Genome Wide Association Mapping of Yield And Various Desirable Agronomic Traits In Rice

Muhammad Ashfaq (✉ ashfaq.iags@pu.edu.pk)

University of the Punjab

Abdul Rasheed

University of the Punjab

Muhammad Sajjad

COMSAT University Islamabad

Muhammad Ali

University of the Punjab

Bilal Rasool

Government College University Faisalabad

Muhammad Arshad Javed

University of the Punjab

Shabnum Shaheen

Lahore College for Women University

Alia Anwar

University of the Punjab

Sami Ul Allah

Bahuddin Zakaria University Bahudar Campus Layyah

Muhammad Shafiq Ahmad

University of the Punjab

Urooj Mubashar

Government Training Education Academy

Research Article

Keywords: Rice, GWAS, Germplasm, PIC, QTLs, SSR, Agronomic traits

Posted Date: February 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1280628/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Rice (*Oryza sativa* L.) is one of the staple foods worldwide. To feed the growing population, improvement of rice cultivar is important. To make the improvement in rice breeding program, it is imperative to understand the similarity of the existing rice accessions. Previous studies demonstrated the existence of abundant elite genes in rice landraces. A genome-wide association study (GWAS) was performed for yield and yield related traits to find the elite genes.

Design: Experimental study

Methods and results: A total of 204 SSRs markers were used among 17 SSRs were found to be located on each chromosome in the rice genome. The diversity was analyzed using different genetic characters i.e., the total number of alleles (TNA), polymorphic information content (PIC) and gene diversity by Pwer markers and the values for each genetic character per marker ranged from 2 to 9, 0.332 to 0.887 and 0.423 to 0.900 respectively across the whole genome. The results of population structure identified four main groups. MTA identified several markers associated with many agronomically important traits. These results will be very useful for the selection of potential parents, recombinants and MTAs that governing the improvements and developments of new high yielding rice varieties.

Conclusion: Analysis of diversity in germplasm is important for improvement of cultivars in breeding program. In the present study, the diversity was analyzed with different methods and found that enormous diversity was present in studied rice germplasm. The structure analysis found the presence of 4 genetic groups in the existing germplasm. A total of 129 marker-trait associations (MTAs) have been found in this study.

Introduction

Rice is one of the most important cereal crops and staple foods for almost half of the world's population [1]. Finding the genetic basis of natural diversity in an important agronomic trait in rice landraces and it is crucial for ensuring world's food security [1]. In a crop population, natural diversity or genomic variation is the outcome of mutation, outcrossing, crossing over and unsystematic genetic drift [2, 3]. Conservation and characterization of germplasm is essential to screen the desired genetic material for crop improvement program [4]. Moreover, Crop improvement is highly dependent on the availability of genetic variation in existing germplasm and identification of genetic variation for target traits [5]. Therefore, exploiting the genetic diversity in rice is of utmost importance.

Rice breeders have used various kinds diversity panel like old cultivars, wild species, modern cultivars, advanced lines and commercially released varieties for improvement of cultivated rice varieties. As a results of both natural and artificial selection under diverse habitats, the cultivated rice varieties possess high levels of morphological, physiological and genetic diversity resulting in >120,000 distinct rice varieties to be developed till date [6, 7]. Therefore, it is necessary to asses this genetic diversity to make decision about its conservation and to select the genotypes which could be utilized as potential parents

in upcoming breeding program. This kind of information will be helpful in selecting the genetically diverse parents that can be used for the development of new hybrids rice varieties.

The microsatellites markers also known as Simple-sequence repeats (SSRs) are now been gaining importance for assessment of genetic diversity. The main key features of SSRs over other molecular markers are genetic co-dominance, high level of polymorphism, multi allelic, relative abundant, easily and automatically scored and more interesting wide spread across the genome. Over the past decades, these polymorphic SSR markers have been extensively used in assessing the genetic diversity analysis [8, 9]. Along with that SSR has also been utilized for species identification [10], and mapping genetic linkage and association mapping [11]. The present study will also focus on the assessment of genetic diversity and association mapping using polymorphic SSR markers.

Conventional method of characterization of genetic diversity is the use of bi-allelic mapping populations or QTL mapping [12–14]. However, with the advancements in whole genome sequencing technology, Genome wide association study (GWAS) has now become extremely popular to map natural genetic variation in diverse populations, not only in rice but in other crop species as well [15, 16]. GWAS has advantage over linkage mapping as it enables the plant breeder to exploits the genetic diversity by using modern genetic technologies [17]. GWAS also maximize the diversity of the alleles and identifies the elite genes in the genome [18]. Furthermore, many candidate genes controlling agronomically important traits have been identified in rice through GWAS [19–20]. Using the same GWAS approach the present study was conducted to identify the agronomically important gene/ QTLs associated with rice yield and yield-related agronomic traits using diverse collection of genotypes. Moreover, this study will also identify the traits specific genotype that would be utilized in rice breeding program for the development of super rice.

Materials And Methods

Plant material

A panel of 100 rice genotypes representing different ecotypes and geographical origins were used (Supplementary Table. 1). Seeds of the rice were collected from Rice Research Institute, Kala Shah Kaku, United States Department of Agriculture (USDA) the USA.

Field Evaluation

The chosen rice accessions were sown for the screening of rice lines on the basis of various morphological characters and genetic diversity study. Seeds of the rice lines were soaked and sown in puddled field conditions through the broadcasting method. After 25 days the nursery of each rice line was prepared for transplantation into the next field with three replications under a Randomized Complete Block Design (RCBD). Fertilizer applications were done during 25 - 40 days after the transplant of seedlings. NPK, the primary crop stimulants were applied with a ratio of 120:60:60 kg/ha. However, P and K were smeared basally after transplant; N was smeared in three splits. Crop protection measures were

practiced preventing pests and unwanted wild plants occasionally when necessary. The data was collected for two consecutive years from 2016-17 and 2017-18.

Phenotypic Data Collection

The traits measured in the current study includes Days to 50 percent flowering (DF), Days to maturity (DM), Plant Height (PH), Tillers/Plant (TP), Panicle length (PL), Number of grains/panicle (G/P), Unfilled grain/panicle (UG/P), Seed setting %age (SS), Thousand grain weight (TGW), Yield kg/plot (Y/P), Yield kg/ha (Y/H). Details of measurement procedures of each trait are given in supplementary file (Supplementary File 1)

Statistical Analysis Of The Phenotypic Data

Mean, range, and standard deviation were determined using statistical tools offered by Microsoft Excel. Analysis of variance was completed using the SPSS statistical package to check the genetic variance amongst the rice lines for all their characters studied.

Isolation Of Genomic Dna

Genomic DNA of each rice genotype was extracted from 2 g of freshly taken leaf tissues using the modified CetylTrimethyl Ammonium Bromide (CTAB) method [21]. The extracted genomic DNA samples were dissolved and stored in TE buffer (10 mM Tris-base, 1 mM EDTA, pH 8.0). The quality of DNA was checked by running on agarose gel electrophoresis and further the concentration and quality was measured by nanodrop (ND1000, Thermo Scientific, USA).

Genotyping By Ssr

Collection of 204 SSR markers was screened to check polymorphism between the 100 rice accessions. The SSR markers were chosen from previously available rice genetic linkage and genome sequence (physical) maps [22, 23] (<http://www.gramene.org>). The PCR was carried out in PCR tube with 20 µL reaction volume. The PCR components included 2 µL genomic DNA (template), 2 µL 10 X buffer containing 1.5 mg MgCl₂, 0.1 µL Taq polymerase, 0.4 µL 10 mM dNTPs, and 2 µL (forward and reverse) primer. The PCR reactions were carried out on ABI thermal cycler (ProFlex PCR) using the following cycler conditions: initial denaturing temperature at 94°C for 4 min, followed by 35 cycles of 94°C for 45 sec, specific temperature for each primer for 45 sec, 72°C for 60 sec with a final extension at 72°C for 10 min. Amplified PCR products were examined by using 2-3% metaphore gel electrophoresis) (Bio Whittaker Molecular Applications, Vallensbaek Strand, Denmark) and visualized under UV light.

Molecular Data Analysis

Polymorphic bands amplified from SSR markers were recorded in numerical format like, 1 for presence and 0 for absence. Each amplified band from respective marker is designated as an allele. This data was further utilized for assessing the genetic diversity in the form of allelic variation, total number of total alleles (TNA), unique alleles (UA), private alleles (PA), expected alleles (Ne), expected heterozygosity (He) and these are measured by Power-Marker Ver 3.25 [24]. 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 .

Population Structure Analysis

The genetic structure (Q) prediction was analyzed with the program STRUCTURE version 2.2 (Pritchard et al., 2000) to identify clusters of genetically similar lines. The process of STRUCTURE organizes the diversity panel into n cluster to form a structure matrix (Q-matrix). The putative number of K (groups) was estimated from $K=1$ to 10 using 100,000 burn-in iterations and MCMC (Markov chain Monte Carlo). Web-based analysis "Structure Harvester v0.6.93" was applied to obtain maximum value or peak of " K ." for validation to understand the STRUCTURE results which were based on ad-hoc techniques [25]. The K values ranged from 1-10 and 6 independent runs to attain the reliable effects. The dissimilarity coefficient was analyzed by un-weighted pair group method with arithmetic mean (UPGAMA) through DARWIN (dissimilarity analysis and representation for windows) Version 5.0.148 software (<http://darwin.cirad.fr/darwin>). The linkage disequilibrium (LD) in studied diversity panel was evaluated using squared Pearson's correlation coefficients (r^2) using the r^2 command in the software PLINK [26].

Association Mapping

The association mapping method was used to map QTLs for the yield and other agronomic traits examined in the two growing seasons for a total of 204 polymorphic SSR markers. The average phenotypic data of studied traits during both years were used in genome-wide association mapping. The mix linear model (MLM) approach was used for linear unbiased estimates. The marker trait association was measured by incorporating Q+K matrices using TASSEL version v2.0.1 [27]. The association analysis was performed at 50000 times per mutations for the correction of multiple testing [28]. Markers with the P value ≥ 0.05 were regarded as significant. The default run parameters of the convergence criterion set at 1.09×10^{-4} and the maximum number of iterations set at 200 were used.

Results

Phenotypic variation

The grain yield is one of the most important traits in rice crop improvement which is interlinked with other agronomic traits. Significant differences between the genotypes for yield and yield-related traits were observed in the analysis of variance (ANOVA). Significant differences were observed in DF in season-II and DM in both seasons, and the remaining traits were highly significant differences in both seasons among studied rice germplasm (Table 1). This showed that for the studied attributes, a high level of phenotypic variability was observed among all accessions in the field condition. Phenotypic data showed a large variance in the studied rice diversity panel, making it useful for recording genotypic variability in the population.

Genome wide allelic pattern

The genome wide allelic pattern showed enormous diversity like the number of average alleles per locus (N_a) is 6.015, number of different alleles with a frequency $\geq 5\%$ is 4.735, number of effective alleles (N_e) is 4.162, Shannon's information index (I) is 1.491, expected heterozygosity (H_e) is 0.723 and unbiased expected heterozygosity (uH_e) is 0.727. Private allele means that are particularly unique to a single population, and in the present study there was no private alleles found (Table 2)

Genetic diversity based on PIC, TNA and SRBP

Out of 204 polymorphic SSRs microsatellite markers, 17 SSRs were found to be located on each chromosome which showed that SSRs are covering the whole genome of rice to check the diversity in the rice genome. The TNA values per marker ranged from 2 to 9, PIC values ranged from 0.332 to 0.887 and gene diversity values from 0.423 to 0.900 across the whole genome (Supplementary Table 2). The highest PIC value was observed i.e. 0.823, against RM8048 marker at chromosome 1 with 9 TNA, having size range base pairs (SRBP) 190-280 with 0.846 gene diversity. The maximum PIC value i.e 0.802 was detected against RM207 marker found on chromosome 2 with the 7 TNA, having the SRBP of 190-220 and level of gene diversity is 0.829. The markers on chromosomes 3 (RM5813) and 4 (RM5506) depicted the similar results and showed the maximum PIC value of 0.812 with the 8 TNA, having the SRBP 140-260, with 0.838 gene diversity. The markers RM439 at chromosome 6 showed the maximum PIC value of 0.887 having the 9 TNA, with SRBP was 110-240 and gene diversity was 0.900. The markers RM548, RM7110, and RM8266 which situated at chromosomes 5, 7, 8, respectively depicted the similar behavior and showed the maximum PIC value was 0.823 with the 9 TNA, having the SRBP was 140-260, with the value of 0.846 for gene diversity. The highest PIC value was recorded as 0.831 with 8 TNA, having the SRBP was 130-250 and value 0.851 for gene diversity in both markers namely RM5652 and RM4069 located at chromosome 9 and 11 respectively. The markers namely, RM5944 and RM2851 found on chromosomes at 10 and 12 exhibited the highest PIC values were 0.842 and 0.866 having the SRBP was 140-220 and 170-280 with the values of TNA 8 and 9, and gene diversity values of 0.863 and 0.866 f in the current study. The lowest gene diversity and PIC values across the whole genome in this study found against the markers RM317 followed by RM6973, RM3668, and RM5444 (Supplementary Table 2).

Population Structure analysis

The web-based analysis using 'Structure Harvester v0.6.93' showed the peak value of 4 ($K=4$) indicating the presence of four subgroups among the studied rice accessions (Fig. 1a). The population structure predicted by 'STRUCTURE' Software also divided the germplasm into four groups which is depicted by four different colors (Fig. 1b). Population structure is a consequence of departures from random mating in the sampling population that result in some individuals being more closely related than others. Based on the different structure analysis, the current diversity panel of 100 rice genotypes was clustered into 4 groups (Fig. 1). Further assessment of each cluster or group revealed that genotypes G1–G10 and G27, G28 found to fall into the first group; whereas the G11–G26 and G29, G33 genotypes completely appeared in the second group the third group contained G34–G72 genotypes. The fourth group having that is G73–G100 genotypes. Distance between each cluster specifies the differences among 100 rice genotypes and all clusters genetically different from each other. The distribution of rice germplasm in the four groups was illustrated by UPGAMA DARWIN tree (Fig. 2).

Linkage disequilibrium

The linkage disequilibrium (LD) pattern for 100 rice accessions was generated using genotypic data (Fig. 3). Each cell in the figure represents a relationship between two markers with the color codes for the presence of significant linkage disequilibrium (LD). The LD values were plotted, above the diagonal displays r^2 values and below the diagonal displays P values with color codes such as >0.01 represented by white color, <0.01 shown in blue color, <0.001 represented by green color and values <0.0001 are in red color. The frequency of linkage disequilibrium is associated with value at $P<0.0001$, there is strong LD between two markers and is represented with red color. The green color shows the LD between two markers at $P<0.001$. Similarly, blue color is associated with the lowest LD between two markers with $P<0.01$. At $P>0.01$, the white color represents no linkage between any two markers. The r^2 indicates the percentage of total variation explained. The values of r^2 range from 0.00–1.00 represented by different colors. Similar findings were obtained from the earlier reported results in diverse rice genotypes [39].

Association mapping

Among the 100 rice genotypes, significant phenotypic and genotypic differences were observed for all characters in the two growing seasons. A total of 129 marker-trait associations (MTAs) have been found in this study at $P<0.05$, $P<0.01$, $P<0.001$, and $P<0.0001$ significant levels. Among them, 41 MTAs were observed at $P<0.05$ significant levels, while 57 MTAs at $P<0.01$ followed by 21 MTAs at $P<0.001$ and 10 MTAs at $P<0.0001$ significant level (Table 3).

In this study, the total phenotypic variance represented by r^2 ranged from 6.08% to 27.15% among 11 studied traits using 204 polymorphic SSR markers (supplementary Table 3). Trait-wise highest numbers of markers-trait associations (MTAs) were identified in TGW (15) and Y/H (15) followed by TP (14), PH (14), UG/P (13), SS (13), G/P (11) PL (11), DM (10), Y/P (8), and DF (5). Chromosome-wise multi-trait-loci were perceived on chromosome number 1. Pleiotropic locus is associated with and affects the expression of more than one phenotypic trait. In the present study, several pleiotropic loci were identified

(Supplementary Table 3) like pleiotropic locus (RM5951) on chromosomes 4 at 56.1cM were significantly linked with PH, TGW, TP, Y/H, and SS). Another pleiotropic locus (RM5348) on chromosome 10 at the position 15.7cM was associated with PL, TGW, TP, and Y/H. The studied traits like PL, TGW, TP and Y/H were controlled by a pleiotropic locus (RM233B) on chromosome 2 at 16.3cM. Similarly, the Markers (RM237) on chromosome 1 showed pleiotropic effects for PL, TGW, Y/H, and Y/P in the present study. The explanation of each trait with their associated MTA was given below:

Days to 50 percent flowering (DF)

In total, five SSR markers were found to be significantly associated with DF which are located on chromosomes 1, 2, 4, 10, and 12 (Supplementary Table 3). The markers namely RM6973, RM6745, RM207, RM5853, and RM5886 were significantly associated with this trait. The phenotypic variation explained (PV) by the DF associated markers was 27.15 % to 9.63% of the total phenotypic variation (PV) of this trait. The marker (RM6973) explained the maximum value of trait variability (27.15%) on chromosome 12 while the marker (RM5886) on chromosome 4 explained the minimum value (9.63%) of trait variability.

Days to maturity (DM)

In this analysis, a total of 10 SSR markers were strongly correlated with days to maturity (DM). Out of them, three markers from chromosomes 3 (RM7389, RM1004 and RM517), 2 from chromosome 12 (RM6973 and RM3455) and remaining from chromosomes 4, 2, 11, 5 and 1 for the markers RM5635, RM8, RM3625, RM437 and RM1387 respectively (Supplementary Table 3). For this trait, the total PV of these markers ranged from 6.53% to 23.23%. On chromosome 12, the marker (RM6973) had the highest PVE (23.23%) while the marker (RM3455) on the same chromosome explained the least proportion of 6.53 percent of the variability of the trait at the position.

Plant height (PH)

Plant height was highly associated with fourteen SSR markers. The SSR markers namely RM6973, RM8084, RM81A, RM27, RM8257, RM5603, RM543, RM583, RM375, RM526, RM3153, RM3455, RM5951, and RM3153 showed significant associated with this trait. These PH related markers explained 6.48% to 23.23% of the total phenotypic variation in this trait (Supplementary Table 3). The marker (RM6973) explained maximum phenotypic trait variability (23.23%) on chromosome 12 at while the marker RM3153 on chromosome 8 at an explained minimum value (6.48%). Markers Trait Association for PH were distributed across 9 chromosomes, including, 3 SSRs at chromosome 1, 2 SSRs at each chromosome 2, 8 and 12, while 1 SSR at each chromosome 3,4,5,6 and 7.

Number of tillers per plant (TP)

Total 14 significant SSR markers were strongly linked with the number of tillers per plant including, 3 markers located on each chromosome 10 and 2 while, 2 SSR on each chromosome 4 and 9 and the other on 1, 3, 5, and 7 chromosomes (Supplementary Table 3). These markers explained 8.02% to 21.02% of the

phenotypic variation for TP in this study. The marker (RM6987) explained a maximum variation of 21.02 percent on chromosome 10, while the marker (RM5951) explained a minimum variation (8.02 percent) on 4 for this trait. RM6987, RM233B, RM6628, RM569, RM8, RM3625, RM201, RM207, RM5348, RM5853, RM5886, RM1106, RM505, and RM595115 are the significant markers associated with the number of tillers per plant.

Panicle length (PL)

Panicle length was highly associated with 11 SSR markers were located on chromosome 1, 2, 4, and 5, while the other on chromosome 6, 10, and 11. The significantly associated markers are RM233B, RM211, RM6628, RM6666, RM375, RM5348, RM317, RM441, RM421, RM237, and RM348 with panicle length (Supplementary Table 3). All of these SSRs explained 6.28% to 17.02% of the phenotypic variation in this trait. The marker (RM233B) explained maximum variation (17.02%) on chromosome while the marker (RM348) on chromosome 4 explained minimum variation (6.28%) in panicle length.

Number of grains per panicle (G/P)

In total, 11 significant SSR markers were found to be associated with the G/P. The MTA for G/P was distributed across 8 chromosomes, including, 3 SSRs on chromosome 12 and 2 on chromosomes 6, while the other on the chromosomes 1, 3, 4, 7, 9 and 11. The markers RM1004, RM1036, RM2459, RM375, RM3795, RM1106, RM113, RM400, RM2851, RM3455 and RM3558 showed significant association. All these SSRs explained 6.28% to 15.26% of the phenotypic variation in this trait. The marker (RM1004) on chromosome 3 explained maximum phenotypic variation (15.26%) while the marker (RM3558) on 4 explained minimum variation (6.28%) for the G/P (Supplementary Table 3).

Number of unfilled grain per panicle (UG/P)

In association analysis, 13 SSR markers were found to be significantly associated with the UG/P. Overall, 13 significant SSRs distributed across 8 chromosomes, Out of them, 2 SSRs on each chromosome 2, 4, 5, 9, and 11 while other SSRs on 1, 7, and 8. The markers namely RM7033, RM7390, RM7396, RM264, RM116, RM4552, RM317, RM3412, RM3838, RM421, RM300, RM125, and RM279 were significantly associated with Number of unfilled grains per panicles as mentioned in table 4.9. The phenotypic variation explained (PV) by the UG/P associated markers was 8.01% to 21.02% of the total phenotypic variation (PV) of this trait. The marker (RM7033) explained the maximum value of trait variability (21.02%) on chromosome 9 while the marker (RM279) from chromosome 2 explained the minimum value (8.01%) of the (UG/P) trait variability (Supplementary Table 3).

Seed setting percentage (SS)

A total of 13 SSR markers were strongly associated with the SS. Out of them, 3 markers from chromosomes 11 (RM264, RM421, and RM125), 2 from each chromosome 4 (RM7240, RM279), 8 (RM8264, RM6251) and 9 (RM5708, RM116) while the remaining from chromosomes 2, 5, 7 and 10 (RM703, RM7390, RM4552 and RM317 respectively. Total PV by these markers ranged from 6.69 %to

21.02 % for this trait. The marker (RM7033) had maximum PVE (21.02 %) on chromosome 2 while the marker (RM125) on chromosome 12 explained the least proportion 6.69 % of the SS (Supplementary Table 3).

1000 grain weight (TGW)

The TGW was highly associated with 15 SSR markers which include RM6973, RM6987, RM233B, RM7396, RM6745, RM3838, RM1106, RM5348, RM341, RM441, RM300, RM149, RM201, RM5951, and RM237 SSR. Total phenotypic variance ranged from 6.08 %-23.23%. The marker (RM6973) explained maximum phenotypic trait variability (23.23%) on chromosome 12 while the marker RM237 on chromosome 1 explained minimum value (6.08%). Markers trait association for TGW were distributed across 8 chromosomes, including, 3 SSRs at each chromosome 10 and 11 while 2 SSRs at each chromosome 1, 2, 4, and 9 and remaining SSRs at chromosome 5 and 8 (Supplementary Table 3).

Yield per plot (Y/P)

A total of 8 SSR markers were strongly linked to Y/P which includes RM8257, RM6745, RM569, RM8, RM237, RM441, RM505, and RM35588. These markers are located on chromosome, 1, 2, 3, 4, 7 and 11 chromosomes (Supplementary Table 3). The phenotypic variance explained by these markers ranged from 6.28 % to 15.71 %. The maximum variation (15.71%) was found on chromosome 7 against the marker RM8257, while the minimum variation (6.28%) was found on chromosome 4 against the marker RM3558.

Yield per hectare (Y/H)

The 15 significant SSR markers were found to be associated with Y/H which includes RM6987, RM6628, RM548, RM5603, RM569, RM3625, RM5348, RM5853, RM113, RM237, RM207, RM5886, RM5951, RM3153, and RM233B. The marker were distributed across 8 chromosomes, including 3 SSRs on each chromosome 1 and 5, while 2 SSRs on each chromosome 2, 4 and 10, and the other on the chromosomes 3, 8 and 11 (Supplementary Table 3). All these SSRs explained 6.28% to 21.04% of the phenotypic variation in this trait. The marker (RM6987) on chromosome 10 explained maximum phenotypic variation (21.04%) while the marker (RM233B) on 2 explained minimum variation (6.28%) for yield per hectare in 100 rice genotypes.

Discussion

Assessments of genetic diversity have been carried out in recent decades using diverse approaches [29–31]. The SSR markers are type of co-dominance markers are commonly being used for assessment of genetic diversity and association analysis in many crop species including rice [32–35]. The markers with higher PIC values have a greater ability to identify the cultivars having more genetic diversity. A locus with a PIC > 0.5 is considered highly diverse [36]. In the current study, the lowest PIC values were observed against RM317 marker and highest PIC value was 0.887 against RM439 marker. The genetic diversity can

also be detected by total number of alleles (TNA). Marker detecting the lowest total number of alleles showed the lowest gene diversity than those detected maximum total number of alleles revealed the higher gene diversity [29, 37]. In this study, the highest number of alleles per markers was 9 which was detected on chromosomes 1, 5, and 6, while the lowest total number of alleles 2 at chromosomes 4, 2, and 1. The PIC value also exhibited a significant correlation with the TNA and gene diversity for microsatellites SSR markers (Supplementary Table 2). Shannon's Information Index mean value is 1.491 having the standard value 0.027. Similar results were also reported by many researchers to exhibits the genetic diversity [31, 38–40]. In this study, significant amount of rare alleles were also identified which indicates that these rare alleles contribute well to the overall genetic diversity of the population. Collectively, all the results from PIC value, TNA, Shannon's Information Index and presence of rare alleles showed that there was enormous diversity present in the current collection of rice germplasm, which in turn can be used in the future breeding programs for selection of diverse parents.

The analysis of population structure provides the information about genetic similarities and dissimilarities within a set of genotypes under study. The population structure analysis formed the cluster based on this genetic similarities and dissimilarities. So, clustering a large number of germplasms represents a simple and effective process for assessing the genetic variability of germplasm [41]. The analysis of population structure based on Bayesian clustering model is one of the most frequent method of population clustering [35, 42, 43]. In the present study the selected genotypes were grouped into 4 sub-populations, as viewed in STRUCTURE analysis (Figure 1b). But according to given origin and pedigree record, there were three types of the population but on a genetic basis, the population divided into four clusters, so the results are unexpected. Consistent with the STRUCTURE analysis results, the DARWIN tree (UPGMA cluster) analysis also distribute the diversity panel into 4 clusters (Figure 2). Distance between each cluster specifies the differences among 100 rice genotypes and all clusters are genetically different from each other. The more genetic distance among clusters showed genetic diversity of the genotypes. The presence of 4 groups in the germplasm under study showed that there might be some genetic diversity present which needs to be known and yet investigated.

Genome wide association study (GWAS) also known as association analysis, is a method that relies on non-random combination of alleles at two genetic loci to study the natural variation present among the population. Recently, it has been widely used for the identification of elite gene / QTLs related to various traits such as yield, quality, biotic and abiotic stress tolerance in cultivated crops [44–45]. Therefore, present research is analogous to previous studies which focused on the small population with limited marker [40, 46]. There are two models used in association study to find the marker-trait associations (MTAs), one is GLM and other one is MLM. The GLM identified the more false positive association while MLM is more accurate to find the MTAs. So, in the present study MLM approach was utilized to find MTA. In total, 129 MTAs have been found in this study at different ($P < 0.05$, $P < 0.01$, $P < 0.001$, and $P < 0.0001$) significant levels (Table 3). Among them 15 MTA were found for TGW, 15 for Y/H, 14 for each PH and TP, 13 for each UG/P and SS, 11 for each G/P and PL, 10 for DM, 8 for Y/P and 5 for DF. Similar studies have been conducted previously and several MTAs have been found associated with yield and yield related traits like marker, RM5749 present on chromosome 4 was associated with numbers of the fertile tiller and

TGW [47]. The RM8257, RM6745, and RM569 found to be linked to yield and yield-related traits [48, 49]. Similarly, RM207 and RM540 found to be associated with G/P in rice [50]. Moreover, some pleiotropic locus have also been identified such as pleiotropic locus RM5951 on chromosome 10 was found to be associated with PH, TGW, TP, Y/H, and SS. Similarly, the pleiotropic locus (RM233B) on chromosome 2 found to be associated with PL, TGW, TP and Y/H. these MTAs will be helpful in the marker assisted selection.

Conclusion

Diversity in germplasm is imperative in a breeding program for improvement of cultivars. In the present study, the diversity was analyzed with different analysis using SSR markers. The 17 markers out of total 204 SSRs, found to be located on each chromosome in the rice genome. Different genetic characters were studied i.e. TNA, values per marker ranged from 2 to 9, PIC values ranged from 0.332 to 0.887 and gene diversity values from 0.423 to 0.900 across the whole genome. At chromosome 1 with 9 total numbers of alleles having the size range 190-280 base pairs with 0.846 gene diversity, the maximum PIC value was observed at 0.823 in the RM8048 marker. The structure analysis found the presence of 4 genetic groups in the existing germplasm. This showed that enormous genetic diversity is present in current collection of rice germplasm. This genetic diversity will be helpful in selecting the best parent to be utilized in breeding program. A total of 129 marker-trait associations (MTAs) have been found in this study. Among them, 43 MTAs were observed at $P < 0.05$ significant levels, while 57 MTAs at $P < 0.01$ followed by 21 MTAs at $P < 0.001$, and 10 MTAs at $P < 0.0001$ significant level. These MTA will be helpful in marker assisted selction breeding.

Declarations

Funding:

The project was supported by University of the Punjab, Lahore Pakistan and Higher Education Commission, Government of Pakistan Islamabad.

Acknowledgements

The authors would like to very thankful to the College for Agriculture and Biotechnology, Zhejiang University China for supporting and helping to complete this research work timely. Many thanks and greetings to Higher Education Commission (HEC), Government of Pakistan financially supported to complete the project in early dates.

Ethical statement

This research work presented in this paper is new, original and for public benefit. No human/animal sampling or material is used. All plant material was publically available and property rights were

infringed. No potential risks are speculated from the outcome of the research and the marker trait associations were identified in the project is commonly studied in labs/field and have no potential risks.

Autors countributions

MA, AR and MA design the experiment. MA, AR, MS, BR and SS collect and analyze the data. AA and SS write the manuscripts and analyze the data. SA, MAS and MAJ revise the manuscripts.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement: All necessary data is provided in supplementary file.

References

1. Zhang P, Zhong K, Zhong Z, Tong H (2019) Genome-wide association study of important agronomic traits within a core collection of rice (*Oryza sativa* L.) BMC plant bio. 19:1–12. <https://doi.org/10.1186/s12870-019-1842-7>
2. Shah SM, Naveed SA, Arif M (2013) Genetic diversity in basmati and non-basmati rice varieties based on microsatellite markers. Pak J Bot 45:423–431. <http://www.pakbs.org/.../56.pdf>
3. Bhandari H, Bhanu A, Srivastava K, Singh M, Shreya HA (2017) Assessment of genetic diversity in crop plants-an overview. Adv plants agric Res 7:00255. DOI: 10.15406/apar.2017.07.00255
4. Chakravarthi BK, Naravaneni R (2006) SSR marker based DNA fingerprinting and diversity study in rice (*Oryza sativa*. L). Afr J Biotec5. <https://www.ajol.info/index.php/ajb/article/view/42772>
5. Bevan MW, Uauy C (2013) Genomics reveals new landscapes for crop improvement. Geno bio 14:1–11. <https://doi.org/10.1186/gb-2013-14-6-206>
6. Vaughan DA, Lu BR, Tomooka N (2008) The evolving story of rice evolution. Plant sci 174:394–408. <https://doi.org/10.1016/j.plantsci.2008.01.016>
7. Roy S, Marndi B, Mawkhlieng B, Banerjee A, Yadav R, Misra A, Bansal K (2016) Genetic diversity and structure in hill rice (*Oryza sativa* L.) landraces from the North-Eastern Himalayas of India. BMC genet 17:1–15. <https://doi.org/10.1186/s12863-016-0414-1>
8. Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FM (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC plant bio 13:1–17. <https://doi.org/10.1186/1471-2229-13-39>
9. Ren X, Jiang H, Yan Z, Chen Y, Zhou X, Huang L, Lei Y, Huang J, Yan L, Qi Y (2014) Genetic diversity and population structure of the major peanut (*Arachis hypogaea* L.) cultivars grown in China by SSR markers. PLoS ONE 9:e88091. <https://doi.org/10.1371/journal.pone.0088091>

10. Shirasawa K, Ishii K, Kim C, Ban T, Suzuki M, Ito T, Muranaka T, Kobayashi M, Nagata N, Isobe S (2013) Development of Capsicum EST–SSR markers for species identification and in silico mapping onto the tomato genome sequence. *Mol Breed* 31:101–110. <https://doi.org/10.1007/s11032-012-9774-z>
11. Lu J, Wang S, Zhao H, Liu J, Wang H (2012) Genetic linkage map of EST-SSR and SRAP markers in the endangered Chinese endemic herb *Dendrobium* (Orchidaceae). *Genet Mol Res* 11:4654–4667. <http://dx.doi.org/10.4238/2012.December.21.1>
12. Adriani DE, Lafarge T, Dardou A, Fabro A, Clément-Vidal A, Yahya S, Dingkuhn M, Luquet D (2016) The qTSN positive effect on panicle and flag leaf size of rice is associated with an early down-regulation of tillering. *FPLS* 6:1197. <https://doi.org/10.3389/fpls.2015.01197>
13. Xu J, Henry A, Sreenivasulu N (2020) Rice yield formation under high day and night temperatures—A prerequisite to ensure future food security. *Plant, cell env* 43: 1595-1608. <https://doi.org/10.1111/pce.13748>
14. Chen L, Wang Q, Tang M, Zhang X, Pan Y, Yang X, Gao G, Lv R, Tao W, Jiang L (2021) QTL mapping and identification of candidate genes for heat tolerance at the flowering stage in rice. *Front Genet* 1840. <https://doi.org/10.3389/fgene.2020.621871>
15. Huang X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat genet* 42:961–967. <https://doi.org/10.1038/ng.695>
16. Kadam NN, Tamilselvan A, Lawas LM, Quinones C, Bahuguna RN, Thomson MJ, Dingkuhn M, Muthurajan R, Struik PC, Yin X (2017) Genetic control of plasticity in root morphology and anatomy of rice in response to water deficit. *Plant phys* 174:2302–2315. <https://doi.org/10.1104/pp.17.00500>
17. Kraakman AT, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* 168:435–446. <https://doi.org/10.1534/genetics.104.026831>
18. Zhao L, Lei J, Huang Y, Zhu S, Chen H, Huang R, Peng Z, Tu Q, Shen X, Yan S (2016) Mapping quantitative trait loci for heat tolerance at anthesis in rice using chromosomal segment substitution lines. *Breed sci* 15084. <https://doi.org/10.1270/jsbbs.15084>
19. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat genet* 48:927–934. <https://doi.org/10.1038/ng.3596>
20. Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, Li W, Zhan Q, Cheng B, Xia J (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat com* 6:1–9. <https://doi.org/10.1038/ncomms7258>
21. Murray M, Thompson WF, Rapid isolation of high molecular weight plant DNA (1980) *Nucl acids res* 8:4321–4326. <https://doi.org/10.1093/nar/8.19.4321>
22. Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation,

- transposon associations, and genetic marker potential. *Geno res* 11:1441–1452.
<http://www.genome.org/cgi/doi/10.1101/gr.184001>
23. McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L. *DNA res* 9:199–207.
<https://doi.org/10.1093/dnares/9.6.199>
 24. Liu J, Muse S, PowerMarker (2017) V3. 0 Manual. PowerMarker V3. 0 Manual
<http://www.powermarker.net>
 25. Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Cons genet reso* 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
 26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, PLINK (2007) a tool set for whole-genome association and population-based linkage analyses. *Ameri J hum genet* 81:559–575. <https://doi.org/10.1086/519795>
 27. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinfo* 23:2633–2635.
<https://doi.org/10.1093/bioinformatics/btm308>
 28. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genet* 155:945–959. <https://doi.org/10.1093/genetics/155.2.945>
 29. Rana MM, Islam MA, Imran S, Rubani S, Hassan L (2018) Genetic diversity analysis of NERICA lines and parents using SSR markers. *Int J plant & soil sci* 1–10.
<https://journalbank.org/index.php/IJPSS/article/view/607>
 30. Ahmed S, Anik TR, Islam A, Uddin I, Haque MS (2019) Screening of some rice (*Oryza sativa* L.) genotypes for salinity tolerance using morphological and molecular markers. *Biosci biotec res asia* 16:377–390. <http://dx.doi.org/10.13005/bbra/2753>
 31. Verma H, Borah J, Sarma R (2019) Variability assessment for root and drought tolerance traits and genetic diversity analysis of rice germplasm using SSR markers. *Sci Rep* 9:1–19.
<https://doi.org/10.1038/s41598-019-52884-1>
 32. Dirlewanger E, Cosson P, Tavaud M, Aranzana M, Poizat C, Zanetto A, Arus P, Laigret F (2002) Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theo Appli Genet* 105:127–138.
<https://doi.org/10.1007/s00122-002-0867-7>
 33. Hasnaoui N, Buonamici A, Sebastiani F, Mars M, Zhang D, Vendramin GG (2012) Molecular genetic diversity of *Punica granatum* L.(pomegranate) as revealed by microsatellite DNA markers (SSR). *Gene* 493:105–112. <https://doi.org/10.1016/j.gene.2011.11.012>
 34. Shiferaw E, Pè M, Porceddu E, Ponnaiah M (2012) Exploring the genetic diversity of Ethiopian grass pea (*Lathyrus sativus* L.) using EST-SSR markers. *Mol Bree* 30:789–797.
<https://doi.org/10.1007/s11032-011-9662-y>

35. Nachimuthu VV, Muthurajan R, Duraialaguraja S, Sivakami R, Pandian BA, Ponniah G, Gunasekaran K, Swaminathan M, Sabariappan R (2015) 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* 8:30. doi:10.1186/s12284-015-0062-5
36. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer J of hum genet* 32:314. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1686077/>
37. Babu BK, Meena V, Agarwal V, Agrawal P (2014) Population structure and genetic diversity analysis of Indian and exotic rice (*Oryza sativa* L.) accessions using SSR markers. *Mol bio rep* 41:4329–4339. <https://doi.org/10.1007/s11033-014-3304-5>
38. Shahriar M, Robin A, Begum S, Hoque A (2014) Diversity analysis of some selected rice genotypes through SSR-based molecular markers. *J Bangla Agric uni* 12:307–311. <https://www.banglajol.info/index.php/JBAU/article/view/28689>
39. Roy PS, Rao GJN, Jena S, Samal R, Patnaik A, Patnaik SSC, Jambhulkar NN, Sharma S, Mohapatra T (2016) Nuclear and chloroplast DNA variation provides insights into population structure and multiple origin of native aromatic rices of Odisha, India. *PLoS ONE* 11:e0162268. <https://doi.org/10.1371/journal.pone.0162268>
40. Anandan A, Anumalla M, Pradhan SK, Ali J (2016) 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* 11:e0152406. <https://doi.org/10.1371/journal.pone.0152406>
41. Belamkar V, Selvaraj MG, Ayers JL, Payton PR, Puppala N, Burow MD (2011) A first insight into population structure and linkage disequilibrium in the US peanut minicore collection. *Genet* 139:411. DOI 10.1007/s10709-011-9556-2
42. Zhao WG, Chung JW, Kwon SW, Lee JH, Ma KH, Park YJ (2013) Association analysis of physicochemical traits on eating quality in rice (*Oryza sativa* L.). *Euphytica* 191:9–21. <https://doi.org/10.1007/s10681-012-0820-z>
43. Edzesi WM, Dang X, Liang L, Liu E, Zaid IU, Hong D (2016) Genetic diversity and elite allele mining for grain traits in rice (*Oryza sativa* L.) by association mapping. *FPLS* 7:787. <https://doi.org/10.3389/fpls.2016.00787>
44. Donde R, Kumar J, Gouda G, Gupta MK, Mukherjee M, Baksh SY, Mahadani P, Sahoo KK, Behera L, Dash SK (2019) Assessment of genetic diversity of drought tolerant and susceptible rice genotypes using microsatellite markers. *Rice Sci* 26:239–247. <https://doi.org/10.1016/j.rsci.2019.01.004>
45. Zhang P, Zhong K, Shahid MQ, Tong H (2016) Association analysis in rice: from application to utilization. *FPLS* 7:1202. <https://doi.org/10.3389/fpls.2016.01202>
46. Pandit E, Tasleem S, Barik SR, Mohanty DP, Nayak DK, Mohanty SP, Das S, Pradhan SK (2017) Genome-wide association mapping reveals multiple QTLs governing tolerance response for seedling stage chilling stress in indica rice. *FPLS* 8:552. <https://doi.org/10.3389/fpls.2017.00552>

47. Suji K, Biji K, Poornima R, Prince KSJ, Amudha K, Kavitha S, Mankar S, Babu RC (2012) Mapping QTLs for plant phenology and production traits using indica rice (*Oryza sativa* L.) lines adapted to rainfed environment. *Mol biotec* 52:151–160. <https://doi.org/10.1007/s12033-011-9482-7>
48. Horii H, Nemoto K, Miyamoto N, Harada J (2006) Quantitative trait loci for adventitious and lateral roots in rice. *Plant Bree* 125: 198-200. <https://doi.org/10.1111/j.1439-0523.2006.01124.x>
49. Qu Y, Mu P, Zhang H, Chen CY, Gao Y, Tian Y, Wen F, Li Z (2008) Mapping QTLs of root morphological traits at different growth stages in rice. *Genetica* 133:187–200. DOI 10.1007/s10709-007-9199-5
50. Singh N, Choudhury DR, Tiwari G, Singh AK, Kumar S, Srinivasan K, Tyagi R, Sharma A, Singh N, Singh R (2016) Genetic diversity trend in Indian rice varieties: an analysis using SSR markers. *BMC genet* 17:1–13. DOI 10.1186/s12863-016-0437-7

Tables

Table 1 is available in the Supplemental Files section.

Table 2: Allelic pattern for 204 SSR markers in 100 rice genotypes

Name	Description	Mean	SD
Na	Na = No. of Average Alleles	6.015	0.146
Na Freq. $\geq$ 5%	Na (Freq $\geq$ 5%) = No. of Different Alleles with a Frequency $\geq$ 5%	4.735	0.105
Ne	Ne = No. of Effective Alleles = $1 / (\text{Sum } p_i^2)$	4.162	0.099
I	I = Shannon's Information Index = $-1 * \text{Sum } (p_i * \ln(p_i))$	1.491	0.027
No. P.A.	No of Private Alleles = No. of Alleles Unique to whole Population	0.00	0.00
He	He = Expected Heterozygosity = $1 - \text{Sum } p_i^2$	0.723	0.008
uHe	uHe = Unbiased Expected Heterozygosity = $(2N / (2N-1)) * He$	0.727	0.008

Table 3: Marker-trait associations at different significance level

Name	No. of Associations	($P<0.05$)	($P<0.01$)	($P<0.001$)	($P<0.0001$)
DF	5	1	2	1	1
DM	10	4	3	2	1
NG	11	4	5	2	0
PH	14	5	6	2	1
PL	11	4	6	1	0
SS	13	3	4	4	2
TG	15	6	5	2	2
TP	14	2	10	1	1
UG	13	5	5	2	1
YH	15	4	7	3	1
YP	8	3	4	1	0
Total	129	41	57	21	10

Figures

Figure 1

The population structure of rice diversity panel of 100 rice genotypes using 204 SSR marker data. (a) Peak value observed by 'Structure Harvester v 0.6.93' indicates the maximum value of $K=4$. (b) Population structure of rice diversity panel based on Bayesian method analyzed with 204 polymorphic SSRs detecting 4 groups. The four different colors represent the four different groups or subpopulations showing the similar genetic background.

Figure 2

The UPGAMA DARWIN tree displaying the distribution of rice genotypes in four groups, and presenting the genetic similarities and dissimilarities within and between groups.

Figure 3

Linkage disequilibrium (LD) patterns among 100 accessions genotyped with 204 SSR markers. The squared correlation coefficients (r^2) for each pair of markers are presented in the upper triangle and their corresponding tests in the lower triangle: As shown in the LD patterns display diagram, $P < 0.01$ for blue, $P < 0.001$ for green, $P < 0.0001$ for red. R^2 indicates the percentage of total variation explained. The values of r^2 range from 0.00-1.00 represented by different colors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.docx](#)
- [SupplementaryFilesMBR.doc](#)