

Article Type: Research

J Name: Modern Phytomorphology

Short name: MP

ISSN: ISSN 2226-3063/eISSN 2227-9555

Year: 2025

Volume: 19

Page numbers: 332-342

DOI: 10.5281/zenodo.17658913
(10.5281/zenodo.2025-19-332-342)



Short Title: Optimized genotyping-by-sequencing for high-resolution SNP discovery in rice (*Oryza sativa* L.)

RESEARCH ARTICLE

Optimized genotyping-by-sequencing for high-resolution SNP discovery in rice (*Oryza sativa* L.)

Ayesha Farooq¹, Muhammad Ali², Muhammad Asrar³, Muhammad Bilal Chatham⁴, Shahbaz Ahmad², Shabnum Shaheen¹, Disna Ratnasekera⁵, Zhiyong Xu⁶, Sammina Mahmood⁷, Qurban Ali^{8*}, Muhammad Ashfaq^{8*}, Farah Khan¹

¹Department of Botany, Lahore College for Women University, Lahore, Pakistan

²Department of Entomology, University of the Punjab, Lahore, Pakistan

³Department of Zoology, Faculty of Life Science, Government College University, Faisalabad, Pakistan

⁴Department of Agronomy, University of the Punjab, Lahore, Pakistan

⁵Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, Matara, Sri Lanka

⁶WuhanYansheng Agriculture Technology Ltd., Wuhan 430076, China

⁷Department of Botany, Division of Science and Technology, University of Education, Lahore, Pakistan

⁸Department of Plant Breeding and Genetics, University of the Punjab, Lahore, Pakistan

***Corresponding author:** Qurban Ali, Department of Plant Breeding and Genetics, University of the Punjab, Lahore, 54590, Pakistan, Email: saim1692@gmail.com

Muhammad Ashfaq, Department of Plant Breeding and Genetics, University of the Punjab, Lahore, 54590, Pakistan, Email: ashfaq.iags@pu.edu.pk

Received: 13.10.2025, Manuscript No. mp-25-171827 | **Editor Assigned:** 15.10.2025, Pre-QC No. mp-25-171827 (PQ) | **Reviewed:** 29.10.2025, QC No. mp-25-171827 (Q) | **Revised:** 05.11.2025, Manuscript No. mp-25-171827 (R) | **Accepted:** 14.11.2025 | Published: 21.11.2025

Abstract

Rice (*Oryza sativa* L.) is a staple crop vital to food security worldwide, particularly in South Asia. Enhancing its yield potential and disease resistance through genomic tools is a pressing necessity for sustainable agriculture. This study aimed to identify genome-wide Single Nucleotide Polymorphisms (SNPs) associated with key agronomic traits in Pakistani rice germplasm using Genotyping-By-Sequencing (GBS) and Genome-Wide Association Studies (GWAS). Fourteen diverse rice varieties were genotyped using ApeKI-based GBS, generating high-quality sequencing data. Following quality filtering and alignment to the MSU v6.0 Nipponbare reference genome, a total of 208 highly significant SNPs ($-\log_{10}P \geq 4$) were identified. GWAS was conducted using a mixed linear model incorporating population structure and kinship, while STRUCTURE analysis was employed to assess subpopulation stratification. Significant associations were identified for fifteen agronomic and yield-related traits, including days to flowering, plant height, grain weight, and disease resistance. Key pleiotropic loci such as OsGRb06041, OsGRg05186, and OsGRb22352 were linked with multiple traits. Population structure analysis revealed two main genetic clusters ($K=2$), reflecting the divergence between traditional basmati types and improved cultivars. This study delivers a robust genomic resource for Marker-Assisted Selection (MAS) and future functional studies in rice. It highlights specific genomic regions with potential utility in breeding programs aimed at enhancing productivity, resilience, and quality traits in Pakistani rice germplasm.

Keywords: Rice, GBS, GWAS, SNPs, MAS, Agronomic traits, Yield, Bacterial panicle blight, Genetic diversity

Abbreviations: GBS: Genotyping-By-Sequencing, GWAS: Genome-Wide Association Studies, SNPs: Single Nucleotide Polymorphisms, MAS: Marker-Assisted Selection: GDP: Gross Domestic Product

Introduction

Rice (*Oryza sativa* L.) feeds over half the global population and remains the cornerstone of food security, particularly in South and Southeast Asia. In Pakistan, rice not only constitutes a major dietary staple but also serves as a significant export commodity, contributing substantially to national GDP. However, climate change, emerging diseases, and resource limitations present increasing challenges to rice productivity and quality. The development of high-yielding, disease-resistant cultivars is therefore a strategic priority to ensure both local food security and global market competitiveness. With an increase in population there is an alarming gap between increasing demand and present production of rice causing a huge challenge for breeders. In order to meet the following challenge, we need to increase either area under production or to increase its yield or both. It appears difficult to increase area under production due to restraints like water logging, drought, salinity etc. Whereas, increasing the yield seems to be a potential option which is possible by using better management practices and enhanced varieties with better yield potential. Management practices are efficient only if a variety has a genetic potential for higher yield. Enhancement in genetic potential depends on the genetic diversity and genetic maps of the variety that carries information of gene, responsible for different agronomic characteristics (Reynolds and Borlaug, 2006).

Genomic variation analysis is an important part of plant genetics and improvement of crop programs. DNA polymorphisms can be linked to associated phenotype differences; they can also be genetically related to its causative factor and can also identify relationships between individuals of a population. Sudden emergence of next generation sequencing has caused technological revolution in agricultural genomics. These advanced technologies along with reducing sequence information generation limitations have also facilitated in classification of genomes and genes, and offers a much more comprehensive view of gene functions and diversity of plants. A new idea, named GBS, has come forward, in which the discovery of SNPs in large segregating population is associated with scoring, quick and direct study of plant diversity along with mapping of a trait or mutation of interest (Deschamps, et al. 2012).

Development in the technology of sequencing has significantly enhanced our ability to determine and genotype genetic markers. Some of the approaches may aim to lessen the genome complexity by sequencing the portion of targeted genome. Such GBS approach has been proved to be a cost efficient and effective tool in QTL detection, genome selection, population genetics and high-resolution mapping (Furuta, et al. 2017). Many crops have been studied using GBS to help in breeding programs, e.g. *Cicer arietinum* (Kujur, et al. 2015), *Brassica napus* (Bayer, et al. 2015, Bus, et al. 2012), *Zea mays* (Elshire, et al. 2011, Gore, et al. 2009), *Solanum tuberosum* (Uitdewilligen, et al. 2013), *Oryza sativa* (Huang, et al. 2009, Spindel, et al. 2015), *Sorghum bicolor* (Morris, et al. 2013) and *Triticum aestivum* (Poland, et al. 2012). Combined with phenotypic data, GBS offers a strong basis for quick mapping and identification of genes essential agronomic traits, which can then be introgressed into crop germplasm (Abe, et al. 2012, Edwards, et al. 2013).

Till date categorization of genetic basis of rice has been based on population linkage mapping developed from biparental crosses. While being useful in identifying allelic variation (between two phenotypically distinctive parents) it cannot be used for sampling of large pool genetic differences that may play a role in phenotypic variation within a species. Whereas, GWAS utilizing populations of distinct individuals can possibly reveal a greater and more descriptive set of loci that contribute to phenotypic variation. As revealed by different latest studies (Huang, et al. 2010, Zhao, et al. 2011, Yano, et al. 2016), Genome-Wide Association Studies in rice allows quick and efficient detection of significant loci for phenotypic related traits. Due to the accessibility of genome wide molecular markers, cheaper facilities for genotyping have made it feasible to explore natural populations to categorize considerable marker trait association and also GWAS. GWAS can be a practical method to map loci for many different traits and also to identify SNPs or genes related to the traits due to enhanced mapping resolution. It has also been emerged as the powerful tool to identify genes underlying complex diseases at an extraordinary rate (Swamy, et al. 2017, Zheng, et al. 2018).

In this study, we employed ApeKI-based GBS to genotype a diverse panel of Pakistani rice varieties and performed GWAS to identify SNPs associated with agronomically important traits. Our objectives were to: (i) construct a high-quality SNP dataset; (ii) identify loci significantly associated with phenotypic variation in yield-related and stress-resistance traits; and (iii) analyze the genetic structure of local rice germplasm. The findings of this research are expected to inform breeding strategies aimed at developing resilient, high-yielding rice cultivars for sustainable agriculture in Pakistan and beyond.

Materials and Methods

Phenotyping and trait measurement

Fifteen agronomic and yield-related traits were recorded for GWAS analysis, including: Days to 50% Flowering (DF), days to maturity (DM), Plant Height (PH), number of Effective Tillers (ET), number of Tillers per Plant (TP), Panicle Length (PL), Panicle Weight (PW), Flag Leaf Area (FLA), number of seeds per panicle (G/P), Weight of Seeds per Panicle (WtS/P), 1000 Grain Weight (TGW), Seed Length (SL), Seed Width (SW), Seed Length/Width ratio (S/LW), Yield (Y). In addition, disease severity for Bacterial Panicle Blight (BPB) was evaluated for resistance analysis.

DNA extraction

DNA was extracted from ten-day-old rice leaves using the CTAB method (Doyle and Doyle, 1987) with minor modifications, and quality/quantity was assessed spectrophotometrically. DNA integrity was confirmed by 0.8% agarose gel electrophoresis in 1X TAE buffer, with bands visualized under a UV trans-illuminator and documented using a gel documentation system.

Restriction enzyme digestion

In order to carryout Genotyping-by-Sequencing Elshire, et al. 2011 protocol was followed with some modifications:

Genomic DNA was digested using ApeKI (New England Biolabs), a methylation-sensitive restriction enzyme that generates 5' overhangs and targets low-copy regions, avoiding repetitive genomic sequences. The digestion reaction included 2 μL of DNA (100 ng/ μL), 2 μL of 10X NEB Buffer 3, 1 μL of ApeKI (4 U/ μL), and 15 μL of nuclease-free water, yielding a 20 μL reaction volume. The mix was gently vortexed, briefly centrifuged, and incubated at 75 °C for 2 hours.

Adapter ligation

Two adaptors were ligated to the digested fragments: A barcode adaptor with a 3' overhang of 4-8 bp and a 5' overhang complementary to ApeKI ends, and a common adaptor with an ApeKI-compatible sticky end. The ligation mix (18 μL) contained 8 μL of 5X T4 DNA ligation buffer, 1 μL of T4 DNA ligase, 7 μL of distilled water, and 2 μL each of barcode and common adaptors. This mixture was added to the digested DNA and incubated at 22°C for 1 hour, followed by enzyme inactivation at 65°C for 30 minutes and cooling to 4°C. Ligated products were stored at -20°C.

Library cleanup and double size selection

A two-step bead-based cleanup was employed using Polyethylene Glycol (PEG) and NaCl buffers. First, 60 μL of MilliQ water and 50 μL of Ampure Mix 1 were added to each sample and incubated at room temperature for 30 minutes. The supernatant was transferred to a new plate with 20 μL of Ampure Mix 2, incubated again, and washed twice with 80% ethanol. After brief air drying, 30 μL of MilliQ water was used to elute the DNA.

PCR amplification of ligated libraries

The PCR reaction (50 μL) comprised 10 μL of ligated DNA, 25 μL of 2X NEB Taq Master Mix, 2 μL of primer mix (12.5 pmol/ μL), and 13 μL of distilled water. The thermal profile included an initial step at 72°C for 5 minutes and 98°C for 30 seconds, followed by 18 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 30 seconds, ending with a final extension at 72°C for 5 minutes. The primers used were:

Forward: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

Reverse: 5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCTGCTGAACCGCTCTTCCGATCT

Library purification

PCR products were cleaned using 75 μL (1.5X) Ampure XP beads. The mix was incubated at room temperature for 15 minutes and placed on a magnet. Two ethanol washes (80%) were followed by air drying, and elution was performed in 30 μL of MilliQ water.

Sample pooling and size selection

Equal molar amounts (12 μL) from each sample were pooled. A 1.5% agarose gel (400 mL) was prepared using GelRed stain. Samples were loaded with bromide-free loading dye and run at 80 V for 2 hours. Target bands were visualized under UV light, excised, and purified with the Qiagen gel extraction kit. DNA was eluted in 30 μL elution buffer.

Final quantification and normalization

Library size distribution was analyzed using a LabChip system, and concentrations were measured using a Qubit fluorometer. Qubit working solution was prepared according to the manufacturer's instructions. DNA concentration (ng/ μL) was converted to molarity (nM) using the formula:

$$\text{Concentration (nM)} = (\text{conc. in ng}/\mu\text{L} / 660 \times \text{average library size in bp}) \times 10^6$$

Libraries were diluted to 10 nM with 10 mM Tris-HCl (pH 8.5) containing 0.1% tween-20 for sequencing.

Statistical analysis

Filtering of raw sequence data: High-quality raw sequencing data were obtained in qseq format and underwent rigorous filtering to ensure downstream analytical accuracy. Reads were retained only if they perfectly matched expected barcode sequences and

contained the expected ApeKI cut-site. Reads with adapter–adapter dimers or ambiguous nucleotides (“N”) in the first 72 bases were discarded. The filtered reads were demultiplexed, sorted by barcode, and trimmed to 64 bases. Adapter sequences were removed using `process_radtags`, and unpaired reads were excluded from the dataset.

GBS data processing and SNP calling: SNP discovery was conducted using the TASSEL 3.0 GBS pipeline. High-quality reads were aligned to the MSU Rice Genome Annotation Project version 6.0 (Nipponbare) reference genome using Bowtie 2 (Spindel and Wright, et al. 2013). SNPs with more than 10% missing data or monomorphic variants were excluded from analysis. Remaining SNP data were numerically encoded as follows: 1 for homozygotes (reference), 0 for heterozygotes, and -1 for homozygotes. Minor Allele Frequencies (MAF) was calculated to assess allele distribution (Spindel, et al. 2015). A mixed linear model incorporating population structure and cryptic kinship was employed to evaluate SNP–trait associations (Norton, et al. 2018).

Population structure analysis

To infer population stratification, STRUCTURE v2.3.1 (Garris, et al. 2005) was used. Genotypic data were analyzed under a Bayesian clustering model with a burn-in period of 30,000 and 30,000 MCMC iterations. The number of subpopulations (K) was tested from 1 to 10. The optimal K value was determined using the Evanno method in Structure Harvester based on changes in LnP(D) (Yu and Buckler, 2006).

Genome-Wide Association Analysis (GWAS)

GWAS was performed using the PIQUE pipeline for data preprocessing and the Efficient Mixed-Model Association eXpedited (EMMAX) software for SNP–phenotype association analysis (Kang, et al. 2010). A significance threshold of $P < 0.05$ ($-\log_{10} P \geq 3$) was used to identify meaningful associations. Manhattan and quantile–quantile (Q–Q) plots were generated to visualize SNP distributions and assess the model fit (Huang, et al. 2015).

Results

Genomic library sequencing and alignment metrics

The sequencing metrics confirmed robust library preparation and high-quality data generation across 14 rice genotypes. After preprocessing, an average of over 97% of reads was retained, and alignment rates to the reference genome ranged from 86.2% to 88.9%. The proportion of uniquely mapped reads exceeded 96% for all samples, indicating minimal alignment ambiguity and ensuring reliable downstream SNP calling (Tab. 1).

Sample	Number of raw reads (bp)	Number of bases (bp)	GC content (%)	Percentage of data (%)	Raw read length	Number of pre-processed reads (bp)	Percentage of pre-processed reads (%)	Number of reads aligned (bp)	Percentage of reads aligned (%)	Number of uniquely aligned reads (bp)	Percentage of uniquely aligned reads (%)
BAS 370	4 203 835	420 380 000	42.1	91.36	100 × 2	4 146 804	98.6	3 685 298	88.9	3 594 266	97.5
BAS Pak	3 219 412	321 940 000	41.7	92.51	100 × 2	3 149 872	97.8	2 713 875	86.2	2 610 014	96.2
BAS 198	3 158 027	315 800 000	43.5	92.73	100 × 2	3 096 924	98.1	2 695 973	87.1	2 596 873	96.3
BAS 385	2 861 670	286 160 000	41.8	91.66	100 × 2	2 774 894	97.0	2 417 346	87.1	2 330 122	96.4
Super Basmati	3 696 550	369 650 000	40.7	91.02	100 × 2	3 589 908	97.1	3 174 548	88.4	3 096 194	98.0
Basmati 2000	2 266 041	226 600 000	41.4	91.75	100 × 2	2 194 896	97.0	1 899 685	86.6	1 824 209	96.0
Basmati 515	2 513 749	251 370 000	43.8	92.04	100 × 2	2 443 927	97.2	2 142 183	87.7	2 062 408	96.3
PK 1121 aromatic	4 301 519	430 150 000	42.3	92.21	100 × 2	4 239 731	98.6	3 708 401	87.5	3 587 535	96.7
PK 386	4 511 260	451 120 000	42.7	91.53	100 × 2	4 456 808	98.8	3 916 462	87.9	3 830 112	97.8
Kisan Basmati	3 061 030	306 100 000	40.2	91.07	100 × 2	2 986 972	97.6	2 649 407	88.7	2 579 874	97.3
Chenab Basmati	4 101 513	410 150 000	40.6	92.33	100 × 2	3 996 816	97.4	3 519 904	88.1	3 407 691	97.0

Punjab Basmati	3 462 772	346 270 000	42.7	91.14	100 × 2	3 365 804	97.2	2 914 773	86.6	2 799 641	96.1
Basmati 2019	3 566 129	356 610 000	41.1	92.25	100 × 2	3 500 906	98.2	3 102 577	88.6	3 057 386	98.5
Super Gold	2 901 422	290 140 000	42.4	92.44	100 × 2	2 863 412	98.7	2 488 793	86.9	2 431 426	97.7

Table 1. Summary and alignment statistics of sequencing reads mapped of fourteen rice varieties with reference genome Nipponbare.

Genome-Wide Association Study (GWAS)

A total of 208 significant SNPs ($-\log_{10}P \geq 4$) were associated with 15 agronomic and yield-related traits using a mixed linear model. Manhattan and Q-Q plots (Figure 1a and 1b) demonstrated the chromosomal distribution and statistical significance of these associations. SNPs forming distinct peaks on the Manhattan plots were flagged as key candidate loci due to their strong phenotypic associations. Specifically, Days to 50% Flowering (DF) was associated with 51 SNPs across chromosomes 2, 3, 7, 8, and 11, with Phenotypic Variation Explained (PVE) ranging from 6% (OsGRb17256) to 28.3% (OsGRb15811). Days to Maturity (DM) showed association with 9 SNPs on chromosomes 3, 4, 6, 7, 8, 9, and 10 (PVE: 9.1-18.4%). The number of Seeds per Panicle (G/P) was influenced by 7 SNPs on chromosome 9 (PVE: 7.2-22.5%). Plant Height (PH) was associated with 19 SNPs across six chromosomes, with PVE values between 7.3% and 23.2%. Panicle Length (PL) had 22 SNPs with PVE values from 7.2% to 19.5%, and 1000 Grain Weight (TGW) showed associations with 10 SNPs located on chromosomes 1, 2, 5, 8, 9, and 11 (PVE: 6.2-17.3%). Tiller per Plant (T/P) was linked to 4 SNPs (PVE: 10.2-16.4%), while 46 SNPs influenced Yield (PVE: 12.9-24.7%). Seed Length (SL) and Seed Width (SW) were influenced by 5 and 6 SNPs, respectively, with SL showing PVE values from 14.5% to 15.2%, and SW from 14.8% to 19.8%. Flag Leaf Area (FLA) had 7 significant SNPs (PVE: 10-24.9%), Panicle Width (PW) had 5 (PVE: 14-22.7%), and Effective Tillers (ET) had 6 (PVE: 8.8-16.7%). For Weight of Seeds per Panicle, 7 SNPs were detected (PVE: 7.9-22.2%), while Bacterial Panicle Blight (BPB) was associated with 4 SNPs (PVE: 6.8-11.2%). Notably, several pleiotropic loci were discovered, including OsGRb06041 (Chr 2) linked with both DF and TGW, OsGRg05186 (Chr 3) associated with DM, PW, and T/P, OsGRb19642 (Chr 9) influencing FLA, G/P, and Yield, and OsGRb22352 (Chr 11) affecting DF, PL, Yield, and Weight of Seeds per Panicle. The average interval between SNPs was 6.0 kb, with 83.2% of the SNPs located within 10 kb of their nearest neighbor, indicating dense genome coverage suitable for high-resolution association mapping (Tabs. 2 and 3).

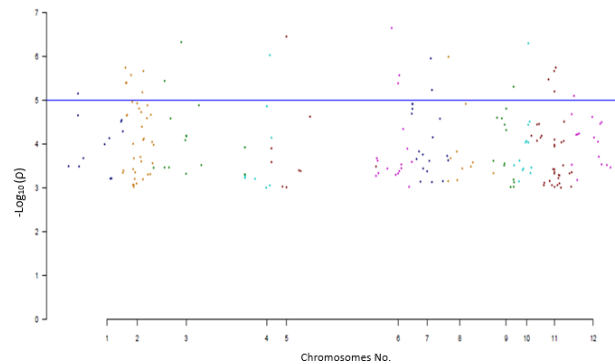


Figure 1a. Manhattan plots showing the location of significant SNPs and $\log_{10}(p)$ across 12 rice chromosomes associated with all traits under study. The blue horizontal line designates the threshold of significance.

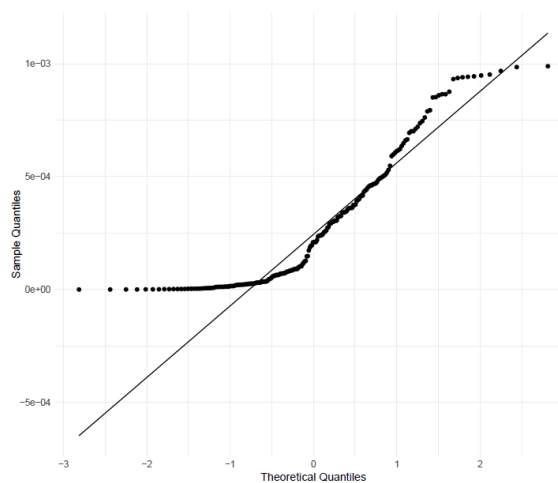


Figure 1b. Q-Q Plot for assessing normality of residuals with sample quantiles (Y-axis) vs. theoretical quantiles (X-axis) for all traits.

Trait	Total SNPs	Chromosome(s)	Significant SNPs	Position (cM)	Pleiotropic Loci
Days to 50% Flowering (DF)	51	2, 3, 7, 8, 11	OsGRb15811, OsGRb24059, OsGRg02612, OsGRb15223 and OsGRb24280	30180192-8928822	OsGRb22352 (Chr 11, 19075686cM)
Days to Maturity (DM)	09	3, 4, 6, 7, 8, 9, 10	OsGRg05186, OsGRg01577	65194661-2941637	OsGRg05186 (Chr 3, 65194661cM)
No. of Seeds/Panicle (G/P)	07	9	OsGRb33106, OsGRb09511	17085262-12926281	OsGRb19642 (Chr 9, 19116041cM)
Plant Height (PH)	19	1, 3, 4, 6, 7, 8	OsGRg06248, OsGRb24963	25546570-4852563	-
Panicle Length (PL)	22	1, 2, 3, 5, 6, 10, 11	OsGRb29201, OsGRb38929, OsGRg25714, OsGRb11525, OsGRb16808, OsGRb22951, OsGRg29683 and OsGRg07384	25437901-25732844	OsGRb22352 (Chr 11, 19075686cM)
1000 Grain Weight (TGW)	10	1, 2, 5, 8, 9, 11	OsGRb20539, OsGRg04122	10256143-2942452	OsGRb06041 (Chr 2, 16915416cM)
Tiller/Plant (T/P)	04	2, 3, 6, 11	OsGRg15277, OsGRb25369	28593541-8588445	OsGRg05186 (Chr 3, 65194661cM)
Yield	46	1, 3, 4, 7, 8, 9, 10, 11, 12	OsGRg01045, OsGRb26721, OsGRb32601, OsGRb24777, OsGRb06726, OsGRb10649	19116041-20179676	OsGRb22352 (Chr 11, 19075686cM), OsGRb19642 (Chr 9)
Seed Length (SL)	05	1, 8, 9, 11	OsGRb21802, OsGRb21889, OsGRb21859	12348419-10361467	-
Seed Width (SW)	06	5, 6, 8	OsGRb25807, OsGRb08496	2856240-4554966	-
Flag Leaf Area (FLA)	07	1, 4, 5, 9	OsGRg18870, OsGRb28544	33053796-13030751	OsGRb19642 (Chr 9, 19116041cM)
Panicle Width (PW)	05	1, 3, 6	OsGRg01909, OsGRb23923	39577604-10247220	OsGRg05186 (Chr 3, 65194661cM)
Effective Tillers (ET)	06	5, 6, 7, 8	OsGRg04446, OsGRb13097	6519407-18839766	-
Weight of Seeds/Panicle	07	5, 6, 7, 9, 11	OsGRb14645, OsGRb38178, OsGRb311652	25793671-30878662	-
Bacterial Panicle Blight (BPB)	04	1, 2, 3	OsGRg28835, OsGRg201457	6048532-37552902	-

Table 2. Summary of Significant SNPs and Pleiotropic Loci for various traits in Rice.

Trait	SNP (Lowest & Highest)	P-Value (Lowest & Highest)	SNP (Lowest & Highest)	PVE (%) (Lowest & Highest)	SNP (Lowest & Highest)	LOD (Lowest & Highest)
Days to 50% Flowering (DF)	OsGRg03819 (Chrm #2) OsGRb23391 (Chrm #11)	1.8×10^{-6} - 9.48×10^{-4}	OsGRb17256 (Chrm #8) OsGRb15811(Chrm #2)	06 -28.3	OsGRb23391 (Chrm #11) OsGRg03819 (Chrm #2)	3.02- 5.75
Days to Maturity (DM)	OsGRb14447 (Chrm #9) OsGRg01577 (Chrm #6)	2.59×10^{-5} - 5.3×10^{-4}	OsGRg05186 (Chrm #3)OsGRg01577 (Chrm #6)	9.1-18.4	OsGRg04475 (Chrm #6) OsGRb17439 (Chrm #4)	3.12- 4.86
No. of Seeds/Panicle (G/P)	OsGRg16274 (Chrm #9) OsGRb33106 (Chrm #9)	1.55×10^{-5} - 9.52×10^{-4}	OsGRb33106 (Chrm #9) OsGRb09511(Chrm #9)	7.2-22.5	OsGRb33106 (Chrm #9) OsGRg16274 (Chrm #9)	3.02- 4.81

Plant Height (PH)	OsGRg16147 (Chrm #3) OsGRg06248 (Chrm #6)	4.71×10^{-7} - 9.44×10^{-4}	OsGRg06248 (Chrm #6)OsGRb24963 (Chrm #7)	7.3-23.2	OsGRg06248 (Chrm #6) OsGRg16147 (Chrm #3)	3.03-6.33
Panicle Length (PL)	OsGRb20572 (Chrm #6) OsGRb22951 (Chrm #5)	2.24×10^{-7} - 9.32×10^{-4}	OsGRb29201 (Chrm #11) OsGRb38929(Chrm #11)	7.2-19.5	OsGRb34669 (Chrm #5) OsGRb20572 (Chrm #6)	2.09-6.11
1000 Grain Weight (TGW)	OsGRg04122 (Chrm #8) OsGRb17840(Chrm #11)	1.02×10^{-6} - 7.89×10^{-4}	OsGRb20539(Chrm #2) OsGRg04122(Chrm #8)	6.2-17.3	OsGRb03785 (Chrm #5) OsGRb18430 (Chrm #1)	2.40-4.65
Tiller/Plant (T/P)	OsGRb25369 (Chrm #11) OsGRg03852 (Chrm #3)	3.48×10^{-5} - 4.92×10^{-4}	OsGRg15277 (Chrm #2) OsGRb25369(Chrm #11)	10.2-16.4	OsGRg15277 (Chrm #2)+OsGRg03852 (Chrm #3) OsGRb25369(Chrm #11)	3.31-4.46
Yield	OsGRb21043(Chrm #12) OsGRg16932 (Chrm #4)	7.94×10^{-6} - 9.89×10^{-4}	OsGRg01045 (Chrm #9) OsGRb26721(Chrm #12)	12.9-24.7	OsGRg16932 (Chrm #4) OsGRb21043 (Chrm #12)	3.00-5.10
Seed Length (SL)	OsGRb21889+OsGRb21859 (Chrm #11) OsGRb21802 (Chrm #8)	6.44×10^{-5} - 3.6×10^{-4}	OsGRb21889+OsGRb21859(Chrm #11) OsGRb21987 (Chrm #9) OsGRb21889+OsGRb21859(Chrm #11) OsGRb21987 (Chrm #9)	14.5-15.2	OsGRb21802 (Chrm #8) OsGRb21889+OsGRb21859 (Chrm #11)	3.44-4.19
Seed Width (SW)	OsGRb25807 (Chrm #5) OsGRb08496 (Chrm #6)	1.25×10^{-4} - 4.61×10^{-4}	OsGRb08496 (Chrm #6) OsGRb25807 (Chrm #5)	14.8-19.8	OsGRb08496 (Chrm #6) OsGRb25807 (Chrm #5)	3.34-4.90
Flag Leaf Area (FLA)	OsGRb28544 (Chrm #5) OsGRg28569 (Chrm #5)	3.5×10^{-7} - 9.68×10^{-4}	OsGRg18870 (Chrm #1) OsGRb28544 (Chrm #5)	10-24.9	OsGRb01831 (Chrm #1) OsGRb28544 (Chrm #5)	23.7-6.46
Panicle Width (PW)	OsGRg01909 (Chrm #1) OsGRb23925 (Chrm #3)	3.06×10^{-5} - 5.04×10^{-4}	OsGRb23923 (Chrm #3)OsGRg01909 (Chrm #1)	14-22.7	OsGRb23925 (Chrm #3) OsGRg01909 (Chrm #1)	3.30-4.51
Effective Tillers (ET)	OsGRb13087 (Chrm #6) OsGRg03954 (Chrm #8)	2.69×10^{-6} - 4.61×10^{-4}	OsGRb13097 (Chrm #6) OsGRg04446 (Chrm #7)	8.8-16.7	OsGRg03954 (Chrm #8) OsGRb13087 (Chrm #6)	3.34-5.57
Weight of Seeds/ Panicle	OsGRb38929 (Chrm #11) OsGRb38178 (Chrm #11)	1.13×10^{-4} - 9.85×10^{-4}	OsGRb14645 (Chrm #7)+OsGRb38178 (Chrm #11)OsGRb311652 (Chrm #11)	7.9-22.2	OsGRb38178 (Chrm #11) OsGRb38929 (Chrm #11)	3.01-3.95
Bacterial Panicle Blight (BPB)	OsGRg25613 (Chrm #1) OsGRg28835 (Chrm #2)	2.08×10^{-4} - 3.14×10^{-4}	OsGRg28835 (Chrm #2)OsGRg201457 (Chrm #3)	6.8-11.2	OsGRg28835 (Chrm #2) OsGRg28214 (Chrm #3)	3.05-3.38

Table 3. Summary of SNPs, P-Values, PVE (%) and LOD values for key agronomic traits and disease resistance in rice.

Genome-wide SNP associations with agronomic and resistance traits in rice

The study revealed several SNPs significantly associated ($FDR < 0.1$) with phenological, morphological, yield-related, grain quality, and Bacterial Panicle Blight (BPB) resistance traits. The identified loci were distributed across 11 rice chromosomes, with Minor Allele Frequencies (MAF) ranging from 0.013 to 0.55 and adjusted FDR values ranging from 0.014 to 0.09.

Phenological traits: For Days to 50% Flowering (DF), significant associations were identified on chromosome 2. The SNP OsGRg01715 (T/C; MAF=0.03; FDR=0.039) represented the lowest frequency associated variant, whereas OsGRb05532 (T/C; MAF=0.49; FDR=0.02) marked a common variant. These SNPs may influence flowering-time regulatory genes such as Hd1, Ehd1, or Ghd7, which respond to photoperiod and environmental cues. Similarly, Days to Maturity (DM) was associated with OsGRg05186 (chr 3; A/G; MAF=0.03; FDR=0.03) and OsGRg04475 (chr 6; A/G; MAF=0.36; FDR=0.02), implicating genes possibly related to hormonal pathways or developmental timing regulation.

Plant architecture and yield components: Significant SNPs for Plant Height (PH) were located on chromosomes 6 and 7. OsGRg06248 (A/C; MAF=0.021; FDR=0.023) may correspond to loci affecting gibberellin biosynthesis, while OsGRb24963 (A/G; MAF=0.363; FDR=0.052) may influence stem elongation. Panicle Length (PL) was associated with OsGRg04319 (chr 1; T/C; MAF=0.021; FDR=0.031) and OsGRb22951 (chr 5; T/C; MAF=0.44; FDR=0.037). Panicle Width (PW) exhibited associations with OsGRb23923 (chr 3; A/G; MAF=0.02; FDR=0.022) and OsGRb24065 (chr 1; A/G; MAF=0.33; FDR=0.07), highlighting their relevance in panicle morphology and grain capacity. Number of Seeds per Panicle (G/P) was significantly associated with OsGRg12638 (chr 9; A/C; MAF=0.02; FDR=0.02)

and OsGRg15175 (A/C; MAF=0.04; FDR=0.03), potentially influencing spikelet fertility and reproductive success. Tiller per Plant (T/P) and Effective Tillers (ET), both key yield-contributing traits, showed associations with SNPs on chromosomes 2, 3, 6, and 8. Notably, OsGRg15277 (chr2; T/C; MAF=0.02; FDR=0.033) and OsGRg03954 (chr8; A/G; MAF=0.013; FDR=0.18) represent rare alleles, which may contribute to tillering capacity under specific environmental conditions.

Grain quality and yield-related traits: Significant SNPs for 1000 Grain Weight (TGW) were observed on chromosomes 2 (OsGRb24166; A/G; MAF=0.02; FDR=0.017) and 11 (OsGRb17840; A/G; MAF=0.36; FDR=0.019), suggesting these loci are linked to grain filling and size regulation mechanisms. Grain yield per plant was associated with OsGRg01045 (chr9; T/A; MAF=0.02; FDR=0.06) and OsGRb02133 (chr1; A/G; MAF=0.53; FDR=0.014), with the latter being a high-frequency allele, suggesting its stability and potential for breeding value. Seed Length (SL) and Seed Width (SW) were linked to SNPs on chromosomes 11 and 6, respectively. OsGRb21889 (SL; A/G; MAF=0.02; FDR=0.04) and OsGRb08457 (SW; A/C; MAF=0.016; FDR=0.02) may correspond to grain morphology genes controlling endosperm development. The Weight of Seeds per Panicle trait was associated with OsGRb09045 (chr5; T/C; MAF=0.08; FDR=0.02) and OsGRb38178 (chr11; A/C; MAF=0.55; FDR=0.09), indicating both rare and common allelic contributions to sink strength.

Disease resistance: Bacterial Panicle Blight (BPB): Two SNPs were significantly associated with Bacterial Panicle Blight (BPB) severity. OsGRg201457 (chr3; T/G; MAF=0.013; FDR=0.02) and OsGRg25613 (chr1; T/A; MAF=0.25; FDR=0.03) suggest the involvement of resistance genes, possibly encoding receptor-like kinases, WRKY transcription factors, or other pathogen-response elements. The presence of a low-frequency allele with a strong FDR value highlights its potential as a rare but valuable resistance source (Tab. 4).

Trait	SNP (Lowest)	Allele (Lowest)	MAF (Lowest)	FDR value (Lowest)	SNP (Highest)	Allele (Highest)	MAF (Highest)	FDR value (Highest)
Days to 50% Flowering (DF)	OsGRg01715 (chr2)	T/C	0.03	0.039	OsGRb05532 (chr2)	T/C	0.49	0.02
Days to Maturity (DM)	OsGRg05186 (chr3)	A/G	0.03	0.03	OsGRg04475 (chr6)	A/G	0.36	0.02
No. of Seeds/Panicle (G/P)	OsGRg12638 (chr9)	A/C	0.02	0.02	OsGRg15175 (chr9)	A/C	0.04	0.03
Plant Height (PH)	OsGRg06248 (chr6)	A/C	0.021	0.023	OsGRb24963 (chr7)	A/G	0.363	0.052
Panicle Length (PL)	OsGRg04319 (chr1)	T/C	0.021	0.031	OsGRb22951 (chr5)	T/C	0.44	0.037
1000 Grain Weight (TGW)	OsGRb24166 (chr2)	A/G	0.02	0.017	OsGRb17840 (chr11)	A/G	0.36	0.019
Tiller/Plant (T/P)	OsGRg15277 (chr2)	T/C	0.02	0.033	OsGRg03852 (chr3)	A/C	0.057	0.025
Yield	OsGRg01045 (chr9)	T/A	0.02	0.06	OsGRb02133 (chr1)	A/G	0.53	0.014
Seed Length (SL)	OsGRb21889 (chr11)	A/G	0.02	0.04	OsGRb21987 (chr9)	A/C	0.14	0.017
Seed Width (SW)	OsGRb08457 (chr6)	A/C	0.016	0.02	OsGRb08496 (chr6)	A/G	0.32	0.02
Flag Leaf Area (FLA)	OsGRb28544 (chr5)	A/G	0.02	0.015	OsGRb01831 (chr1)	A/G	0.34	0.03
Panicle Width (PW)	OsGRb23923 (chr3)	A/G	0.02	0.022	OsGRb24065 (chr1)	A/G	0.33	0.07
Effective Tillers (ET)	OsGRg03954 (chr8)	A/G	0.013	0.18	OsGRb13097 (chr6)	A/G	0.05	0.08
Weight of Seeds per Panicle	OsGRb09045 (chr5)	T/C	0.08	0.02	OsGRb38178 (chr11)	A/C	0.55	0.09
BPB	OsGRg201457 (chr3)	T/G	0.013	0.02	OsGRg25613 (chr1)	T/A	0.25	0.03

Table 4. Comparison of the lowest and highest MAF and FDR values for various traits in rice.

Population structure analysis

Structure analysis of 208 SNPs revealed two distinct subpopulations ($K=2$), with Pop1 comprising varieties like BAS 370, Super Basmati, and PK 386 (depicted in red), and Pop2 including genetically distinct types such as Punjab Basmati, Basmati 2019, and Super Gold (depicted in green). The Delta K method supported $K=2$ as the most informative model of population structure. This differentiation reflects breeding history and geographical origin and offers a framework for stratified association studies and future varietal development (Fig. 2).

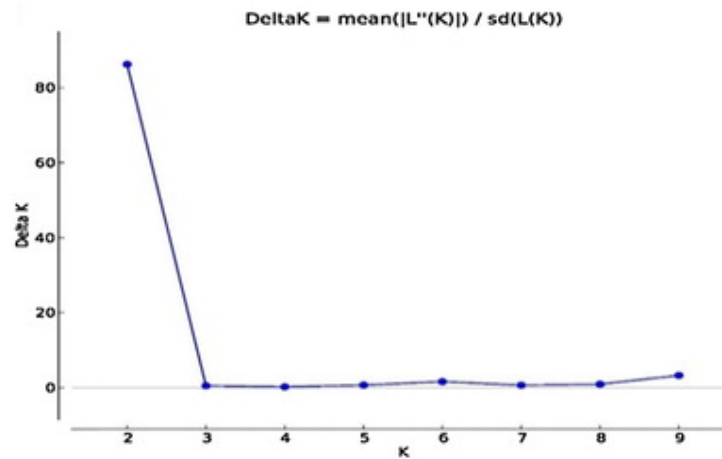


Figure 2a. Population structure analysis of 14 rice varieties; (a) Delta K showing the number of populations.



Figure 2b. Population structure analysis of 14 rice varieties; (b) Bar plot of populations sorted by kinship matrix (Pop 1=population 1, Pop 2=population 2).

Collectively, the results provide novel insights into the genetic basis of complex traits in rice and support the identification of key loci for marker-assisted selection and varietal improvement.

Discussion

The present study employed a genome-wide association approach using GBS to explore genetic diversity, trait-marker associations, and population structure in 14 Pakistani rice varieties. High-throughput sequencing generated a robust dataset characterized by consistent quality, minimal read loss during preprocessing, and high alignment efficiency. Most varieties retained over 97% of reads post-processing, with alignment rates exceeding 86% and unique alignment proportions typically above 96%, underscoring the reliability of sequencing and mapping to the reference genome. Such high-quality metrics align with established standards for GBS-based studies and are critical for downstream applications, including SNP discovery and trait association mapping.

The variation in raw reads and GC content (ranging from ~40.2% to 43.8%) across genotypes suggested minor differences in genomic composition, yet all values remained within an optimal range for efficient sequencing. Notably, 'PK 386' yielded the highest raw read count, whereas 'super gold' retained the highest percentage of reads post-trimming. These differences could influence sequencing depth and SNP detection efficiency across genomes, particularly in regions of low coverage. Similar metrics have been reported by Biscarini, et al. 2016 in temperate rice, emphasizing the utility of such datasets for high-resolution association studies.

A total of 208 significant SNPs were identified through a Mixed Linear Model (MLM) at a threshold of $-\log_{10}(P) \geq 4$, demonstrating associations with 15 agronomic and yield-related traits. The Phenotypic Variation Explained (PVE) ranged from 6% to 28.3%,

confirming both minor and major genetic effects underlying complex traits. The highest PVE values were recorded for traits such as Days to 50% flowering (28.3%), plant height (23.2%), and yield (24.7%), identifying strong candidate loci for Marker-Assisted Selection (MAS). Manhattan plots revealed significant peaks across several chromosomes, particularly 2, 3, 7, 8, and 11, marking these regions as potential QTL hotspots for agronomic traits. Complementary Q-Q plots validated the statistical significance of these associations.

Notably, several pleiotropic SNPs were identified, such as OsGRb06041 (Chr 2), influencing both Days to 50% Flowering and 1000 Grain Weight, and OsGRg05186 (Chr 3), associated with Days to Maturity, Panicle Width, and Tillers per Plant. The presence of such loci highlights genomic regions where regulatory elements may simultaneously govern multiple traits, presenting a strategic opportunity for multi-trait improvement in breeding programs. Similar findings have been reported in other rice GWAS studies (Subedi, et al. 2019, Malik, et al. 2022), further validating the reliability of detected associations.

The SNP density observed in this study—an average inter-marker distance of 6.0 kb with 83.2% of SNPs falling within 10 kb of their nearest neighbor—facilitates fine-mapping and minimizes the confounding effects of linkage disequilibrium. This resolution aligns with or surpasses previous studies (Biscarini, et al. 2016, Volante, et al. 2020) and ensures a high degree of precision in identifying trait-associated loci. Quality control and statistical analyses using PLINK and R Studio further reinforced the robustness of the results, with FDR adjustments accounting for multiple testing errors and MAF thresholds improving marker reliability.

The population structure analysis, based on the 208 SNPs, resolved the rice germplasm into two primary genetic clusters (Pop1 and Pop2) at $K=2$, as indicated by the highest Delta K value. Pop1 included traditional Basmati cultivars like BAS 370, Super Basmati, and BAS Pak, showing high genetic similarity and strong association with typical Basmati traits (e.g., aroma, long grain). In contrast, Pop2 encompassed genetically divergent varieties such as Punjab basmati, basmati 2019, and super gold, which may have been developed through modern breeding approaches for enhanced yield or stress resilience. The clear divergence between these groups suggests opportunities for exploiting heterosis in future crosses. This binary clustering contrasts with the higher K values (e.g., $K=5$ or $K=6$) reported by Biscarini, et al. 2016 and Sao, et al. 2024, indicating a narrower genetic base within Pakistani Basmati varieties, possibly due to localized breeding strategies.

Phenotypic evaluations revealed that genotypes such as BAS 370, BAS 198, BAS 385, and basmati 2019 were superior in disease resistance and yield potential, supported by high PVE values for associated SNPs. In contrast, super basmati and super gold showed higher susceptibility to diseases and lacked SNPs with significant PVE values, suggesting limited utility under stress-prone conditions. These findings provide clear direction for breeding programs: varieties with high PVE-linked SNPs are strong candidates for introgression into elite backgrounds, while susceptible varieties may benefit from genetic enhancement or exclusion from core breeding pools.

The implications of this study are multifold. Genomic regions harboring high-impact SNPs represent key targets for MAS, potentially accelerating the development of climate-resilient, high-yielding rice cultivars. By reducing reliance on chemical inputs and improving yield stability, the deployment of these improved genotypes supports goals of sustainable agriculture and food security. Additionally, the integration of SNP information into breeding pipelines can reduce breeding cycles and enhance selection efficiency.

Economically, improved rice varieties can bolster farmer incomes, reduce production costs, and contribute to national GDP growth. For rice-exporting countries like Pakistan, enhanced Basmati varieties with superior grain quality and disease resistance can improve competitiveness in global markets. By aligning agronomic improvement with economic and environmental sustainability, this research offers a model for integrating molecular tools into crop improvement programs.

Conclusion

This study integrates GBS with genome-wide association analysis to unveil the genetic architecture of key agronomic and yield-related traits in Pakistani rice germplasm. Through high-resolution SNP discovery and association mapping, 208 significant loci were identified, some of which demonstrated pleiotropic effects across multiple traits, including yield, flowering time, grain weight, and disease resistance. The results confirmed the presence of valuable alleles within local rice varieties and delineated two genetically distinct subpopulations, offering crucial insights into the structure and diversity of regional germplasm.

By revealing strong candidate markers for MAS, this research lays a foundation for precision breeding programs aimed at improving both productivity and resilience in rice. The identification of elite varieties harboring superior alleles further enables the development of next-generation cultivars capable of withstanding biotic and abiotic stresses under changing environmental conditions. Moreover, the integration of GBS and GWAS in this study provides a cost-effective model that can be adopted in other crops for rapid genetic improvement. Ultimately, this work contributes significantly to sustainable rice cultivation and long-term food and economic security in rice-growing regions.

References

- Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol.* 30:174-178.
- Bayer PE, Ruperao P, Mason AS, Stiller J, Chan CKK, Hayashi S, Long Y. (2015). High resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*. *Theor Appl Genet.* 128:1039-1047.

- Biscarini F, Cozzi P, Casella L, Riccardi P, Vattari A, Orasen G, Perrini R, Tacconi G, Tondelli A, Biselli C, Cattivelli L, Spindel J, McCouch S, Abbruscato P, Vale G, Piffanelli P, Greco R. (2016).** Genome-wide association study for traits related to plant and grain morphology, and root architecture in temperate rice accessions. *PLoS One*. 11:e0155425.
- Bus A, Hecht J, Huettel B, Reinhardt R, Stich B. (2012).** High throughput polymorphism detection and genotyping in *Brassica napus* using next generation RAD sequencing. *BMC Genom*. 13:281.
- Deschamps S, Llaca V, May GD. (2012).** Genotyping-by-sequencing in plants. *Biology*. 1:460-483.
- Doyle JJ, Doyle JL. (1987).** A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*. 19:11-15.
- Edwards D, Batley J, Snowdon RJ. (2013).** Accessing complex crop genomes with next generation sequencing. *Theor Appl Genet*. 126:1-11.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. (2011).** A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS One*. 6:e19379.
- Furuta T, Ashikari M, Jena KK, Doi K, Reuscher S. (2017).** Adapting genotyping-by-sequencing for rice F2 populations. *G3*. 7:881-893.
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. (2005).** Genetic structure and diversity in *Oryza sativa* L. *Genetics*. 169:1631-1638.
- Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA. (2009).** A first generation haplotype map of maize. *Science*. 326:1115-1117.
- Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan, J. (2009).** High throughput genotyping by whole genome resequencing. *Genome Res*. 19:1068-1076.
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B. (2010).** Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet*. 42:961-967.
- Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, Li W, Zhan Q, Cheng B, Xia J, Chen N, Hao Z, Liu K, Zhu C, Huang T, Zhao Q, Zhang L, Fan D, Zhou, C, Lu Y, Weng Q, Wang ZX, Li J, Han B. (2015).** Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nature Commun*. 6:6258.
- Kujur A, Upadhyaya HD, Shree T, Bajaj D, Das S, Saxena MS, Badoni S. (2015).** Ultra high density intra specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. *Sci Rep*. 5:9468.
- Malik P, Huang M, Neelam K, Bhatia D, Kaur R, Yadav B, Singh J, Sneller C, Singh K. (2022).** Genotyping-by-sequencing based investigation of population structure and genome wide association studies for seven agronomically important traits in a set of 346 *Oryza rufipogon* accessions. *Rice*. 15:37.
- Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera Lizarazu O. (2013).** Population genomic and genome wide association studies of agroclimatic traits in Sorghum. *Proc Natl Acad Sci USA*. 110:453-458.
- Norton GJ, Travis AJ, Douglas A, Fairley S, Alves EP, Ruang-areerate P, Naredo MEB, McNally KL, Hossain M, Islam MR, Price AH. (2018).** Genome wide association mapping of grain and straw biomass traits in the rice Bengal and Assam Aus Panel (BAAP) grown under alternate wetting and drying and permanently flooded irrigation. *Front Plant Sci*. 9:1223.
- Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sanchez-Villeda H, Sorrells M, Jannink JL. (2012).** Genomic selection in wheat breeding using genotyping-by-sequencing. *TPG*. 5:103-113.
- Reynolds MP, Borlaug NE. (2006).** Impacts of breeding on international collaborative wheat improvement. *J Agri Sci*. 144:3-17.
- Sao A, Nair SK, Gauraha D, Premi V, Sahu H, Mannade AK, Chandel G. (2024).** Genome-wide association studies for important agronomic yield attributing traits and resistance to major diseases in rice (*Oryza sativa* L.). *AMA*. 55:17039-17059.
- Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redona E, Atlin G, Jannink JL, McCouch SR. (2015).** Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genetics*. 11:e1005350.
- Subedi SR, Sandhu N, Singh VK, Sinha P, Kumar S, Singh SP, Ghimire SK, Pandey M, Yadaw RB, Varshney RK, Kumar A. (2019).** Genome-wide association study reveals significant genomic regions for improving yield, adaptability of rice under dry direct seeded cultivation condition. *BMC Genomics*. 20:471.
- Swamy BPM, Shamsudin NAA, Rahman SNA, Mauleon R, Ratnam W, Cruz MTS, Kumar A. (2017).** Association mapping of yield and yield related traits under reproductive stage drought stress in rice (*Oryza sativa* L.). *Rice*. 10:21.
- Uitdewilligen JG, Wolters AM, D'Hoop BB, Borm TJ, Visser RG, Eck HJ. (2013).** A next generation sequencing method for genotyping by sequencing of highly heterozygous autotetraploid potato. *PLoS One*. 8:e62355.
- Volante A, Tondelli A, Desiderio F, Abbruscato P, Menin B, Biselli C, Casella L, Singh N, McCouch SR, Tharreau D, Zampieri E, Cattivelli L, Vale G. (2020).** Genome wide association studies for japonica rice resistance to blast in field and controlled conditions. *Rice*. 13:71.
- Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L. (2016).** Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature Genetics*. 48:927-936.
- Yu J, Buckler ES. (2006).** Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol*. 17:155-160.
- Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH. (2011).** Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Commun*. 2:467.
- Zheng XM, Gong T, Ou HL, Xue D, Qiao W, Wang J, Liu S, Yang Q, Olsen KM. (2018).** Genome-wide association study of rice grain width variation. *Genome*. 61:233-240.