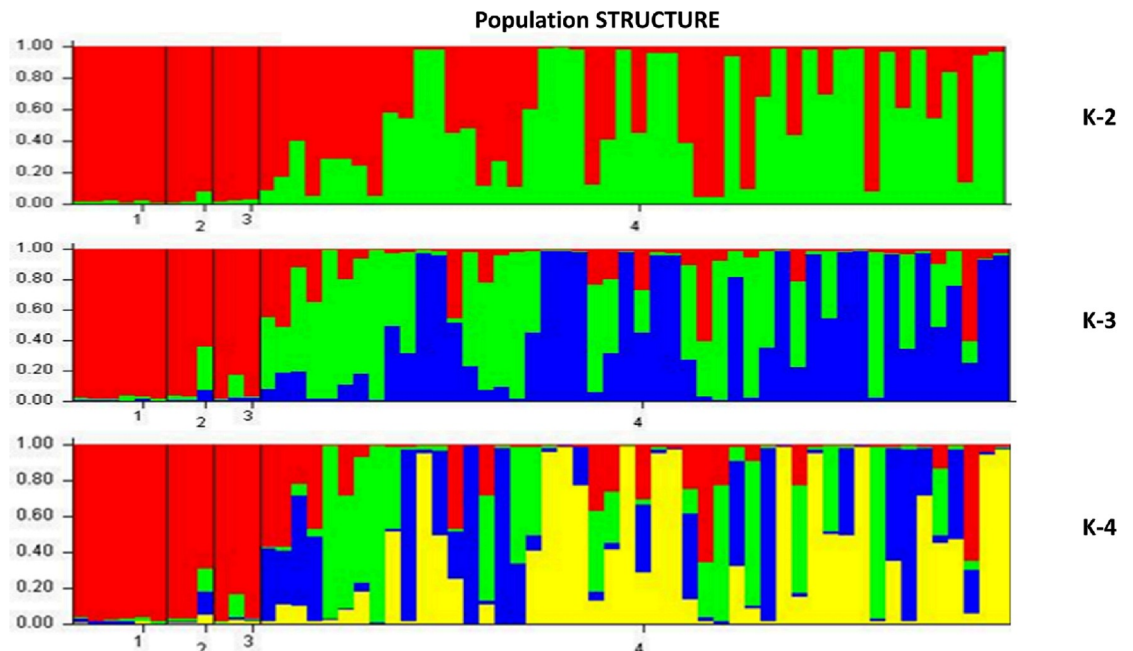


Fig 7. The true value of K was determined by STRUCTURE harvester in K-2 Plot of change in the likelihood of the data, L(K), at values of K from 1 to 10 K = 2.

<https://doi.org/10.1371/journal.pone.0227785.g007>

The MLM analysis identified 10 SSRs that are significantly associated (p -value < 0.005) with 16 QTLs, which were associated with six traits based on the mean data of four seasons at 5% FDR (Table 5). Four SSRs were found to be associated with QTLs considering four seasons data separately as well as their mean data (viz., $qPHT4-a$; $qPL-4.1$ ($qSPP-4.1$); $qFLW-4.1$ and $qTGW-2.1$ ($qTGW-2.1$, $qGn2.1$, $qYLD-2.1$). The SSR marker RM5709 was found to be associated with nine traits, i.e. DFF ($qFD-4.1$), PH ($qPHT4-a$), PL ($qPL-4.1/ qSPP-4.1$) FLL, FLW ($qFLW-a4.1$ and $qFLW-4.1$), FG ($qFGP4.1$), TG ($qTG2.2$) SLBR ($qSlb-4.1$) and YLD ($qYLD4.1$) indicating its association with these major traits. Three markers RM285, RM3827 and RM5709 were found to be associated with QTLs for plant height, $qPHT-6.1$, $qPHT-9.1$ and $qPHT4-a$, respectively (Table 5; Fig 8). Similar marker-QTL association was also recorded for complex traits like TGW ($qTGWa1.1$ and $qTGW-2.1$, $qGn2.1$, $qYLD-2.1$), FLL ($qFLL-9.1$), TG ($qTG-2.1$) and FG ($qFG-7.1$) (Table 5). As two different models have different association with respective traits and therefore, reliability of marker-traits, association would be on the basis of the number of times it was showing association with respective traits in different seasons.

The phenotypic variance contributed by QTLs/SSRs were found to be 13.51% (RM20285) to 26.16% (RM5709, $qFD4.1$) for DFF; 15.15% (RM3827, $qPHT6.1$) to 32.96% (RM5709, $qPHT4a$) for PH; 13.71% (RM3276, $qTL4.1$) to 17.34% (RM287, $qTL11.1/qgpl11.1$) for TL; 13.89% (RM171) to 45.85% (RM5709, $qPL4.1$, $qSPP4.1$) for PL; 16.84% (RM447, $qFLL8.1$) to 25.43% (RM5709) for FLL; 0.7% (RM447, $qFLW-8.1a$) to 30.72% (RM5709, $qFLW4.1$) for FLW; 14.74% (RM324, $qFG7.1$) to 26.19% (RM5709, $qFGP4.1$) for FG; 13.58% (RM324, $qTG2.1$) to 24.30% (RM5709, $qTG2.2$) for TG; 13.20% (RM171, $qTGWa10.1$) to 35.74% (RM263, $qTGW2.1$, $qYLD2.1$, $qGn2.1$) for TGW; 13.57% (RM287, $qSlb11.1$) to 26.79% (RM5709, $qSlb4.1$) for SLBR and 13.14% (RM26499, $qYLD11.1$) to 37.85% (RM5709, $qYLD4.1$) for grain yield (Table 5; Fig 8). More than 25% phenotypic variance was explained

Table 5. Identification of SSR markers associated with grain yield and yield-related traits by using GLM and MLM models.

Traits	Marker	Marker	Chr	cM	GLM at p<0.005			MLM at p<0.005			QTLs associated	Years/Seasons	Reported by
					P-Value	R2	q-Value	P-Value	R2	q-Value			
DFP	212	RM_212**	1	180.2	2.95E-03	0.1849	5.00E-03	-	-	-	<i>qFD-1.1</i>	WS11, WS13	Nov.
DFP	20285	RM_20285	6	121.6	3.86E-03	0.1351	4.00E-03	-	-	-	<i>qFD-6.1</i>	WS11	#
DFP	25	RM_25	8	52.2	1.65E-03	0.2046	0.00E+00	-	-	-	<i>qFD-8.1</i>	WS13	#
DFP	5709	RM_5709 @****	4	109.9	7.67E-04	0.2616	0.00E+00	7.43E-04	0.33692	1.00E-03	<i>qFD-4.1</i>	WS11, WS13, WS14, MEAN	#
DFP	219	RM_219***	9	11.7	3.06E-03	0.1437	0.00E+00	-	-	-	<i>qFD-9.1</i>	WS11, WS14, MEAN	#
PH	3827	RM_3827 @	6	80.4	1.92E-03	0.1566	0.00E+00	4.42E-03	0.15149	1.00E-03	<i>qPHT-6.1</i>	WS14	Nov.
PH	285	RM_285 @@**	9	1.8	1.62E-04	0.26777	0.00E+00	2.36E-03	0.23285	2.00E-03	<i>qPHT-9.1</i>	WS13, WS14	Nov.
PH	5709	RM_5709 @@@@*****	4	109.9	5.96E-05	0.32965	0.00E+00	9.37E-04	0.3262	1.00E-03	<i>qPHT4-a</i>	WS11, WS12, WS13, WS14, MEAN	# [27]
TL	26499	RM_26499	11	0	2.99E-03	0.14213	0.00E+00	-	-	-	<i>qTL-11.1</i>	WS11	Nov.
TL	3276	RM_3276	4	102.4	3.89E-03	0.13714	0.00E+00	-	-	-	<i>qTL-4.1</i>	WS12	Nov.
TL	3827	RM_3827	6	70.4	3.81E-03	0.13769	0.00E+00	-	-	-	<i>qTL-6.1</i>	MEAN	Nov.
TL	287	RM_287 @	11	68.6				2.46E-03	0.17314	3.00E-03	<i>qTL-11.1/gpl 11.1</i>	WS12	#
PL	168	RM_168	3	171.2	3.66E-05	0.30126	3.00E-03	-	-	-	<i>qPL-3.1</i>	WS11	#
PL	171	RM_171	10	92.8	3.36E-03	0.13887	1.00E-03	-	-	-	<i>qPL-10.1</i>	WS11	#
PL	297	RM_297****	1	161.3	6.15E-05	0.2473	0.00E+00	-	-	-	<i>qPL-1.1</i>	WS11, WS12, WS13, MMEAN	Nov.
PL	154	RM_154**	2	4.8	3.36E-03	0.21864	0.00E+00	-	-	-	<i>qPL-2.1</i>	WS11, MEAN	#
PL	551	RM_551**	4	15	1.79E-03	0.15855	0.00E+00	-	-	-	<i>qPL-4.1</i>	WS11, MEAN	#
PL	5709	RM_5709 @@****	4	109.9	1.96E-07	0.45846	0.00E+00	3.23E-03	0.2663	2.00E-03	<i>qPL-4.1, qSPP-4.1</i>	WS11, WS12, WS13, WS14, MEAN	# [27]
PL	5575	RM_5575****	5	0	8.08E-06	0.29704	0.00E+00	-	-	-	<i>qPL-5.1</i>	WS11, WS12, WS13, MMEAN	Nov.
PL	204	RM_204**	6	25.1	3.43E-04	0.28381	0.00E+00	-	-	-	<i>qPL-6.1</i>	WS11, MEAN	# [3]
PL	5711	RM_5711**	7	24.2	2.40E-03	0.19376	0.00E+00	-	-	-	<i>qPL-7.1</i>	WS11, MEAN	Nov.
PL	234	RM_234****	7	88.2	2.14E-04	0.21538	0.00E+00	-	-	-	<i>qPL-7.2</i>	WS11, WS12, WS14, MEAN	#
PL	22899	RM_22899**	8	0	3.20E-03	0.14251	0.00E+00	-	-	-	<i>qPL-8.1</i>	WS11, MEAN	Nov.
PL	447	RM_447	8	124.6	2.44E-03	0.15	0.00E+00	-	-	-	<i>qPL-8</i>	MEAN	# [46]
PL	17	RM_17**	12	109.1	6.48E-04	0.18603	1.00E-03	-	-	-	<i>qPL-12b</i>	WS12, MEAN	# [47]

(Continued)

Table 5. (Continued)

Traits	Marker	Marker	Chr	cM	GLM at p<0.005			MLM at p<0.005			QTLs associated	Years/Seasons	Reported by
					P-Value	R2	q-Value	P-Value	R2	q-Value			
FLW	447	RM_447	8	124.6	1.25E-03	0.16838	1.00E-03	-	-	-	<i>qFLW-8.1</i>	WS14	Nov.
FLW	297	RM_297***	1	161.3	8.23E-04	0.17964	1.00E-03	-	-	-	<i>qFLW-1.1</i>	WS13, WS14, MEAN	# [28]
FLW	5709	RM_5709**	4	109.9	9.92E-04	0.25432	1.00E-03	-	-	-	<i>qFLW-4.1a</i>	WS11, WS12, MEAN	Nov.
FLW	285	RM_285 @@	9	1.8				3.05E-03	0.22189	3.00E-03	<i>qFLW-9.1</i>	WS13	Nov.
FLW	204	RM_204	6	25.1	1.37E-04	0.30374	0.00E+00	-	-	-	<i>qFLW-6.1</i>	WS11	#
FLW	20285	RM_20285	6	121.6	3.39E-03	0.13864	0.00E+00	-	-	-	<i>qFLW-6.2</i>	WS11	Nov.
FLW	206	RM_206	11	102.9	4.51E-03	0.17541	1.00E-03	-	-	-	<i>qFLW-11.1</i>	WS12	#
FLW	19	RM_19	12	20.9	4.33E-03	0.13413	1.00E-03	-	-	-	<i>qFLW-12.1</i>	WS12	#
FLW	297	RM_297****	1	161.3	7.77E-06	0.29796	1.00E-03	-	-	-	<i>qFLW-1.1</i>	WS11, WS12, WS13, MEAN	# [28]
FLW	154	RM_154**	2	4.8	2.79E-03	0.22419	1.00E-03	-	-	-	<i>qFLW-2.1</i>	WS11, MEAN	#
FLW	168	RM_168**	3	171.2	2.38E-03	0.19404	1.00E-03	-	-	-	<i>qFLW-3.1</i>	WS12, MEAN	#
FLW	470	RM_470 @**	4	115.5	6.68E-04	0.18523	1.00E-03	7.78E-04	0.2171	3.00E-03	<i>qFLW-4.1a</i>	WS13, MEAN	#
FLW	5709	RM_5709 @***	4	109.9	1.43E-04	0.30717	1.00E-03	4.22E-03	0.25441	3.00E-03	<i>qFLW4.1</i>	WS11, WS12, MEAN	# [27]
FLW	5575	RM_5575***	5	0	2.19E-04	0.21472	1.00E-03	-	-	-	<i>qFLW-5.1</i>	WS11, WS12, MEAN	Nov.
FLW	1132	RM_1132	7	83.3	2.59E-03	0.22647	1.00E-03	-	-	-	<i>HFLW-7</i>	MEAN	# [30]
FLW	22899	RM_22899****	8	0	6.24E-04	0.18707	2.00E-03	-	-	-	<i>qFLW-8.1</i>	WS11, WS12, WS13, MEAN	Nov.
FLW	5711	RM_5711 @	7	24.2				3.46E-03	0.21637	3.00E-03	<i>qFG-7.1</i>	WS12	Nov.
FLW	447	RM_447	8	124.6	0.0013	0.073	0.001				<i>qFLW-8.1a</i>	WS14	#
FG	1132	RM_1132	7	83.3	1.30E-03	0.2465	1.00E-03	-	-	-	<i>qFG-7.1</i>	WS14	Nov.
FG	5709	RM_5709	4	109.9	7.60E-04	0.26185	2.00E-03	-	-	-	<i>qFGP-4.1</i>	MEAN	# [27]
FG	324	RM_324 @	7	68.9				4.59E-03	0.14743	4.00E-03	<i>qFG-7.1</i>	WS11	#
TG	1132	RM_1132	7	83.3	4.34E-03	0.21091	2.00E-03	-	-	-	<i>qFG-a7.1</i>	WS14	Nov.
TG	324	RM_324 @	2	68.9	4.08E-03	0.13583	2.00E-03	3.43E-03	0.15788	3.00E-03	<i>qTG-2.1</i>	MEAN	#
TG	5709	RM_5709**	4	109.9	1.47E-03	0.24301	2.00E-03	-	-	-	<i>qTG-2.2</i>	WS13, MEAN	Nov.
TGW	297	RM_297 @**	1	161.3	9.08E-05	0.23383	0.00E+00	4.24E-03	0.15024	3.00E-03	<i>qTGW-a1.1</i>	WS11, WS12	Nov.
TGW	201	RM_201**	9	81.2	2.29E-03	0.14928	0.00E+00	-	-	-	<i>qTGW-a9.1</i>	WS11, WS12	#

(Continued)

Table 5. (Continued)

Traits	Marker	Marker	Chr	cM	GLM at p<0.005			MLM at p<0.005			QTLs associated	Years/Seasons	Reported by
					P-Value	R2	q-Value	P-Value	R2	q-Value			
TGW	219	RM_219	9	11.7	2.87E-03	0.14557	2.00E-03	-	-	-	<i>qTGW-a9.2</i>	WS13	Nov.
TGW	171	RM_171	10	92.8	4.68E-03	0.13202	2.00E-03	-	-	-	<i>qTGW-a10.1</i>	WS14	#
TGW	263	RM263 @****	2	127.5	5.75E-07	0.3574	2.00E-03	1.58E-03	0.18982	3.00E-03	<i>qTGW-2.1, qGn2.1, qYLD-2.1</i>	WS12, WS13, WS14, MEAN	# [48]
TGW	5575	RM_5575***	5	0	2.12E-03	0.15383	2.00E-03	-	-	-	<i>qTGW-5.1</i>	WS11, WS12, MEAN	Nov.
TGW	22899	RM_22899**	8	0	1.99E-03	0.15563	2.00E-03	-	-	-	<i>qTGW-8.1</i>	WS12, MEAN	Nov.
SLBR	204	RM_204	6	25.1	9.68E-04	0.25113	0.00E+00	-	-	-	<i>qSlb-a6.1</i>	WS11	#
SLBR	1256	RM_1256*****	3	48.2	3.95E-04	0.19923	4.00E-03	-	-	-	<i>qSlb-3.1</i>	WS11, WS12, WS13, WS14, MEAN	Nov.
SLBR	5709	RM_5709 @@@@*****	4	109.9	1.47E-03	0.24306	5.00E-03	3.18E-03	0.26788	4.00E-03	<i>qSlb-4.1</i>	WS11, WS12, WS13, WS14, MEAN	#
SLBR	287	RM_287****	11	68.6	4.09E-03	0.13571	5.00E-03	-	-	-	<i>qSlb-11.1</i>	WS12, WS13, WS14, MEAN	#
YLD	6100	RM_6100 @	10	53.9				3.10E-03	0.16153	2.00E-03	<i>qYLD-10.1</i>	WS11	Nov.
YLD	1132	RM_1132**	7	83.3	3.86E-03	0.21448	3.00E-03	-	-	-	<i>qYLD-a7.1</i>	WS13, WS14	Nov.
YLD	222	RM_222	10	11.3	2.33E-03	0.15124	4.00E-03	-	-	-	<i>qYLD-a10.1</i>	WS13	#
YLD	297	RM_297**	1	161.3	2.76E-03	0.14664	2.00E-03	-	-	-	<i>qYLD-1.1</i>	WS13, MEAN	#
YLD	154	RM_154****	2	4.8	4.40E-04	0.27707	3.00E-03	-	-	-	<i>qYLD-2.1, qts1</i>	WS12, WS13, WS14, MEAN	# [26]
YLD	168	RM_168***	3	171.2	4.79E-03	0.17369	3.00E-03	-	-	-	<i>qYLD-3.1</i>	WS13, WS14, MEAN	#
YLD	551	RM_551****	4	15	3.27E-03	0.14196	3.00E-03	-	-	-	<i>qYLD-4, qPL-4.1</i>	WS11, WS13, MEAN	# [25]
YLD	5709	RM_5709*****	4	109.9	7.93E-06	0.37847	3.00E-03	-	-	-	<i>qYLD-4.1</i>	WS11, WS12, WS13, WS14, MEAN	#
YLD	5575	RM_5575*****	5	0	6.43E-06	0.30246	3.00E-03	-	-	-	<i>qYLD-5.1</i>	WS11, WS12, WS13, WS14, MEAN	Nov.
YLD	20285	RM_20285*****	6	121.6	5.89E-04	0.18858	3.00E-03	-	-	-	<i>qYLD-6.1</i>	WS11, WS12, WS13, WS14, MEAN	Nov.
YLD	5711	RM_5711***	7	24.2	2.54E-03	0.19217	3.00E-03	-	-	-	<i>qYLD-7.1</i>	WS12, WS13, MEAN	# [24]
YLD	234	RM_234*****	7	88.2	1.43E-03	0.16474	3.00E-03	-	-	-	<i>qPpl7.1</i>	WS11, WS12, WS13, WS14, MEAN	# [29]
YLD	26499	RM_26499	11	0	4.78E-03	0.1314	4.00E-03	-	-	-	<i>qYLD-11.1</i>	MEAN	Nov.
YLD	19	RM_19****	12	20.9	1.10E-03	0.17176	4.00E-03	-	-	-	<i>qYLD-12.1, qSpn-12.1</i>	WS12, WS13, WS14, MEAN	# [49]

(Continued)

Table 5. (Continued)

Traits	Marker	Marker	Chr	cM	GLM at p<0.005			MLM at p<0.005			QTLs associated	Years/Seasons	Reported by
					P-Value	R2	q-Value	P-Value	R2	q-Value			
YLD	204	RM_204**	6	25.1	0.00027	0.28637	0.00E+00	-	-	-	qYLD-6.1	WS11, WS14	#
Total Association of Markers					70			16					

MLM = @ -Markers associated with MLM in the specific -season /environment

** = Indicates SSR markers associated in two seasons/ environments

*** = Indicates SSR markers associated in three seasons/ environments

**** = Indicates SSR markers associated in four seasons/ environments

= Indicates the QTLs association with SSRs, estimated map-position (cM) and physical position (bp) were reported in a gramene web (<http://www.gramene.org/markers>) and NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Nov. = Indicates Novel QTLs identified, supposed to be linked to respective traits

WS-Wet season

<https://doi.org/10.1371/journal.pone.0227785.t005>

by QTLs bracketing RM5709 for each of nine traits, DFF, PH, PL, FLL, FLW, FG, TG, SLBR and YLD. This indicated that RM5709 would be useful for transfer of above nine traits into popular rice varieties.

2D Plot

Marker	Chr.	DFF	PH	TL	PL	FLL	FLW	FG	TG	TGW	YLD	SLBR
RM20285	6										*****	
RM212	1	**										
RM25	8											
RM1132	7										**	
RM324	7							@	@			
RM5709	4	**** @	@.@.@*****		@.@*****	**	@*****		**		*****	@.@.@*****
RM285	9		@.@**			@@						
RM447	8											
RM154	2				**		**				****	
RM19	12										****	
RM204	6				**						**	
RM206	11											
RM470	4						@**					
RM5575	5				****		***		***		*****	
RM5711	7				**		@				***	
RM168	3						**				***	
RM17	12				**							
RM171	10											
RM22899	8				**		****		**			
RM234	7				****						*****	
RM551	4				**						***	
RM1256	3											*****
RM201	9											
RM219	9	***										
RM297	1				****	@.@	****		@**		**	
RM263	2								@*****			
RM26499	11											
RM287	11			@								****
RM3276	4											
RM3827	6		@									
RM222	10											
RM6100	10										0 @	

Fig 8. The 2D-Plot showing SSR marker association with respective phenotypic traits using GLM (p-0.005) and MLM (p-0.005) models. MLM = @ -Markers associated with MLM in the specific -season /environment; **/@@ = Indicates SSR markers associated in two seasons/ environments; ***/@@@ = Indicates SSR markers associated in three seasons/ environments; ****/@@@@ = Indicates SSR markers associated in four seasons/ environments.

<https://doi.org/10.1371/journal.pone.0227785.g008>

Table 6. Identification of SSR markers associated with more than one trait.

Marker	Chr	Position	Co-localization Traits	Gene ID	Locus ID	Function	Reference
RM_297	1	32,099,566–32,099,760	PL, FLL, FLW, TGW, YLD				[50]
RM_154	2	1,083,820–1,084,090	PL, FLW, YLD	Os02g0120800	LOC_Os02g02840.1	Similar to Small GTP-binding protein, EST Rice root cDNA, mRNA	[50–52]
RM_168	3	28,091,534–28,091,727	PL, FLW, YLD				[50, 53]
RM_5709	4	31,875,481–31,875,663	DFF, PH, PL, FLL, FLW, FG, TG, YLD, SLBR	Os04g0619500	LOC_Os04g52850.1	Ovarian tumour, otubain domain containing protein.	[50–52]
RM_551	4	177,080–177,271	PL, YLD				[50, 53]
RM_5575	5	23,345,264–23,345,401	PL, FLW, TGW, YLD	Os05g0474600	LOC_Os05g39680.1, LOC_Os05g39690.1	Similar to Aldose reductase-related protein	[50–52]
RM_20285	6	22,420,889–22,421,088	DFF, FLW, YLD	Os06g0561800	LOC_Os06g36650.1	ABC transporter, transmembrane domain containing protein	[50–52]
RM_204	6	3,168,314–3,168,547	PL, FLW, YLD, SLBR	Os06g0162800	LOC_Os06g06750.1	MADS-box transcription factor, Control of spikelet morphogenesis, Regulation of floral meristem	[50–52]
RM_3827	6	22,297,146–22,297,320	PH, TL	Os06g0560300	LOC_Os06g36480.1	Similar to NAC-type transcription factor.	[50–52]
RM_1132	7	23,984,489–23,984,581	FLW, FG, TG, YLD				[50]
RM_324	7	11,389,704–11,389,942	FG, TG				[50]
RM_5711	7	3,141,181–3,141,329	PL, FLW, YLD	Os07g0157700	LOC_Os07g06390.2, LOC_Os07g06390.3, LOC_Os07g06390.4	Conserved hypothetical protein.	[50–52]
RM_234	7	25,472,688–25,472,820	PL, YLD				[50]
RM_447	8	26,546,992–26,547,102	PL, FLL, FLW	Os08g0531200	LOC_Os08g41900.1	WD40 repeat domain containing protein	[50–52]
RM_22899	8	14,762,955–14,763,118	PL, FLW, TGW	Os08g0331900	LOC_Os08g24310.1	Similar to Leucine Rich Repeat family protein, expressed	[50–53]
RM_285	9	1.8–1.8 cM	PH, FLL				[50]
RM_171	10	19,048,795–19,049,123	PL, TGW	Os10g0489900	LOC_Os10g34820.1, LOC_Os10g34820.2	Similar to CDT1a protein	[50–52]
RM_26499	11	11,001,140–11,001,342	TL, YLD				[50] IRGSP-1
RM_287	11	16,767,319–16,767,617	TL, SLBR				[50] IRGSP-1
RM_19	12	20.9–20.9 cM	FLW, YLD				[50]/IRGSP-1 (2005)

<https://doi.org/10.1371/journal.pone.0227785.t006>

Twenty SSR markers were found to be significantly associated with more than one trait (Table 6). RM5709 was found to be associated with nine traits while RM275 was found to be associated with five traits. Similarly, RM5575, RM204 and RM1133 were found to be associated with four traits each, while RM154, RM168, RM20285, RM5711, RM447, RM22899 were found to be associated with three traits each. Nine SSR markers were found to be associated with two traits (Table 6).

Nineteen SSR markers were found to be associated with different traits in more than one season. Four SSRs i.e., RM5709, RM5575, RM20285 and RM234 were found to be associated with PL, PH, SLBR and YLD and common for 4 seasons. Among them, RM5709 is co-localized

with PH, PL, SLBR and YLD in four seasons. Ten SSRs were co-localized in three seasons with seven traits i.e., RM5709 (DFF, PH), RM297 (PL), RM5575 (PL), RM234 (PL), RM22899 (FLW), RM263 (TGW), RM1256 (SLBR), RM287 (SLBR), RM154 (YLD) and RM19 (YLD). Similarly, 12 SSRs were found to be associated with two seasons viz., RM212 (DFF), RM219 (DFF), RM285 (DFF), RM297 (FLL, TGW), RM5709 (FLW, FLL), RM201 (TGW), RM5575 (TGW), RM1132 (YLD), RM168 (YLD), RM551 (YLD), RM5711 (YLD) and RM204 (YLD). Thirty QTLs were identified as novel based earlier information. At least one QTL was found to be novel for each of 11 traits. Number of novel QTLs found to be 1, 2, 3, 4, 3, 4, 1, 2, 4, 1 and 5, respectively for DFF, PH, TLL, PL, FLL, FLW, FG, TG, TGW, SLBR and YLD traits (Tables 5 and 6).

The association of traits with markers could be confirmed in the 2D plot and Q-Q Plot (Fig 8, S2 Fig, Table 5, Table 6). The QQ-plot showed a similar distribution of marker-trait association for 11 traits (S2 Fig). The GLM Manhattan plot shows 26 SSR markers associated with grain yield at p-value at 0.05 (S3 Fig). However, seven SSRs were associated with grain yield-related traits in MLM Manhattan plot at p-value 0.005 (S4 Fig). The lowest p-value 8.73×10^{-4} was found with marker RM5709 for plant height trait, followed by 0.00158 (Thousand-grain weight) with RM263 and 0.00266 (Flag leaf width) with RM5709 (Table 5).

In-silico study for marker co-localization

The present study has used the computer and web-based servers' big data to confirm the association of co-localizing genes and QTLs linked with yield-related traits of rice. Twenty SSRs were identified that co-localized with grain yield and related traits (Table 6). Using *in-silico* approach, it was found that 10 out of 20 co-localized SSR markers were in agreement with previous reports. These 10 co-localized SSRs viz., RM154, RM5709, RM5575, RM20285, RM204, RM3827, RM5711, RM447, RM22899 and RM171 were found to be very important because of their association with grain yield-related traits. RM5709 found to be highest co-localized on 4th chromosome (associated with 9 yield-related traits viz., DFF, PH, PL, FLL, FLW, FG, TG, YLD and SLBR), followed by RM297 (associated with PL, FLL, FLW, TGW and YLD) and RM5575 (associated with PL, FLW, TGW and YLD) (Table 6).

Discussion

Phenotypic variance

Improving rice yield potential is one of the primary breeding objectives in many countries for several decades [5]. In 1960s and 1980s, the green revolution was initiated with the development of semi-dwarf High Yielding Varieties (HYVs) like IR 8 and IR 36 [2,9,54–57]. The main objective of the green revolution was to fulfil and achieve self-sufficiency in food requirement, which helped the developing countries around the world especially in South Asia. It was realized in rice due to development of semi-dwarf, lodging resistant and fertilizer responsive high yielding varieties. It led to stability in rice production and mitigating the hunger of growing population. It was accomplished in mid-sixties with the development of miracle variety IR8. Since then, a stagnant in yield potential of semi-dwarf *indica* inbred varieties was noticed in *indica* inbreds, which needs to be cracked [54].

New Plant Types (NPTs) was a potential approach for breaking the yield ceiling. The initial effort on NPT was made by IRRI scientists to develop 2nd generation NPT genotypes accumulating favorable alleles from tropical *japonicas* and popular *indica* for yield-related traits with multi-environment testing [5,54]. The main idea behind NPT development was to develop a plant type endowed with combination of unique traits that would help for efficient photosynthesis and biomass partitioning for very high grain yield in irrigated ecology. In this process, favorable *tropical japonica* genes were accumulated in *indica* background in second-

generation NPT lines. IRRI scientists identified highly potential genotypes, i.e., IR 72967-12-2-3 which reportedly produced 10.16 ton/ha [9]. Our main target area for breaking yield ceiling was in eastern zone of India, which has many climatic constraints particularly low light due cloudy weather in wet season. In current study, the advance generation 2nd generation NPTs were phenotypically screened for high grain yield and associated traits in four seasons at NRRI, Cuttack, India. Phenotyping is the most crucial step for crop improvement. Identification of suitable transgressive segregants for the specific quantitative trait in any crops is a challenging task for the breeders. At the outset, a set of such elite materials of NPTs was chosen (advance generation segregating materials) as initial materials. Trait specific selection and evaluation of these materials subsequently led to identification 48 NPTs with variable grain yield, which were subjected to multi-environment testing. In this study, potential 20 NPTs were identified with an average yield in the range of 5.45 to 8.8 tons/ha. These genotypes could be utilised directly or as prospective parents based on yield *per-se*.

PCA Bi-plot analysis showed association with major yield-related traits (Fig 2B, S5 Table). The PC1 and PC 2 explained 45.67% and 12.18% of the total variance, respectively. Similar variability were also reported for PC1 and PC2 viz. 35.2% and 14.4%, respectively [58]. The distribution pattern of traits clearly differentiated the genotypes and relative importance of traits, which influenced the grain yield (YLD). The positive relation was observed among genotypes in the first quadrant, which showed the importance of traits viz., PL, FLL, FLW, TL, YLD, SLBR, and TGW particularly in NPTs. The genotypes were associated with respective traits in a 1st quadrant, which could be responsible for better average grain yield. The first quadrant consisted of important traits and the genotypes endowed with those traits predominantly, hence could be selected as donor parents for specific traits in NPTs. Similarly, traits viz., TG, FG, DFF, and PH were predominant for the genotypes in quadrant IV (S5 Table), and it could be selected as a donor based on a number of fertile spikelets and effective tiller number.

The present study reported that dominant phenotypic traits such as PL, TL, FLL, FG and TG, had a positive correlation with yield. However, the more focused selection should be done for those traits (PH, FLW, TGW and SLBR) that are showing weak correlation with grain yield, because of environmental effects (Table 1, S3 Table). The best genotypes were assigned for grain yield based on phenotypes which are N-129, N-8, and R-255 (Fig 1A and 1B, S2 Table, S3 Table). The dominant specific traits and genotypes for high grain yield could be selected for designing effective breeding strategy. This would be helpful to the breeder for the proper choice of a parent/donors for bi-parental/multi-parental mating vis-à-vis during the process of selection in segregating generations. Therefore, present study reports phenotyping followed by different statistical analysis which suggests trait-specific genotypes for prospective parents in the hybridization programs for breeding super rice [22,58]

The Heatmap shows a data matrix in the form of the colour pattern due to the numeric differences in multivariate data. In ClustVis, hierarchical clustering can be optionally applied on specific traits those were linked with particular genotypes and observations [22]. The Heatmap analysis showed the order of colour merging with the specific traits that are playing a vital role in the association for targeted trait. The different colouring patterns were linked with respective traits starting from deep green to red (Fig 3). Apart from green, the white and light colour traits indicated relatively poor association with yield-related traits viz., PH, DFF and SLBR traits. As the colour moves along the colour chart from green to red, the association with yield improves with grain yield that means the red colour especially was strongly associated. In this context, the order of association based on the colour intensity starts with FG, followed by TG, TGW, FLL, TL, PL and FLW. The respective genotypes have been depicted with their strong associated traits colours. The dense colour gives ideas for a strong character. So that breeder can easily choose donor parents that are actively associated with specific traits [22] (Fig 3).

Allelic and genetic diversity

The utility of SSR markers for population structure, diversity and association mapping depends on the quality of information they provide. The allelic and genetic diversity helps the breeder to understand genetic constitution of germplasm makeup and target donor selection for designing effective future breeding strategy. The 66 out of 85 SSR markers (77.64%) showed polymorphism, which amplified 154 alleles. Similarly, Anandan and his team reported the 39 polymorphic SSRs which amplified 128 alleles [59]. The average PIC value in this study was found to be 0.70, which was similar to previous reports [22,58,60–63]. The lower rate of the average PIC was reported in association studies by several workers {0.31 [64]; 0.47 [65,66]}. The PIC value showed a positive correlation with the total number of alleles (S6 Table, S7 Table). Similar findings were reported by previous researchers [42,60]. Moderate levels of genetic diversity (i.e. 0.39) was observed among 60 genotypes used in the present association study. Similarly, Cui et al. (2013) detected an average diversity of 0.34 in 347 genotypes used for association mapping in cold tolerance at the booting stage [64]. However, a higher rate of average genetic diversity was reported by some workers (0.69: Zhao et al., 2013; 0.52: Nachimuthu et al., 2015; 0.76: Edzesi et al., 2016) [15,66,67].

Population structure

The population structure analysis helps to understand and differentiate the types of population groups existing in a set of genotypes. The population structure based on Bayesian clustering model [15, 66, 67] has been most frequently used to correct spurious associations. The delta K value was measured by ad-hoc and based on the relative rate of change in likelihood LnP (D). The Delta k = 2 was set to get a higher likelihood optimal number of LnP (D) among groups.

The 60 genotypes were differentiated into four sub-populations. Similar sample sizes were used by several researchers in association analysis in rice [68,69] and alfalfa [70]. The UPGMA cluster analysis grouped 60 genotypes into two major groups at 54% of genetic similarity. The Nei's pairwise genetic distance showed three types of populations, i.e. POP1, POP2, and one admixture population. Similarly, at K = 2, STRUCTURE analysis could differentiate entire populations into two subpopulations (Fig 7). Genotypes in these populations along with high mean values could be utilized as potential parents for transgressive segregants with high yield and yield attributing traits towards breaking yield ceiling

The mining through the Power Marker into the details of individual groups revealed that the first population (POP1) contained hardcore NPTs, which was distinctly different from *indica* (*Ind*), *temperate japonica* (*Temp.*) and *tropical japonica* (*Trop.*) (2nd population, POP2, which also includes few NPTs along with *Ind*, *Temp.* and *Trop.*). However, all NPTs contain the genomes of *indica* as well as *temperate* and *tropical japonicas*. Moreover, the population has one admixture group, which lies in between the two classes comprising the characters of both populations. Therefore, a targeted hybridization between consciously selected parents of these two distant groups might result in transgressive segregants with super rice traits for future yield enhancement. At K = 4, the population was clustered into four groups viz., *Ind* (1st), *Trop* (2nd), *Temp* (3rd) and NPTs (4th) according to their genotypic and evolutionary significance. However, this study suggested that popular varieties clustered together according to their ecology, morphology and inter-varietal hybrid fertility of rice varieties in *indica* and *japonica* [58,71]. Here, almost all the NPTs were grouped separately, except one i.e., N-129. Moreover, the genotypes in the 4th group comprised the genomes from *indica*, and *japonica* and supposed to have a relationship with the first three clusters. The population cluster 1, 2 and 3 were purer and divergent, but in the 4th cluster, genotypes were intermediates of first

three clusters. This could help breeders in devising necessary breeding strategy and choosing parents for yield improvement.

QTL association

QTL association has been widely used for the identification and mapping of QTLs for various traits such as tolerance to biotic and abiotic stresses, quality and grain yield in different crops [11,12,18,30,48,72]. This study also targets findings new QTLs, alleles, and genes [73] and validate the previously reported QTLs. The present association study was conducted on 60 diverse genotypes panel and 85 SSR markers. The present association study focused on identification QTLs associated with yield and related traits in relatively small population and with limited markers [74]. Therefore, our study is analogous to previous reports with a small, focused group of genotypes and limited marker pairs combination [12,13,59,66,68,74–76].

In association studies, both GLM and MLM models are used. Population stratification and cryptic relationships are two common reasons for the inflation of false-positive association [38]. GLM model has more false positive association as compared to MLM model analysis [19,41,70,72,77]. It does not consider to influence the population structure and kinship [70,78]. MLM model has higher accuracy and a smaller number of spurious marker-trait association with genotypes as compared to GLM model. This model is having a powerful algorithm, which systematically increases power, improves calibration and reduce computational cost to structured populations generally used for SNPs in GWAS [45,72]. The MLM model integrates structure and kinship matrix (Q+K) which supposedly corrects the false-positive error to the tune of 62.5%. Hence, MLM model has been popularly used in several cases for marker-traits association [12,18,43,45,71,79–83].

In association mapping, mixed model (Q+K) showed a significant improvement in goodness of fit and reducing spurious associations. The K and Q matrix corrected the association between eleven phenotypic traits with markers [43,70] at permutation value is 1000 at $p < 0.005$ for GLM and MLM for the level of significance. In association mapping, p-value plays an essential role because it controls over the level of false-positive association between traits and markers. It means that if the p-value is minimized, there is less chance of a false positive association of markers with respective traits and vice-versa [45]. The value of p is in agreement with previous reports [70–72,84]. However, some researchers reported their results at $p < 0.05$ – 0.01 value, which is much higher compared to our study, where the number of markers is appreciably high [43,70].

GLM identified 30 SSRs which shows 70 associations with grain yield *per se* and yield-related traits based on the four-season mean data (mean value of 4 seasons) (Fig 8). It was found that 11 common SSRs were found between GLM and MLM model and had a positive association with yield-related traits based on four-season mean data (Table 5). Twenty-three SSR markers have been reported previously and these markers have a positive association with QTL regions based on mean data. There were 15 SSRs linked with different QTLs responsible for grain yield *per se* in four different seasons. Out of them, five SSR loci were in agreement to the previous studies. Previous reports indicated that the markers RM154, RM551, RM5711, RM234 and RM19 were associated with grain yield QTLs, (Table 5) viz., *qYLD2.1* & *qts1*, *qYLD-4* & *qPL-4.1*, *qYLD-7.1*, and *qYLD-12.1*, *qSpn-12.1*, respectively [12,24,25,29,58,85]. Similarly, the marker RM5709 has been well documented in the www.gramene.org database for its association with grain yield QTL. Five SSR markers, RM6100, RM1132, RM5575, RM20285 and RM26499 are linked with novel QTLs, *qYLD10.1*, *qYLDa7.1*, *qYLD-5.1*, *qYLD6.1* and *qYLD11.1* responsible for imparting high grain yield. In case of tiller number, three out of four SSR markers RM26499, RM3276 and RM3827 were found to be associated with the novel

QTLs, i.e. *qTL-11.1*, *qTL-4.1* and *qTL-6.1*, respectively. For panicle length (PL), previous researchers have reported the association of RM5709 with *qPL-4.1* and *qSPP-4.1* [27]; RM204 with *qPL-6.1* [3]; RM234 with *qPPL-7.2* [29]; RM447 with *qPL-8* [20,86] and RM17 with *qPL-12b*. Four out of 13 SSRs viz., RM297, RM5575, RM5711, and RM22899 were found to be associated with four novel QTLs, *qPL-1.1*, *qPL-5.1* and *qPL-8.1*, respectively for panicle length (PL). However, Marathi et al. (2012) reported that RM5709 marker was linked with *qSPP-4.1*, indicating the pleiotropic effect of *qPL4.1* on panicle length. The present study reported that a total of 14 SSRs were associated with FLW and out of them, 3 SSRs (RM297, RM5709, RM1132) were reported previously [27,28,30,66,67]. Four SSRs, RM447, RM297, RM5709 and RM285 have been identified for flag leaf length (FLL) in QTL regions of *qFLL8.1*, *qFLL-1.1* [28], *qFLL-4.1a* and *qFLL9.1*, respectively. Out of the above, two markers, RM297 and RM5709 were also associated with QTLs for flag leaf width (FLW), *qFLW-1.1* and *qFLW4.1*, respectively due to pleiotropic effects [27,28]. Similar reports corroborate the present finding of marker-trait association for RM1132 apart from three other reported markers [30]. It is suggested that one locus may be involved in conferring multiple traits, which may be the result of the gene to gene interactions and pleiotropic effect. Three markers QTL associations viz., RM154~*qFLW2.1*, RM168~*qFLW3.1*, and RM470~*qFLW4.1* were identified for the flag leaf width (FLW) (<http://www.gramene.org/>). Four novel QTLs, *qFLW-6.2*, *qFLW-5.1*, *qFLW-8.1* and *qFLW-7.1* were identified for controlling FLW trait in our study.

The RM5709 was reported to be linked with QTL, *qPHT4-a* for plant height [27]. Number of fertile grains is considered as important trait, because of its link with grain yield. Same SSR marker (RM5709) was also found to be associated with *qFGP4.1* [27]. Thousand-grain weight (TGW) is another crucial trait supposedly linked to yield. Our study identified seven QTLs viz., *qTGW-a1.1*, *qTGW-a9.1*, *qTGW-a9.2*, *qTGW-a10.1*, *qTGW2.1*, *qTGW-5.1* and *qTGW-8.1* associated with SSRs viz., RM297, RM201, RM219, RM171, RM263, RM5575 and RM22899, respectively. These QTLs might be highly useful in the rice breeding programs. Four QTLs were found to be novel i.e., *qTGW-a1.1*, *qTGW-a9.2*, *qTGW-5.1* and *qTGW-8.1*, which could be emphasized because of their better grain filling and boldness leading to higher grain yield. Marri et al. (2005) reported the link of RM263 with *qTGW-2.1*, *qGn2.1*, and *qYLD-2.1*, indicating the possibility of pleiotropic effects of *qGn2.1* [48].

Twenty SSRs were having association with more than one traits and have been reported in the gramene-database (<http://www.gramene.org/>) by earlier studies (Table 6) [27,28,48,87]. These were significantly associated with yield controlling complex traits viz., PH, DFF, SLBR, TL, PL, FLL, FLW, FG, TG, TGW and YLD and supposed to have played a significant role in yield enhancement (Table 5, Table 6, Fig 8). Similar reports by Zhang and team (2014) revealed the pleiotropic effect of gene *LSCHL-4*, in influencing increment of leaf chlorophyll, enlargement of flag leaf size, higher panicle branch and grains per panicles [30]. The previous report suggests that specific marker association with more than one trait might be either due to the linkage of genes or pleiotropic effects of a single locus [88–90]. However, variation in population structure, QTL detection methods and environmental conditions restrict our choices to compare the newly identified QTLs with the already reported QTLs. Therefore, we do have a need for further study on these potential trait-specific genotypes. It would lead to design effective breeding strategy for introgression of high grain yielding traits associated QTLs into popular rice varieties for obtaining super rice targeting to overcome yield ceiling.

The *in-silico* study of co-localization of SSR markers with respective traits would be helpful to the breeders to confirm the association of trait-specific SSRs. The present study has reported 10 SSR markers, which are associated with grain yield-related traits and found with gene IDs (Table 6). Most of the SSRs were co-localized with more than two traits. The highest co-localization was identified in RM5709 linked with nine traits followed by RM297 co-localized with

five traits. Similarly, RM5575, RM204 and RM22899 were also recorded to be co-localized with more than one trait and could be rated as important. This marker could be useful in marker-assisted backcross breeding program to produce next-generation super rice.

Conclusion

Breaking yield ceiling in rice warrants conscious selection of parents. Sufficient variability was available in NPT genotypes, because of the genetic distance of parents using tropical *japonicas* and *indicas*, leading to fixation of distinct lines. In the present study, microsatellite markers were used for association studies for grain yield and ten yield-related traits. Few highly-potential genotypes with high yield along with variation in morpho-physiological traits were identified after conducting the trial in four consecutive years. Wide variations were found in all the traits, which would be helpful for the identification of genotypes required for bi-parental/multi-parental crosses in developing super rice genotypes with higher grain yield. The STRUCTURE and tree diagrams were helpful in the classification of populations into distinct clusters vis-à-vis uniqueness among them and helped in the identification of a diverse gene pool for necessary parental selection for targeted transgressive segregants. The *in-silico* study reported twenty SSR markers; those were associated with more than one trait. Nineteen, SSR markers were found to be associated with different traits in more than one season. Four SSRs such as, RM5709, RM5575, RM20285 and RM234 were found to be associated with PL, PH, SLBR and YLD and common for four seasons. More than 25% phenotypic variance was explained by QTLs bracketing RM5709 for each of nine traits, DFF, PH, PL, FLL, FLW, FG, TG, SLBR and YLD. RM5709 would be useful for transfer above nine traits into popular rice varieties. The present study reported that 16 SSRs were linked with 11 yield traits and were found to be associated with 30 novel QTLs. Some of the QTLs are very important viz., TL (*qTL-6.1*), PL (*qPL-1.1*, *qPL-5.1*, *qPL-7.1*, *qPL-8.1*), FLL (*qFLL-9.1*), FLW (*qFLW-5.1*, *qFLW-8.1*), TGW (*qTG-2.2*, *qTGW-5.1*, *qTGW-8.1*), SLBR (*qSlb-3*) and YLD (*qYLD-5.1*, *qYLD-6.1a*, *qYLD-7.1*, *qYLD-11.1*) because of their association with important yield contributing traits or yield *per se* for multiple seasons. Hence, these could be pyramided in elite background for realization of higher yield and breaking yield ceiling. The study would be immensely helpful for selecting target donors with requisite traits for designing effective future breeding strategy for super rice.

Supporting information

S1 Fig. The true value of K was determined by STRUCTURE harvester in K-2 Plot of change in the likelihood of the data, L (K), at values of K from 1 to 10 K = 2.

(TIF)

S2 Fig. QQ-plot showing distribution of marker-trait association for 11 traits.

(TIF)

S3 Fig. GLM Manhattan plot showing markers associated with grain yield using significant p value at 0.05.

(TIF)

S4 Fig. MLM Manhattan plot showing markers associated with grain yield using significant p value at 0.005.

(TIF)

S1 Table. List of 48 New Plant Types (NPT) and 12 popular DUS reference genotypes used in association mapping analysis.

(DOCX)

S2 Table. Grain Yield performance of 20 best varieties and standard checks under irrigated condition.

(DOCX)

S3 Table. Correlation matrix of grain yield and their association with 10 yield-related traits.

(DOCX)

S4 Table. Calculation of standardized coefficients.

(DOCX)

S5 Table. Distribution pattern of genotypes in Principal Component Analysis (PCA) and Biplot by using morphological-physiological data.

(DOCX)

S6 Table. Molecular diversity among 60 rice genotypes based on the alleles amplified by 66 polymorphic SSR markers.

(DOCX)

S7 Table. Correlation between PIC and different types of alleles.

(DOCX)

Acknowledgments

We thank Director, ICAR-National Rice Research Institute (NRRI), Cuttack, India, for providing all laboratory and field facilities and IRRI for providing the NPT seeds.

Author Contributions

Conceptualization: Ravindra Donde, Khirod Kumar Sahoo, Lambodar Behera, Sushanta Kumar Dash.

Data curation: Ravindra Donde, Shibani Mohapatra, S. K. Yasin Baksh, Barada Padhy, Mitadru Mukherjee, Somnath Roy, Krishnendu Chattopadhyay, A. Anandan, Padmini Swain, Khirod Kumar Sahoo, Onkar Nath Singh, Lambodar Behera.

Formal analysis: Ravindra Donde, S. K. Yasin Baksh, Barada Padhy, Mitadru Mukherjee, Somnath Roy, A. Anandan, Padmini Swain, Khirod Kumar Sahoo, Onkar Nath Singh, Lambodar Behera.

Investigation: Barada Padhy, Padmini Swain, Khirod Kumar Sahoo, Lambodar Behera, Sushanta Kumar Dash.

Methodology: Ravindra Donde, Shibani Mohapatra, S. K. Yasin Baksh, Barada Padhy, Mitadru Mukherjee, Padmini Swain, Lambodar Behera.

Resources: Krishnendu Chattopadhyay, Khirod Kumar Sahoo, Onkar Nath Singh, Lambodar Behera, Sushanta Kumar Dash.

Software: Ravindra Donde, Somnath Roy.

Supervision: Krishnendu Chattopadhyay, Onkar Nath Singh, Lambodar Behera.

Validation: Ravindra Donde, Somnath Roy, Khirod Kumar Sahoo, Lambodar Behera.

Visualization: Ravindra Donde.

Writing – original draft: Ravindra Donde.

Writing – review & editing: Lambodar Behera, Sushanta Kumar Dash.

References

1. Varshney Bansal, Aggarwal Datta, Craufurd. Agricultural biotechnology for crop improvement in a variable climate: hope or hype? Trends in plant science. 2011; 16: 363–371. <https://doi.org/10.1016/j.tplants.2011.03.004> PMID: 21497543
2. Peng Khush, Virk Tang, Zou. Progress in ideotype breeding to increase rice yield potential. Field Crops Research. 2008; 108: 32–38.
3. Suji Biji, Poornima Prince, Amudha Kavitha, et al. Mapping QTLs for plant phenology and production traits using indica rice (*Oryza sativa* L.) lines adapted to rainfed environment. Molecular biotechnology. 2012; 52: 151–160. <https://doi.org/10.1007/s12033-011-9482-7> PMID: 22198727
4. Donde Samal, Rout. Studies on Agrobacterium mediated genetic transformation of rice (*Oryza sativa* L.) for drought tolerance using Dreb 1 gene. 2014; 4: 341–350.
5. Khush. What it will take to feed 5.0 billion rice consumers in 2030. Plant molecular biology. 2005; 59: 1–6. <https://doi.org/10.1007/s11103-005-2159-5> PMID: 16217597
6. Donde Gupta, Gouda Kumar, Vadde Sahoo, et al. Computational characterization of structural and functional roles of DREB1A, DREB1B and DREB1C in enhancing cold tolerance in rice plant. Amino acids. 2019; 51: 839–853. <https://doi.org/10.1007/s00726-019-02727-0> PMID: 30900088
7. Gouda Gupta, Donde Kumar, Vadde Mohapatra, et al. Computational approach towards understanding structural and functional role of cytokinin oxidase/dehydrogenase 2 (CKX2) in enhancing grain yield in rice plant. Journal of Biomolecular Structure and Dynamics. 2019; 1–10.
8. Dingkuhn Schnier, Datta De, Dorffling Javellana. Relationships between ripening-phase productivity and crop duration, canopy photosynthesis and senescence in transplanted and direct-seeded lowland rice. Field Crops Research. 1991; 26: 327–345.
9. Peng Laza, Vesperas Khush, Virk Zhu. Rice: progress in breaking the yield ceiling. In Proceedings of the 4th International Crop Science Congress, vol. 26, 2004.
10. Dash Meher, Behera Anandan, Azharudheen Barik, et al. Genetic diversity of New Plant Type rice selections in relation to indica, tropical japonicas, temperate japonicas and irrigated cultures. ORYZA—An International Journal on Rice. 2015; 52: 266–274.
11. Donde Kumar, Gouda Gupta, Mukherjee Baksh, et al. Assessment of Genetic Diversity of Drought Tolerant and Susceptible Rice Genotypes Using Microsatellite Markers. Rice Science. 2019; 26: 239–247.
12. Abdurakhmonov Abdulkarimov. Application of association mapping to understanding the genetic diversity of plant germplasm resources. International Journal of Plant Genomics. 2008; 2008.
13. Agrama Eizenga, Yan. Association mapping of yield and its components in rice cultivars. Molecular Breeding. 2007; 19: 341–356.
14. Huang Zhao, Li Wang, Zhao Li, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nature genetics. 2012; 44: 32–39.
15. Zhao Chung, Kwon Lee, Ma Park. Association analysis of physicochemical traits on eating quality in rice (*Oryza sativa* L.). Euphytica. 2013; 191: 9–21.
16. Lu Zhang, Niu Wang, Xu Feng, et al. Genetic variation and association mapping for 12 agronomic traits in indica rice. BMC genomics. 2015; 16: 1067. <https://doi.org/10.1186/s12864-015-2245-2> PMID: 26673149
17. Zhang Xu, Akash McCouch, Oard. Identification of candidate markers associated with agronomic traits in rice using discriminant analysis. Theoretical and applied genetics. 2005; 110: 721–729. <https://doi.org/10.1007/s00122-004-1898-z> PMID: 15678327
18. Swamy Shamsudin, Rahman Mauleon, Ratnam Cruz, et al. Association mapping of yield and yield-related traits under reproductive stage drought stress in rice (*Oryza sativa* L.). Rice. 2017; 10: 21. <https://doi.org/10.1186/s12284-017-0161-6> PMID: 28523639
19. Rebolledo Peña, Duitama Cruz, Dingkuhn Grenier, et al. Combining image analysis, genome wide association studies and different field trials to reveal stable genetic regions related to panicle architecture and the number of spikelets per panicle in rice. Frontiers in plant science. 2016; 7.
20. Ma Feng, Wei Mei, Xu Chen, et al. Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. Frontiers in plant science. 2016; 7: 1801. <https://doi.org/10.3389/fpls.2016.01801> PMID: 27965699
21. Bagchi Sanghamitra, Berliner Chattopadhyay, Sarkar Kumar, et al. Assessment of physicochemical, functional and nutritional properties of raw and traditional popped rice. 2016;

22. Metsalu Vilo. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic acids research*. 2015; 43: W566–W570. <https://doi.org/10.1093/nar/gkv468> PMID: 25969447
23. Paradis Claude, Strimmer. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 2004; 20: 289–90. <https://doi.org/10.1093/bioinformatics/btg412> PMID: 14734327
24. Fukuta Konisho, Senoo-Namai Yanagihara, Tsunematsu Fukuo, et al. Genetic characterization of rainfed upland New Rice for Africa (NERICA) varieties. *Breeding science*. 2012; 62: 27–37. <https://doi.org/10.1270/jsbbs.62.27> PMID: 23136511
25. Liu Liu, Wu Zeng, Thi Liang, et al. Identification of a candidate gene for panicle length in rice (*Oryza sativa* L.) via association and linkage analysis. *Frontiers in plant science*. 2016; 7.
26. Liu Meng, He Zhu, Shen Zhang. Fine Mapping and Candidate Gene Analysis of the Tiller Suppression Gene *ts1* in Rice. *PloS one*. 2017; 12: e0170574. <https://doi.org/10.1371/journal.pone.0170574> PMID: 28107441
27. Marathi Guleria, Mohapatra Parsad, Mariappan Kurungara, et al. QTL analysis of novel genomic regions associated with yield and yield related traits in new plant type based recombinant inbred lines of rice (*Oryza sativa* L.). *BMC plant biology*. 2012; 12: 137. <https://doi.org/10.1186/1471-2229-12-137> PMID: 22876968
28. Selvaraju Shanmugasundaram, Mohankumar Asaithambi, Balasaraswathi. Detection of quantitative trait locus for leafroller (*Cnaphalocrocis medinalis* (Guenée)) resistance in rice on linkage group 1 based on damage score and flag leaf width. *Euphytica*. 2007; 157: 35–43.
29. Thomson Tai, McClung Lai, Hinga Lobos, et al. Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theoretical and applied genetics*. 2003; 107: 479–493. <https://doi.org/10.1007/s00122-003-1270-8> PMID: 12736777
30. Zhang Li, Wang Ye, Zeng Rao, et al. LSCHL4 from japonica cultivar, which is allelic to NAL1, increases yield of indica super rice 93–11. *Molecular plant*. 2014; 7: 1350–1364. <https://doi.org/10.1093/mp/ssu055> PMID: 24795339
31. Liu Muse. PowerMarker V3. 0 Manual. 2017;
32. Anderson Churchill, Autrique Tanksley, Sorrells. Optimizing parental selection for genetic linkage maps. *Genome*. 1993; 36: 181–186. <https://doi.org/10.1139/g93-024> PMID: 18469981
33. Raju Mohankumar, Sumanth Rajanna, Udayakumar Prasad, et al. Discovery of QTLs for water mining and water use efficiency traits in rice under water-limited condition through association mapping. *Molecular breeding*. 2016; 36: 35.
34. Rohlf. NTSYS-pc: numerical taxonomy and multivariate analysis system. *Applied Biostatistics*; 1992.
35. Nei Roychoudhury. Sampling variances of heterozygosity and genetic distance. *Genetics*. 1974; 76: 379–390. PMID: 4822472
36. Peakall Smouse. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources*. 2006; 6: 288–295.
37. Hubisz Falush, Stephens Pritchard. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour*. 2009; 9: 1322–32. <https://doi.org/10.1111/j.1755-0998.2009.02591.x> PMID: 21564903
38. Pritchard Stephens, Donnelly. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*. 2000; 155: 945–959. PMID: 10835412
39. Evanno Regnaut, Goudet. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*. 2005; 14: 2611\textendash2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
40. Gilbert Andrew, Bock Franklin, Kane Moore, et al. Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Molecular Ecology*. 2012; 21: 4925–4930. <https://doi.org/10.1111/j.1365-294X.2012.05754.x> PMID: 22998190
41. Bradbury Zhang, Kroon Casstevens, Ramdoss Buckler. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23: 2633–5. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
42. Jin Lu, Xiao Sun, Corke Bao. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor Appl Genet*. 2010; 121: 475–87. <https://doi.org/10.1007/s00122-010-1324-7> PMID: 20364375
43. Qin Chen, Yi Bie, Zhang Zhang, et al. Identification of associated SSR markers for yield component and fiber quality traits based on frame map and upland cotton collections. *PloS one*. 2015; 10: e0118073. <https://doi.org/10.1371/journal.pone.0118073> PMID: 25635680

44. Fahrenkrog Neves, Resende Vazquez, Campos Dervinis, et al. Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *New Phytologist*. 2017; 213: 799–811. <https://doi.org/10.1111/nph.14154> PMID: 27596807
45. Wang Tian, Pan Buckler, Zhang. A SUPER powerful method for genome wide association study. *PloS one*. 2014; 9: e107684. <https://doi.org/10.1371/journal.pone.0107684> PMID: 25247812
46. Ma Bao, Guo Zeng, Li Ji, et al. Quantitative trait Loci for panicle layer uniformity identified in doubled haploid lines of rice in two environments. *J Integr Plant Biol*. 2009; 51: 818–24. <https://doi.org/10.1111/j.1744-7909.2009.00854.x> PMID: 19723240
47. Ahamadi Fotokian, Fabriki-Orang. Detection of QTLs influencing panicle length, panicle grain number and panicle grain sterility in rice (*Oryza sativa* L.). *J Crop Sci Biotech*. 2008; 11: 163–170.
48. Marri Sarla, Reddy Siddiq. Identification and mapping of yield and yield related QTLs from an Indian accession of *Oryza rufipogon*. *BMC genetics*. 2005; 6: 33. <https://doi.org/10.1186/1471-2156-6-33> PMID: 15949048
49. Yue Xue, Xiong Yu, Luo Cui, et al. Genetic basis of drought resistance at reproductive stage in rice: separation of drought tolerance from drought avoidance. *Genetics*. 2006; 172: 1213–1228. <https://doi.org/10.1534/genetics.105.045062> PMID: 16272419
50. International. The map-based sequence of the rice genome. *Nature*. 2005; 436: 793. <https://doi.org/10.1038/nature03895> PMID: 16100779
51. Itoh Tanaka, Barrero Yamasaki, Fujii Hilton, et al. Curated genome annotation of *Oryza sativa* ssp. japonica and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res*. 2007; 17: 175–83. <https://doi.org/10.1101/gr.5509507> PMID: 17210932
52. Ohyanagi Tanaka, Sakai Shigemoto, Yamaguchi Habara, et al. The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. japonica genome information. *Nucleic Acids Res*. 2006; 34: D741–4. <https://doi.org/10.1093/nar/gkj094> PMID: 16381971
53. Annotation Rice. The rice annotation project database (RAP-DB): 2008 update. *Nucleic Acids Research*. 2007; 36: D1028–D1033. <https://doi.org/10.1093/nar/gkm978> PMID: 18089549
54. Peng Cassman, Virmani Sheehy, Khush. Yield potential trends of tropical rice since the release of IR8 and the challenge of increasing rice yield potential. 1999;
55. Venkateswarlu Visperas. Solar radiation and rice productivity. *IRRI Research Paper Series (Philippines)*. 1987;
56. Khush. Green revolution: the way forward. *Nature Reviews Genetics*. 2001; 2: 815–822. <https://doi.org/10.1038/35093585> PMID: 11584298
57. Khush. Breaking the yield frontier of rice. *GeoJournal*. 1995; 35: 329–332.
58. Das Sengupta, Parida Roy, Ghosh Prasad, et al. Genetic diversity and population structure of rice landraces from Eastern and North Eastern States of India. *BMC Genet*. 2013; 14: 71. <https://doi.org/10.1186/1471-2156-14-71> PMID: 23945062
59. Anandan Anumalla, Pradhan Ali. Population structure, diversity and trait association analysis in rice (*Oryza sativa* L.) germplasm for early seedling vigor (ESV) using trait linked SSR markers. *PloS one*. 2016; 11: e0152406. <https://doi.org/10.1371/journal.pone.0152406> PMID: 27031620
60. Behera Mohanty, Pradhan Singh, Singh Sahu, et al. Assessment of genetic diversity of rainfed lowland rice genotypes using microsatellite markers. *Indian J Genet*. 2013; 73: 142–152.
61. Blair Hedetale, McCouch. Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). *TAG Theoretical and Applied Genetics*. 2002; 105: 449–457. <https://doi.org/10.1007/s00122-002-0921-5> PMID: 12582550
62. Garland Lewin, Abedinia Henry, Blakeney. The use of microsatellite polymorphisms for the identification of Australian breeding lines of rice (*Oryza sativa* L.). *Euphytica*. 1999; 108: 53–63.
63. Iqbal Rahman. Identification of marker-trait associations for lint traits in cotton. *Frontiers in plant science*. 2017; 8: 86. <https://doi.org/10.3389/fpls.2017.00086> PMID: 28220132
64. Cui Xu, Tang Yang, Yu Xin-xiang, et al. Genetic structure and association mapping of cold tolerance in improved japonica rice germplasm at the booting stage. *Euphytica*. 2013; 193: 369–382.
65. Borba Brondani, Bressegello Coelho, Mendonça Rangel, et al. Association mapping for yield and grain quality traits in rice (*Oryza sativa* L.). *Genetics and Molecular Biology*. 2010; 33: 515–524. <https://doi.org/10.1590/S1415-47572010005000065> PMID: 21637426
66. Nachimuthu Muthurajan, Duraialaguraja Sivakami, Pandian Ponniah, et al. Analysis of population structure and genetic diversity in rice germplasm using SSR markers: an initiative towards association mapping of agronomic traits in *Oryza Sativa*. *Rice*. 2015; 8: 30. <https://doi.org/10.1186/s12284-015-0062-5> PMID: 26407693

67. Edzesi Dang, Liang Liu, Zaid Hong. Genetic diversity and elite allele mining for grain traits in rice (*Oryza sativa* L.) by association mapping. *Frontiers in plant science*. 2016; 7.
68. Pradhan Barik, Sahoo Mohapatra, Nayak Mahender, et al. Population structure, genetic diversity and molecular marker-trait association analysis for high temperature stress tolerance in rice. *PloS one*. 2016; 11: e0160027. <https://doi.org/10.1371/journal.pone.0160027> PMID: 27494320
69. Roy Banerjee, Mawkhlieng Misra, Pattanayak Harish, et al. Genetic diversity and population structure in aromatic and quality rice (*Oryza sativa* L.) landraces from North-Eastern India. *PloS one*. 2015; 10: e0129607. <https://doi.org/10.1371/journal.pone.0129607> PMID: 26067999
70. Wang Qiang, Zhao Xu, Zhang Gao, et al. Association Mapping for Fiber-Related Traits and Digestibility in Alfalfa (*Medicago sativa*). *Frontiers in plant science*. 2016; 7.
71. Abdurakhmonov Kohel, Yu Pepper, Abdullaev Kushanov, et al. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics*. 2008; 92: 478–87. <https://doi.org/10.1016/j.ygeno.2008.07.013> PMID: 18801424
72. Zhang Zhong, Shahid Tong. Association analysis in rice: From application to utilization. *Frontiers in plant science*. 2016; 7.
73. Huang Wei, Sang Zhao, Feng Zhao, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*. 2010; 42: 961–7. <https://doi.org/10.1038/ng.695> PMID: 20972439
74. Garris McCOUCH, Kresovich. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). *Genetics*. 2003; 165: 759–769. PMID: 14573486
75. Gawenda Schröder-Lorenz, Debener. Markers for ornamental traits in Phalaenopsis orchids: population structure, linkage disequilibrium and association mapping. *Molecular breeding*. 2012; 30: 305–316.
76. Pandit Tasleem, Barik Mohanty, Nayak Mohanty, et al. Genome-wide association mapping reveals multiple QTLs governing tolerance response for seedling stage chilling stress in indica rice. *Frontiers in Plant Science*. 2017; 8: 552. <https://doi.org/10.3389/fpls.2017.00552> PMID: 28487705
77. Xu Bao, He Park. Genome-wide association study of eating and cooking qualities in different subpopulations of rice (*Oryza sativa* L.). *BMC genomics*. 2016; 17: 663. <https://doi.org/10.1186/s12864-016-3000-z> PMID: 27544770
78. Flint-Garcia Thuillet, Yu Pressoir, Romero Mitchell, et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J*. 2005; 44: 1054–64. <https://doi.org/10.1111/j.1365-3113X.2005.02591.x> PMID: 16359397
79. Babu Dinesh, Agrawal Sood, Chandrashekara Bhatt, et al. Comparative genomics and association mapping approaches for blast resistant genes in finger millet using SSRs. *PloS one*. 2014; 9: e99182. <https://doi.org/10.1371/journal.pone.0099182> PMID: 24915067
80. Boyles Cooper, Myers Brenton, Rauh Morris, et al. Genome-wide association studies of grain yield components in diverse sorghum germplasm. *The plant genome*. 2016; 9.
81. Liu Xiu, Chen Jie, others. Effects of low light on agronomic and physiological characteristics of rice including grain yield and quality. *Rice Science*. 2014; 21: 243–251.
82. Su Fan, Li Wei, Wang Wang, et al. Detection of favorable QTL alleles and candidate genes for lint percentage by GWAS in Chinese upland cotton. *Frontiers in plant science*. 2016; 7.
83. Sun Kang, Zhang Tan, Yu He, et al. Genetic diversity and population structure of rice pathogen *Ustilago noidea virens* in China. *PLoS One*. 2013; 8: e76879. <https://doi.org/10.1371/journal.pone.0076879> PMID: 24098811
84. Schulz Schott, Voorrips Smulders, Linde Debener. Genome-wide association analysis of the anthocyanin and carotenoid contents of rose petals. *Frontiers in Plant Science*. 2016; 7.
85. Wei Chen, Ma Zhang, Han Mei, et al. Comparative analysis of expression profiles of panicle development among tolerant and sensitive rice in response to drought stress. *Frontiers in plant science*. 2017; 8.
86. Wen Tsai. Haplotype association analysis of combining unrelated case-control and triads with consideration of population stratification. *Front Genet*. 2014; 5: 103. <https://doi.org/10.3389/fgene.2014.00103> PMID: 24860592
87. Yue Rong-rong, Ze-chuan Li-yong, Xing-hua Shi-hua. Quantitative trait locus analysis for rice yield traits under two nitrogen levels. *Rice Science*. 2015; 22: 108–115.
88. Alexandratos, Bruinsma, others. World agriculture towards 2030/2050: the 2012 revision. 2012.
89. Fu Zhang, Tan Zhu, Ma Fu, et al. Analysis of QTLs for yield-related traits in Yuanjiang common wild rice (*Oryza rufipogon* Griff.). *J Genet Genomics*. 2010; 37: 147–57. [https://doi.org/10.1016/S1673-8527\(09\)60033-3](https://doi.org/10.1016/S1673-8527(09)60033-3) PMID: 20227048

90. Vikram Swamy, Dixit Ahmed, Cruz Singh, et al. qDTY 1.1, a major QTL for rice grain yield under reproductive-stage drought stress with a consistent effect in multiple elite genetic backgrounds. *BMC genetics*. 2011; 12: 89. <https://doi.org/10.1186/1471-2156-12-89> PMID: 22008150