



OPEN ACCESS

EDITED BY

Rajni Parmar,
Tennessee State University, United States

REVIEWED BY

Yang-Jun Wen,
Nanjing Agricultural University, China
Tian Qing Zheng,
Chinese Academy of Agricultural
Sciences, China
Wricha Tyagi,
Central Agricultural University, India
Rajendra Kumar,
Indian Agricultural Research Institute
(ICAR), India

*CORRESPONDENCE

Erbao Liu,
✉ liuerbao@ahau.edu.cn

†These authors have contributed equally
to this work

RECEIVED 13 June 2023

ACCEPTED 16 August 2023

PUBLISHED 04 September 2023

CITATION

Liu Z, Sun H, Zhang Y, Du M, Xiang J, Li X,
Chang Y, Sun J, Cheng X, Xiong M, Zhao Z
and Liu E (2023), Mining the candidate
genes of rice panicle traits via a genome-
wide association study.
Front. Genet. 14:1239550.
doi: 10.3389/fgene.2023.1239550

COPYRIGHT

© 2023 Liu, Sun, Zhang, Du, Xiang, Li,
Chang, Sun, Cheng, Xiong, Zhao and Liu.
This is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Mining the candidate genes of rice panicle traits via a genome-wide association study

Zhengbo Liu[†], Hao Sun[†], Yanan Zhang[†], Mingyu Du[†], Jun Xiang,
Xinru Li, Yinping Chang, Jinghan Sun, Xianping Cheng,
Mengyuan Xiong, Zhe Zhao and Erbao Liu*

College of Agronomy, Anhui Agricultural University, Hefei, China

Panicle traits are important for improving the panicle architecture and grain yield of rice. Therefore, we performed a genome-wide association study (GWAS) to analyze and determine the genetic determinants of five panicle traits. A total of 1.29 million single nucleotide polymorphism (SNP) loci were detected in 162 rice materials. We carried out a GWAS of panicle length (PL), total grain number per panicle (TGP), filled grain number per panicle (FGP), seed setting rate (SSR) and grain weight per panicle (GWP) in 2019, 2020 and 2021. Four quantitative trait loci (QTLs) for PL were detected on chromosomes 1, 6, and 9; one QTL for TGP, FGP, and GWP was detected on chromosome 4; two QTLs for FGP were detected on chromosomes 4 and 7; and one QTL for SSR was detected on chromosome 1. These QTLs were detected via a general linear model (GLM) and mixed linear model (MLM) in both years of the study period. In this study, the genomic best linear unbiased prediction (BLUP) method was used to verify the accuracy of the GWAS results. There are nine QTLs were both detected by the multi-environment GWAS method and the BLUP method. Moreover, further analysis revealed that three candidate genes, *LOC_Os01g43700*, *LOC_Os09g25784*, and *LOC_Os04g47890*, may be significantly related to panicle traits of rice. Haplotype analysis indicated that *LOC_Os01g43700* and *LOC_Os09g25784* are highly associated with PL and that *LOC_Os04g47890* is highly associated with TGP, FGP, and GWP. Our results offer essential genetic information for the molecular improvement of panicle traits. The identified candidate genes and elite haplotypes could be used in marker-assisted selection to improve rice yield through pyramid breeding.

KEYWORDS

rice, panicle traits, genome-wide association study, favorable haplotypes, candidate genes

1 Introduction

Rice (*Oryza sativa* L.) is one of the most important food crop species and serves as the primary nutrition source for almost half of the world's population (Loos et al., 2014). With the increase in the world population, food shortages have become a serious global problem. Therefore, increasing production has become a key target in rice breeding. Rice yield is a complex trait multiplicatively determined by four main factors: number of panicles, number of grains per panicle, grain weight (GW), and proportion of filled grain (Li et al., 2013). Panicle length (PL) is one aspect of panicle architecture and is usually considered a yield-related trait. PL, together with spikelet number and density and seed setting rate (SSR),

determines the grain number per panicle (Liu et al., 2016). An increase in grain number per panicle reflects an increase in yield. In addition, GW is an important agronomic trait that directly affects rice yield (Xing and Zhang, 2010). Elucidation of the genetic mechanism controlling these panicle traits would thus be an efficient way for breeders to improve rice yields.

The panicle trait is a quantitative trait regulated by multiple genes and is greatly affected by the environment. In recent years, many quantitative trait loci (QTLs) and genes associated with panicle traits have been mapped and cloned, which is inseparable from the rapid development of molecular biology techniques however, the molecular mechanisms underlying panicle traits have been insufficiently studied. To date, at a minimum 300 QTLs for panicle traits distributed across 12 chromosomes have been detected (Xing et al., 2002; Kobayashi et al., 2003; Thomson et al., 2003; Mei et al., 2005; Zhang et al., 2015; Zhang et al., 2015; Wang et al., 2019; Bai et al., 2021; Thapa et al., 2021; Zhong et al., 2021). For instance, Agata et al. constructed near-isogenic lines with ST-1 and Koshihikari as genetic backgrounds. Through QTL analysis and fine mapping, two QTLs, *Pr15* and *Pbl6*, were found to independently regulate panicle length in rice (Agata et al., 2020). Zhang et al. conducted a GWAS on 5 panicle traits of 315 rice accessions introduced from the international rice microcore germplasm bank and detected five QTLs related to PL (Zhang et al., 2015). Zhang et al. mapped a QTL regulating rice panicle length, qPL8, by constructing recombinant inbred lines and fine mapping (Zhang et al., 2021). Using 42 single-segment substitution lines (SSSLs) derived from nine donors in the genetic background of HJX74, Wang et al. performed a QTL analysis for PL, and fourteen QTLs for PL were recognized (Wang et al., 2019). Zhong et al. performed a GWAS on 27 traits of 421 homozygous rice accessions, the results of which revealed 21 and 15 QTLs tightly linked to total grain number per panicle (TGP) and PL, respectively (Zhong et al., 2021). In addition, many panicle trait genes are pleiotropic. For example, *OsPIN5b* (Lu et al., 2015) controls PL and affects the SSR and yield of rice; similarly, *Ghd7* (Xue et al., 2008) is a major QTL that controls grain number per panicle, plant height and heading date. Genes including *OsFAD8*, *OsPIN5b*, *HD1*, *Ghd7*, and *EUI1* were found to regulate PL (Xue et al., 2008; Zhang et al., 2008; Huang et al., 2012; Lu et al., 2015; Zhang et al., 2019); among them, the *EUI1* gene positively regulates PL, and the *OsFAD8*, *OsPIN5b*, *HD1*, and *Ghd7* genes negatively regulate PL. A total of five genes controlling grain number per panicle, GW, and SSR have been cloned (Song et al., 2007; Si et al., 2016; Liu et al., 2018; Xiang et al., 2019; Tu et al., 2022). Specifically, *GLW7* encodes a protein that is a positive regulator of cell proliferation and enhances the grain width, grain size, and yield of rice (Si et al., 2016). The GW-related gene *GW2* was found to encode a previously unknown RING-type protein with E3 ubiquitin ligase activity. Loss of *GW2* function enhances the GW, grain width, and yield of rice (Song et al., 2007). Decreased expression of *OsCKX2* leads to the accumulation of cytokinins in inflorescence meristems and increases the number of grains per panicle (Tu et al., 2022). *OsOAT* (Liu et al., 2018) and *LSSR1* (Xiang et al., 2019) are two genes that reportedly regulate the SSR of rice. In the *OsOAT* mutant, metabolic abnormalities induced by nitrogen deficiency in florets result in malformed glumes and anther indehiscence, which affect the pollination process and lead to a low SSR (Liu et al., 2018). *LSSR1* encodes a putative GH5 cellulase

TABLE 1 Description statistics of panicle traits of 162 rice materials.

Phenotype	Year	Mean \pm SD	Range	CV (%)	H _B ² (%)
PL	2019	22.6 \pm 4.7	15.0–35.0	20.8	97.5
	2020	23.3 \pm 5.1	12.9–38.1	21.7	97.8
	2021	23.2 \pm 4.9	12.8–37.7	21.2	98.4
TGP	2019	194.4 \pm 54.8	57.0–426.0	28.2	93.6
	2020	196.2 \pm 56.9	46.0–406.0	29.0	91.9
	2021	192.0 \pm 57.7	53.0–460.0	30.1	87.4
FGP	2019	170.0 \pm 51.3	32.0–383.0	30.2	91.6
	2020	181.0 \pm 52.8	39.0–375.0	29.2	90.5
	2021	174.7 \pm 48.0	43.0–345.0	27.5	97.6
SSR	2019	0.9 \pm 0.1	0.4–1.0	12.7	74.0
	2020	0.9 \pm 0.1	0.6–1.0	6.7	73.5
	2021	0.9 \pm 0.1	0.6–1.0	7.6	94.2
GWP	2019	4.1 \pm 1.1	1.3–7.4	27.4	91.9
	2020	4.0 \pm 1.1	1.1–7.3	28.4	89.2
	2021	4.1 \pm 1.3	0.7–8.2	31.2	79.9

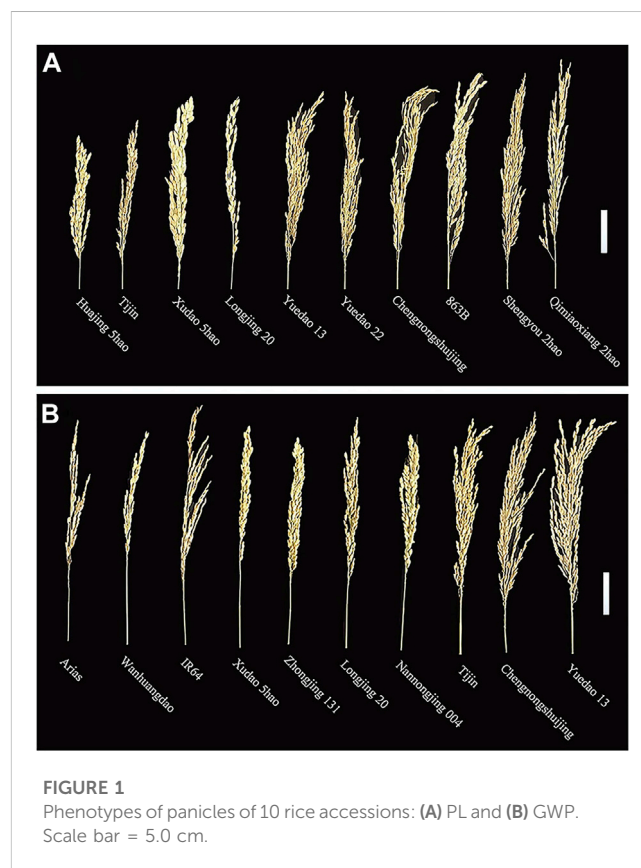


FIGURE 1 Phenotypes of panicles of 10 rice accessions: (A) PL and (B) GWP. Scale bar = 5.0 cm.

that functions during rice fertilization, and *LSSR1* loss-of-function rice mutants presently substantially decreased SSR (Xiang et al., 2019).

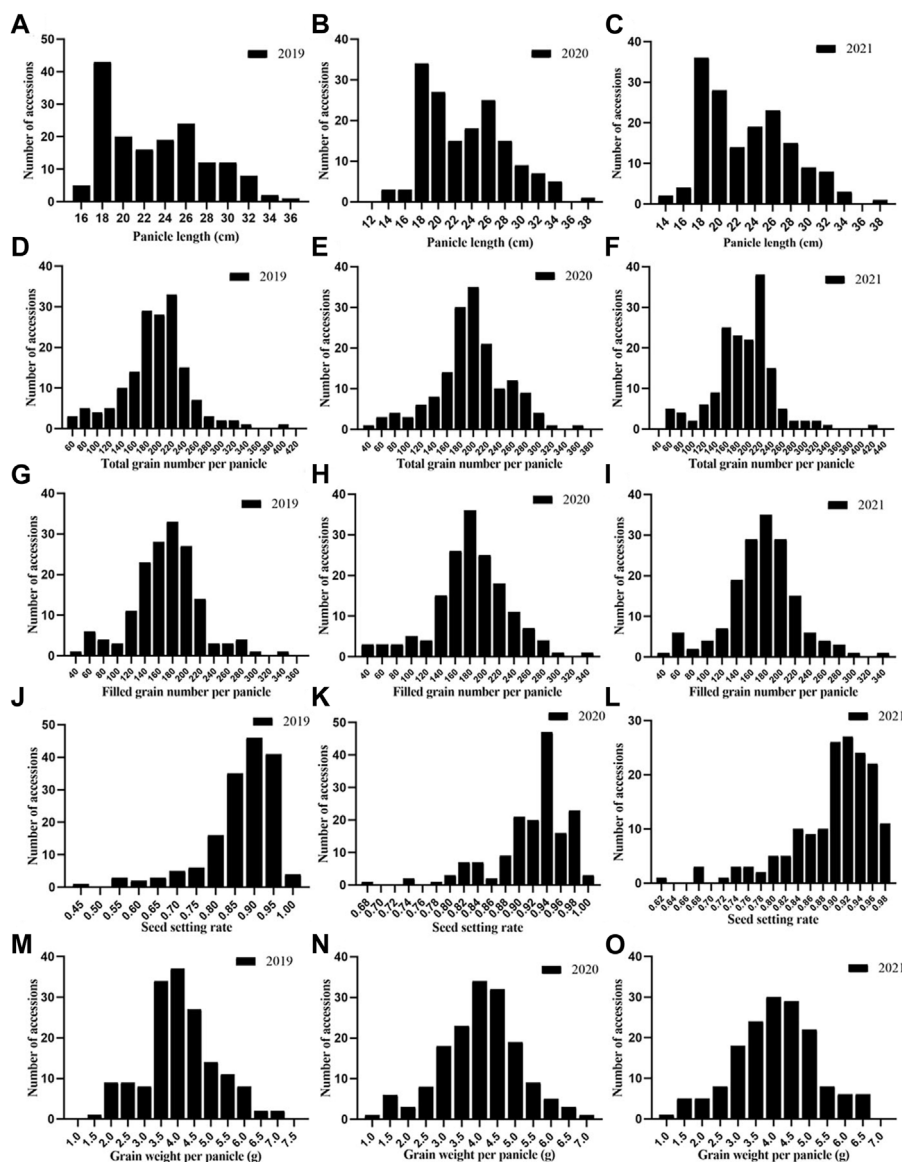
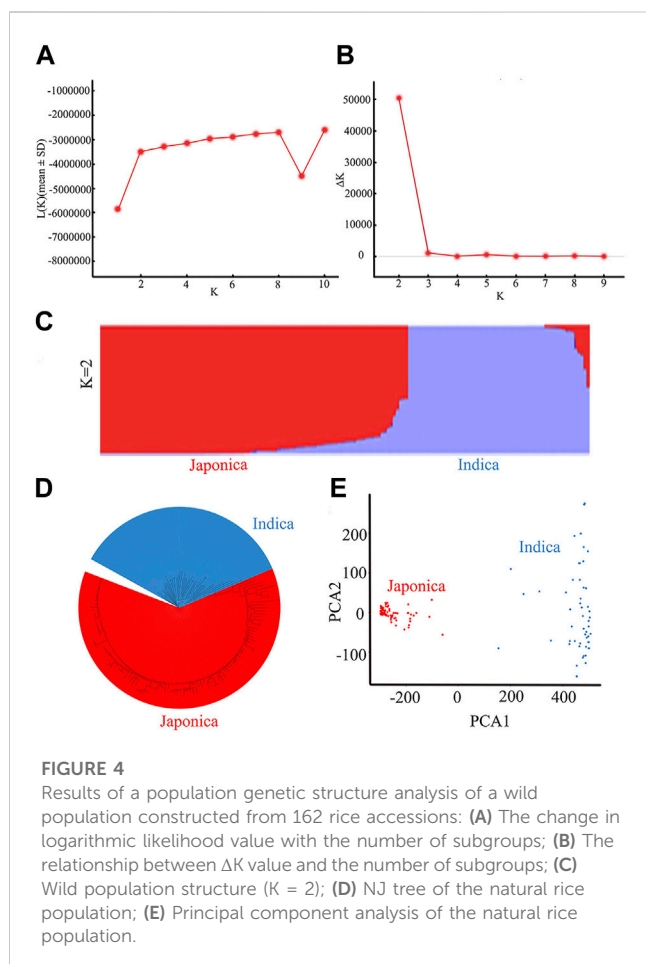
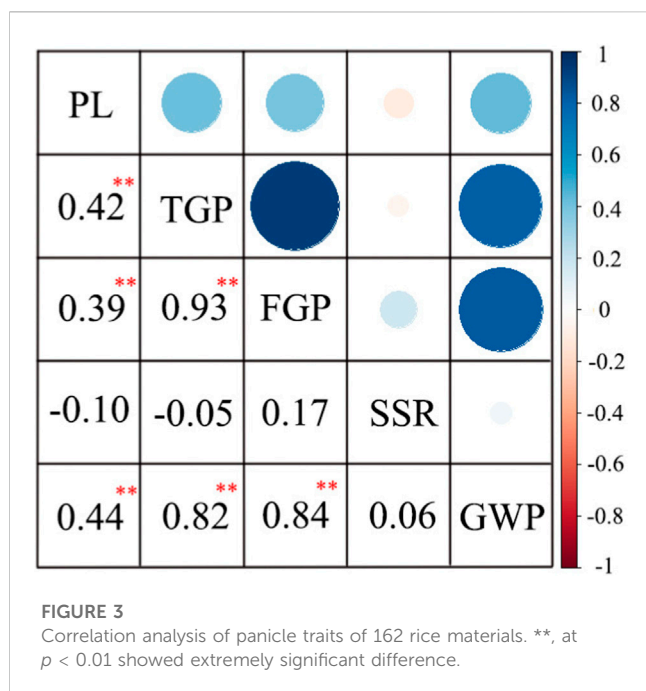


FIGURE 2 Phenotypic frequency distribution of panicle traits of 162 rice germplasms: (A) PL in 2019; (B) PL in 2020; (C) PL in 2021; (D) TGP in 2019; (E) TGP in 2020; (F) TGP in 2021; (G) FGP in 2019; (H) FGP in 2020; (I) FGP in 2021; (J) SSR in 2019; (K) SSR in 2020; (L) SSR in 2021; (M) GWP in 2019; (N) GWP in 2020 and (O) GWP in 2021.

In recent years, genome-wide association studies (GWASs) have become an efficient means for mapping QTLs or genes related to target traits. GWASs based on second-generation sequencing technology enable resequencing of many genomes and provide the possibility of identifying favorable alleles and genetic variations associated with complex traits on a large scale with high accuracy (Huang et al., 2010; Huang et al., 2012). Additionally, GWASs are efficient tools for association mapping of phenotypic and genotypic data, which use the advantage of natural variations to establish associations between traits and single-nucleotide polymorphisms (SNPs) (Zhang et al., 2015; Bai et al., 2016; Han et al., 2016). The two models used to perform GWASs include the general linear model (GLM) and the mixed linear model (MLM), and many genes related to panicle traits have

been successfully mapped (Huang et al., 2012; Crowell et al., 2016; Su et al., 2016; Bai et al., 2021). For example, Bai et al. used genome-wide association analysis (GWAS) to analyze and identify three agronomic traits of 340 rice materials from the 3,000 Rice Genome Project. A total of 153 quantitative trait QTLs were detected, and three new candidate genes that may affect panicle length were found (Bai et al., 2021). A GWAS based on 5 panicle traits of 315 rice accessions introduced from the international rice microcore germplasm bank revealed a total of 36 SNPs (Zhang et al., 2015). Therefore, the research on the key genes involved in the regulation of panicle traits is still ongoing. Therefore, the genes related to panicle traits need to be further studied. In this research, we used a panel of 162 diverse rice accessions to perform a GWAS on five panicle traits. The goals of this research were to (1) identify QTLs associated with



panicle traits, including PL, TGP, filled grain number per panicle (FGP), SSR, and grain weight per panicle (GWP); (2) dissect the genetic architecture of panicle traits and mine the genes; (3) detect

beneficial haplotypes; and (4) identify parents with excellent traits and offer molecular information to improve panicle traits by pyramiding breeding.

2 Materials and methods

2.1 Rice materials

A total of 162 rice accessions were obtained from the State Key Laboratory of Crop and Genetics and Germplasm Innovation of Nanjing Agricultural University (Du et al., 2022). Of the 162 rice germplasms, 128 were from China, and the rest were from other countries, including Vietnam (20), Japan (11), the Philippines (2) and Indonesia (1). More detailed information for each accession, including the accession name, number, latitude, longitude, country of origin and subgroup ancestry of the accession, is given in Supplementary Table S1.

2.2 Phenotype identification

The 162 accessions were planted at the Jiangpu Experimental Station of Nanjing Agricultural University in Nanjing, China, in 2019 and at the Experimental Station of Anhui Agricultural University in Hefei, China, in 2020 and 2021. The varieties were sown in the middle of May and transplanted in the middle of June; each single plant was spaced $26 \text{ cm} \times 16 \text{ cm}$ in all 3 years (2019–2021). A randomized block design with two replications was used for the 3-year experiments. Each variety was planted in 5 rows, with 8 plants in each row. These varieties were subjected to normal fertilizer and water management practices. At maturity, we harvested five panicles of the main culm of each accession and measured the PL, TGP, FGP, SSR, and GWP. The mean phenotypes of each accession were used to calculate the SSR, FGP, and TGP.

$$\text{SSR} (\%) = (\text{FGP}/\text{TGP}) \times 100\%$$

Excel 2010 (Microsoft) software was used for data collation, SPSS software (version 25.0) was used to process the sorted data, and the mean and standard deviation of each agronomic trait were calculated.

Generalized heritability refers to that all genetic variation accounts for the total phenotypic variation. The calculation formula of broad-sense heritability is determined by comparing the relationship between genetic differences among individuals and total variation. In a population, the genetic differences of individuals can be divided into two parts: one is caused by genetic differences, namely, genetic variance (V_A), and the other is caused by environmental factors, namely, environmental variance (V_E). Total variation (V_P) is the sum of genetic variance (V_A) and environmental variance (V_E).

The calculation formula of broad heritability is as follows:

$$H_B^2 = V_A / V_P \times 100\%$$

The value range of generalized heritability is between 0 and 1. The closer the value is to 1, the greater the influence of genetic factors on the trait is. On the contrary, if the generalized heritability is close to 0, it indicates that the trait is mainly affected by environmental factors.

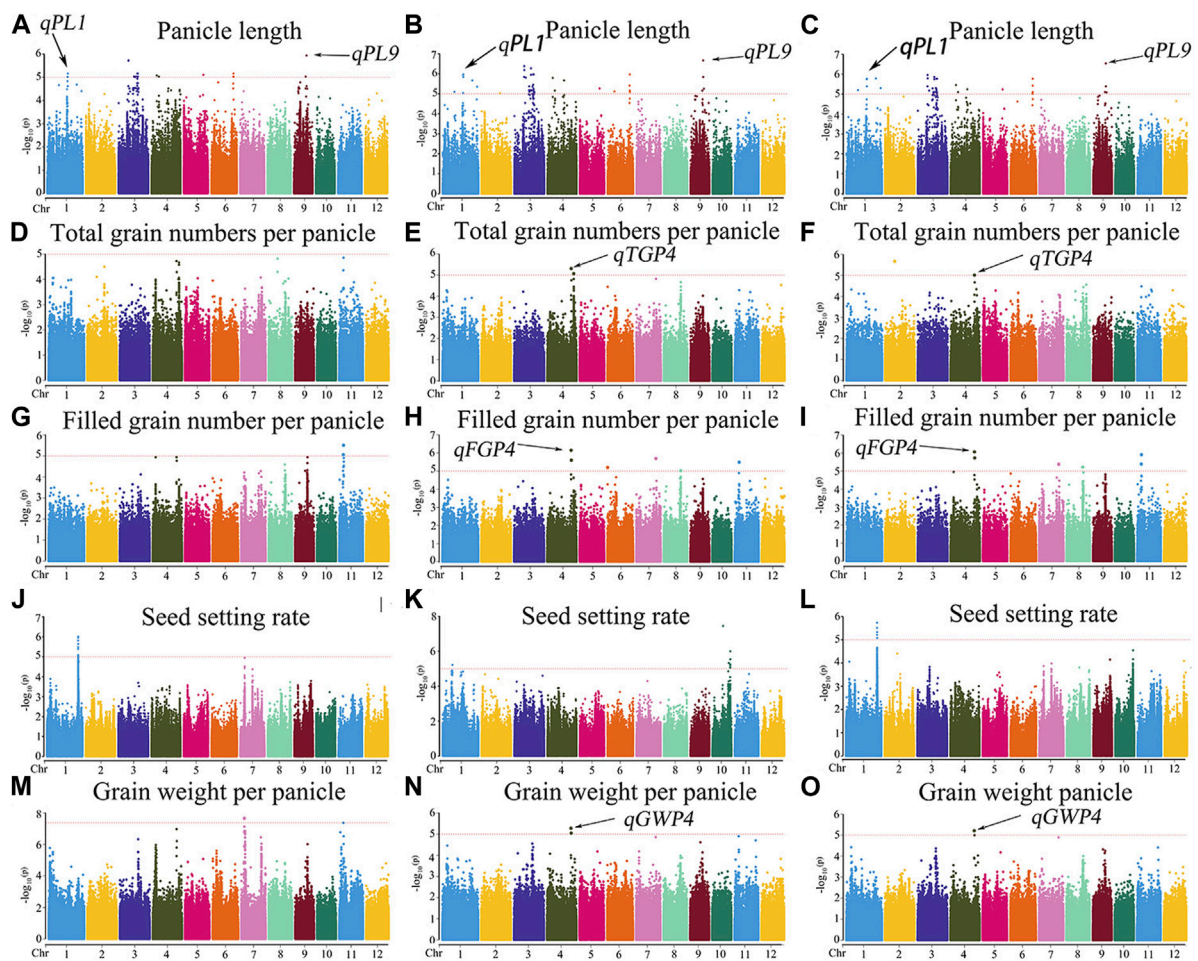


FIGURE 5

Manhattan plots of the GWAS results for PL, TGP, FGP, SSR, and GWP with MLM: (A) PL traits in 2019; (B) PL traits in 2020; (C) PL traits in 2021; (D) TGP traits in 2019; (E) TGP traits in 2020; (F) TGP traits in 2021; (G) FGP traits in 2019; (H) FGP traits in 2020; (I) FGP traits in 2021; (J) SSR traits in 2019; (K) SSR traits in 2020; (L) SSR traits in 2021; (M) GWP traits in 2019; (N) GWP traits in 2020; (O) GWP traits in 2021.

In this study, we performed a correlation analysis on the 3-year data of the five traits. Regarding the threshold, when $|r| \geq 0.8$, it can be considered that the two variables are highly correlated; $0.5 \leq |r| < 0.8$, it can be considered that the two variables are moderately correlated; $0.3 \leq |r| < 0.5$, it can be considered that the two variables are weakly correlated; and $|r| < 0.3$, the two variables can be considered basically irrelevant.

2.3 SNP filter analysis and annotation

To sequence the 162 accessions, young leaves were collected from a single plant at the tillering stage, and genomic DNA was extracted using a standard cetyltrimethylammonium bromide protocol. A double-end sequencing library was constructed using 5 μ g genomic DNA, and the inserted fragment was approximately 350 bp reads. After further processing of the original sequence, the genomic sequence data of 0.532 Tb was obtained after removing joint contamination and low-quality reads. The average genome coverage of each test material was 5.48 times.

Accessions in our population were filtered using PLINK (version 1.9) command and SNPs were further filtered using plink with maf (0.05) and geno (0.2). A total of 1,209,388 SNPs were identified. Library construction, sequencing, and sequence cleaning were performed by staff at Mega Genomics Beijing (<http://www.megagenomics.cn/mobile.php/>, accessed on 10 April 2019).

We used ANNOVAR software (Wang et al., 2010) to perform SNP annotation on the Nipponbare genome sequence. The new annotation results were divided into exon regions, splicing sites, intron regions, and upstream and downstream regions. The SNPs in the coding exons are divided into two types: synonymous and nonsynonymous; nonsynonymous SNPs lead to amino acid changes.

2.4 Population structure and genetic analysis

We used PLINK (version 1.9) software to filter the SNP loci of 162 rice materials, retain the unconnected SNP loci, and convert them into STRUCTURE format (Earl and vonHoldt, 2012). To

TABLE 2 Genome-wide significant association of rice panicle traits.

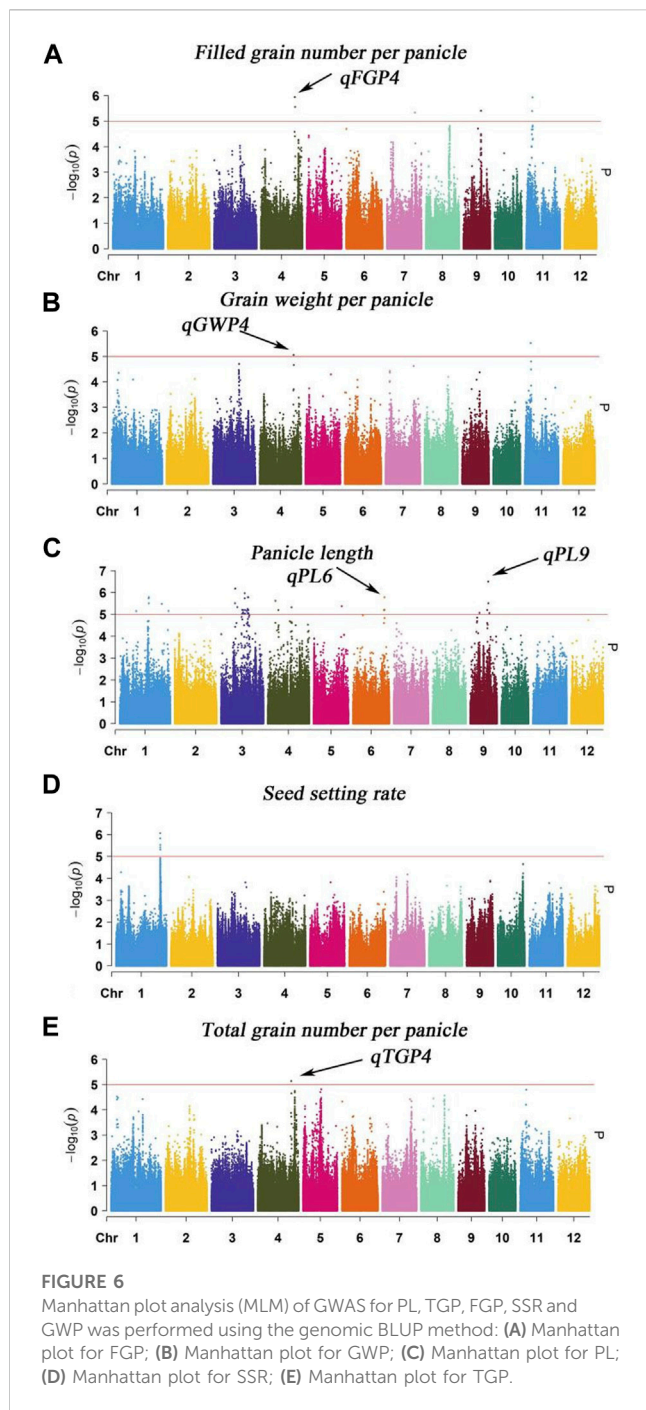
Trait	QTLs	Chr	SNP	Allele	2019	2020	2021	Model	Reported QTLs and genes
					<i>p</i> -value	<i>p</i> -value	<i>p</i> -value		
PL	<i>qPL1</i>	1	25,009,989	A/G	4.1×10^{-6}	5.3×10^{-7}	8.5×10^{-7}	GLM	<i>LPI</i> Liu et al. (2016)
		1	25,009,989	A/G	6.9×10^{-6}	1.1×10^{-6}	1.7×10^{-6}	MLM	
	<i>qPL6</i>	6	26,935,985	A/G	4.5×10^{-6}	8.4×10^{-6}	7.3×10^{-6}	GLM	
		6	26,935,985	A/G	9.3×10^{-6}	7.0×10^{-6}	8.0×10^{-6}	MLM	
	<i>qPL9.1</i>	9	15,393,518	A/C	1.5×10^{-6}	3.7×10^{-7}	3.9×10^{-7}	GLM	
		9	15,393,518	A/C	1.2×10^{-6}	2.2×10^{-7}	2.9×10^{-7}	MLM	
	<i>qPL9.2</i>	9	14,388,593	T/C	4.5×10^{-6}	8.4×10^{-6}	8.4×10^{-6}	GLM	
		9	14,388,593	T/C	9.3×10^{-6}	7.0×10^{-6}	7.0×10^{-6}	MLM	
TGP	<i>qTGP4</i>	4	28,917,371	C/T	5.6×10^{-10}	3.5×10^{-11}	4.9×10^{-10}	GLM	
		4	28,917,371	C/T	—	4.9×10^{-6}	9.7×10^{-6}	MLM	
	<i>qFGP4</i>	4	28,917,371	C/T	3.5×10^{-8}	4.1×10^{-11}	2.4×10^{-10}	GLM	
		4	28,917,371	C/T	—	7.1×10^{-7}	8.6×10^{-7}	MLM	
FGP	<i>qFGP7</i>	7	23,875,734	C/T	—	9.3×10^{-11}	2.1×10^{-9}	GLM	<i>OsBZRI</i> Qiao et al. (2017)
		7	23,875,734	C/T	—	2.0×10^{-6}	4.2×10^{-6}	MLM	
SSR	<i>qSSR1.1</i>	1	37,937,308	G/A	3.5×10^{-7}	—	2.2×10^{-6}	GLM	<i>sd1</i> Ye et al. (2015)
		1	37,937,308	G/A	2.3×10^{-6}	—	8.5×10^{-6}	MLM	
GWP	<i>qGWP4</i>	4	28,917,371	C/T	—	7.5×10^{-10}	1.6×10^{-9}	GLM	
		4	28,917,371	C/T	—	5.3×10^{-6}	6.1×10^{-6}	MLM	

TABLE 3 Genome-wide significant association of rice panicle traits (BLUP).

Trait	QTLs	Chr	SNP	Allele	BLUP	Model	Reported QTLs and genes
					<i>p</i> -value		
PL	<i>qPL1</i>	1	25,009,989	A/G	1.6×10^{-6}	MLM	<i>LPI</i> Liu et al. (2016)
	<i>qPL6</i>	6	26,935,985	A/G	6.3×10^{-6}	MLM	
	<i>qPL9.1</i>	9	15,393,518	A/C	3.2×10^{-7}	MLM	
	<i>qPL9.2</i>	9	14,388,593	T/C	6.3×10^{-6}	MLM	
TGP	<i>qTGP4</i>	4	28,917,371	C/T	7.2×10^{-6}	MLM	
FGP	<i>qFGP4</i>	4	28,917,371	C/T	1.1×10^{-6}	MLM	<i>OsBZRI</i> Qiao et al. (2017)
	<i>qFGP7</i>	7	23,875,734	C/T	4.6×10^{-6}	MLM	
SSR	<i>qSSR1.1</i>	1	37,937,308	G/A	4.8×10^{-6}	MLM	<i>sd1</i> Ye et al. (2015)
GWP	<i>qGWP4</i>	4	28,917,371	C/T	8.7×10^{-6}	MLM	

predict the genetic structure of the entire population, we first used STRUCTURE to analyze the population structure of the germplasm and then analyzed the number of required subgroups (K) through the Structure Harvester website (<http://taylor0.biology.ucla.edu/structureHarvester>, accessed on 8 October 2022). (Bradbury et al., 2007; Falush, Stephens, and Pritchard, 2007; Franois and Durand, 2010). The distance matrix was constructed using VCF2Dis ([https://](https://github.com/BGI-shenzhen/VCF2Dis)

github.com/BGI-shenzhen/VCF2Dis, accessed on 9 October 2022) based on SNPs and an NJ clustering diagram was constructed using MEGA version 7. Principal component analysis was conducted using GCTA (version 1.93.2) software on Linux. PLINK (version 1.9) was used for LD analysis of the genotypic data, and the LD diagram, population structure diagram, NJ tree, and PCA (Zhao et al., 2007) diagram were constructed by R.



2.5 GWAS

In this research, we obtained 1,209,388 SNPs ($MAF > 0.05$) and five groups of phenotypic data. Two models, including GLM and MLM (Yu et al., 2006) in TASSEL (version 5.2.40) software (Bradbury et al., 2007), were used to analyze the associations between the SNPs and phenotypic data. The genome-wide significance thresholds of the GWAS were determined using a modified Bonferroni correction (Li et al., 2012), the threshold of the GLM was set to $p = 4.1 \times 10^{-8}$, and the threshold of MLM was 1.0×10^{-5} . A QTL covers all SNPs located in the same LD region, and

the SNP with the smallest p -value is considered the leading SNP. We used the R package ‘CMplot’ to construct the Manhattan diagram.

In this study, the genomic BEST Linear Unbiased Prediction (BLUP) (Habier et al., 2013) method was used to verify the accuracy of the GWAS results. The BLUP method can integrate multi-environment data, remove environmental effects, and obtain stable genetic phenotypes of individuals. BLUP is a common practice for phenotypic processing, and the lmer function in lme4 in R package is a common method for BLUP analysis. We used the R package ‘CMplot’ to construct the Manhattan diagram, and the threshold of the MLM model is 1.0×10^{-5} , which is the same as the threshold setting of the multi-environment GWAS method.

2.6 Identification of candidate genes and haplotype analysis

Based on the LD decay distance and the results of our GWAS analysis, the candidate areas of genes on chromosomes were estimated, and the genes in the candidate regions were identified at the China Rice Data Center. (<https://www.ricedata.cn/>, accessed on 19 October 2022). Nonsynonymous SNPs lead to amino acid changes, and we focused on nonsynonymous SNPs by comparison with the Nipponbare reference genome sequence (<http://rice.plantbiol-ogy.msu.edu/cgi-bin/gbrowse/rice/>, accessed on 22 October 2022). The identification of candidate genes was mainly through nonsynonymous SNPs in exons and gene function. The nonsynonymous SNPs in exons were selected for haplotype analysis to select favorable haplotypes. The haplotypes of the candidate genes were analyzed based on geographical region and subgroup. Analysis of different phenotypes of candidate gene haplotypes was performed using ANOVA combined with Duncan’s multiple range test. Favorable haplotype analysis of candidate genes was performed using the 3KRG gcHap dataset (<http://bigd.big.ac.cn/gvm/getProjectDetail?project=GVM000123>, accessed on 1 March 2023) and the Rice Functional Genomics and Breeding Database (<https://www.rmbreeding.cn/Public/download>, accessed on 1 March 2023). The ‘favorable’ gcHap of a gene was defined as the one associated with the highest trait value. Five subpopulations of XI (XI-1A, XI-1B, XI-2, XI-3, and XI-adm) and four subpopulations of GJ (temperate GJ [GJ-tmp], subtropical GJ [GJ-sbtrp], tropical GJ [GJ-trp], and GJ-adm) were classified by Wang et al. (2018).

2.7 Forecast of excellent parents

The average positive (negative) haplotype effect (AHE) at a gene locus was calculated as follows:

$$AHE = \sum h_c / n_c$$

Among them, n_c represents the number of haplotypes with positive (negative) effects on gene loci. h_c represents the phenotypic value of haplotypes with positive (negative) effects.

Rice materials with the highest positive haplotype effect were predicted to be the most promising parents for panicle trait improvement in rice breeding at all panicle trait-related gene loci.

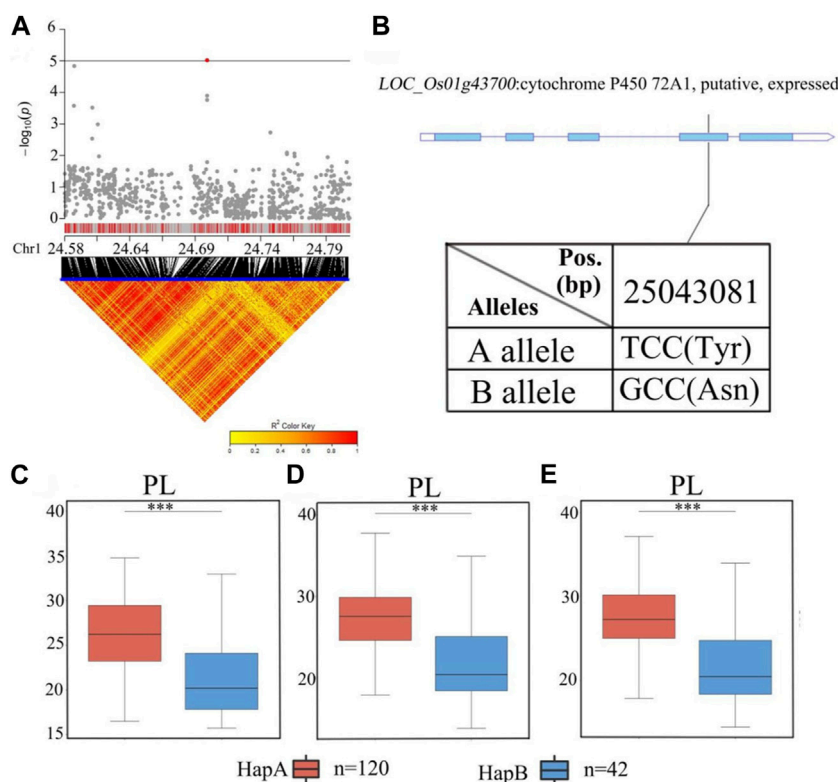


FIGURE 7 The results of haplotype analysis of candidate genes are as follows: (A) Local Manhattan result (top part) and LD heatmap (bottom part); (B) The structure diagram of LOC_Os01g43700 and the SNPs between Hap A and Hap B in LOC_Os01g43700 cDNA. The blue fragment represents the exons; (C) Box line diagram for the PL trait for the two haplotypes (n = 120 versus 42) in 2019; (D) Box plots for PL trait for the two haplotypes (n = 120 versus 42) in 2020; (E) Box plots for PL trait for the two haplotypes (n = 120 versus 42) in 2021. The central lines indicate the median values, and the box edges represent the upper and lower quartiles (***, $p < 0.001$, ANOVA).

3 Results

3.1 Phenotypic variation in panicle traits in the natural population

The panicle traits of 162 rice accessions showed significant differences. Generalized heritability refers to all genetic variation accounting for the total phenotypic variation. All the traits had high generalized heritability, ranging from 73.5% to 98.4% in 2019, 2020 and 2021 (Table 1) Furthermore, it is worth noting that the 3-year broad heritability of PL is greater than 95.0%, which indicates that the variation in this trait is mainly controlled by genes. There were significant differences in PL, TGP, FGP, SSR, and GWP values among varieties in the 3 years, and their coefficients of variation (CV) ranged from 6.7% to 31.2%. Among 162 rice materials, we selected 10 materials to represent the diversity of PL and FWP, including the accession with the greatest PL (Shengyou 2 Hao), the accession with the lowest PL (Huajing 5 Hao), the accession with the greatest GWP (Yuedao 13), and the accession with the lowest GWP (Arias) (Figure 1). The results demonstrated that across the 162 rice accessions, the five panicle traits presented great genetic variation.

In general, the phenotypic changes during all 3 years were similar (Figure 2). Most of the traits appeared to be normally distributed in all 3 years, but SSR and PL showed skewed distributions. The TGP, FGP, and GWP were definitively

normally distributed, indicating that these three traits were not controlled by a single gene but by quantitative traits, which laid a foundation for better understanding the genetic structure of panicle traits. The results above confirmed that there was abundant phenotypic variation across the 162 rice accessions in this study.

The correlation analysis of the 3-year data showed that SSR was negatively correlated with PL and TGP (Figure 3). However, TGP was positively correlated with FGP and GWP, and FGP was more strongly correlated with TGP than with GWP. In general, TGP, FGP, and GWP mirrored one another and were strongly correlated.

3.2 Population structure and linkage disequilibrium (LD) analysis

The 162 rice accessions and SNP markers were used for a genetic structure analysis of the natural population (Figure 4; Supplementary Table S2). Using the Bayesian clustering software STRUCTURE (version 2.3.4) (Evanno, Regnaut, and Goudet, 2005), we calculated varying levels of K means (Figure 4A). When ΔK was the highest, $K = 2$ (Figure 4B). The optimal number of subpopulations was predicted to be $K = 2$ based on a model component analysis by STRUCTURE software (version 2.3.4) and assessing ΔK . Therefore, the rice accessions of 162 were divided into 2 subgroups, namely, an *indica* group and a

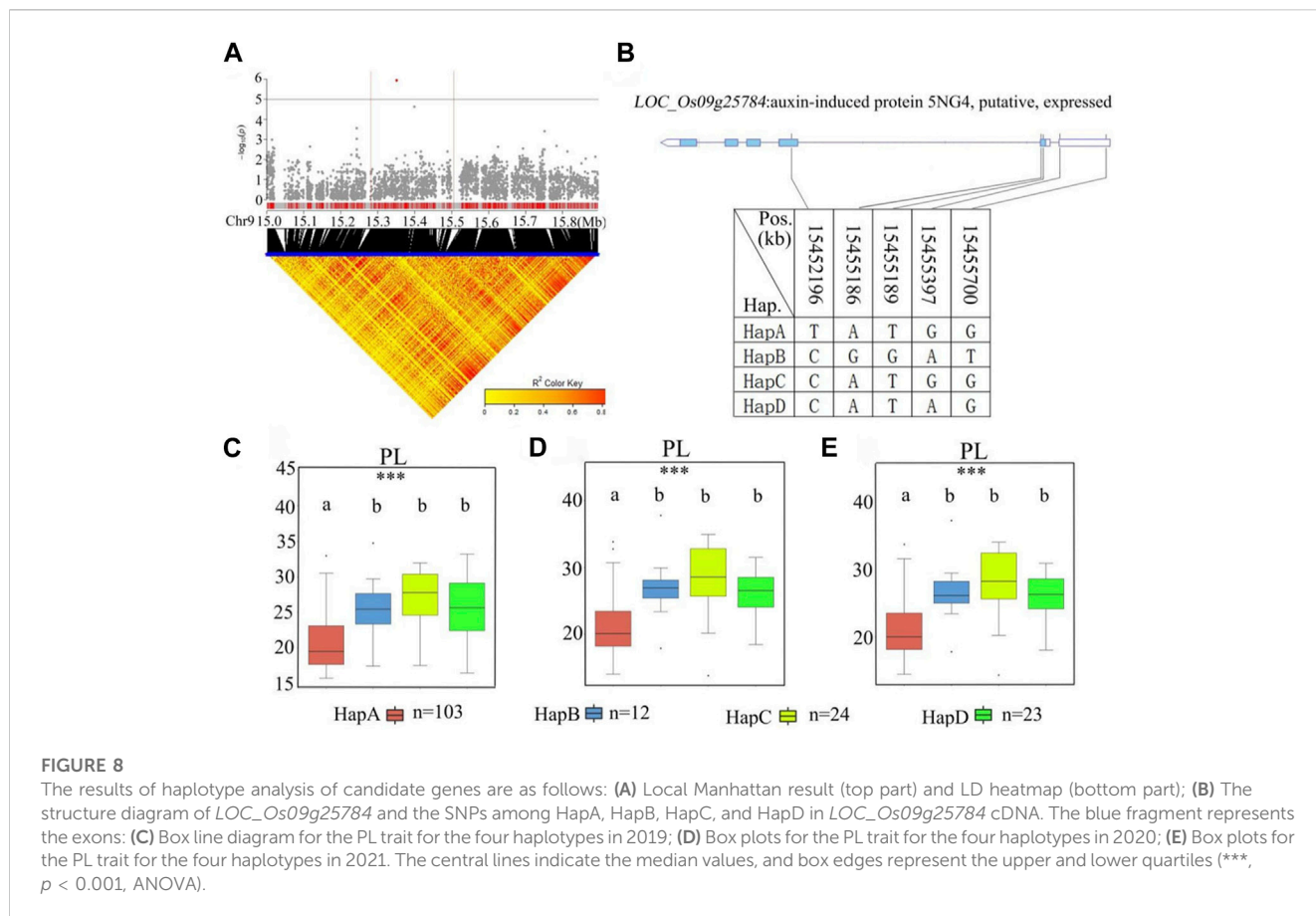


FIGURE 8

The results of haplotype analysis of candidate genes are as follows: (A) Local Manhattan result (top part) and LD heatmap (bottom part); (B) The structure diagram of *LOC_Os09g25784* and the SNPs among HapA, HapB, HapC, and HapD in *LOC_Os09g25784* cDNA. The blue fragment represents the exons; (C) Box line diagram for the PL trait for the four haplotypes in 2019; (D) Box plots for the PL trait for the four haplotypes in 2020; (E) Box plots for the PL trait for the four haplotypes in 2021. The central lines indicate the median values, and box edges represent the upper and lower quartiles (***, $p < 0.001$, ANOVA).

japonica group (Figure 4C). The 162 rice materials included 59 *indica* subpopulations and 103 *japonica* subpopulations. We used the first two principal components (PCs) as covariates within the GWAS model to control for subpopulation structure (Figure 4E). The results of the principal component analysis (PCA) and neighbor-joining (NJ) tree (Figure 4D) were consistent with the structure results. Therefore, the population of 162 rice materials was divided into two subgroups.

The degree of LD was the chromosome distance when the r^2 value was reduced to half of the maximum value. The LD decay distance of the whole population was found to be 112 kb (Du et al., 2022).

3.3 GWAS of PL, TGP, FGP, SSR, and GWP traits

The MLM with correction of kinship bias was mainly used in this study, supplemented by the GLM. We conducted a GWAS between panicle traits and SNPs across the 162 rice accessions. The GWAS results showed that a total of 14 SNPs (QTLs) were significantly associated with PL, TGP, FGP, SSR, and GWP (Figure 5; Supplementary Figure S1; Supplementary Table S3). These QTLs were located on chromosomes 1, 4, 5, 6, 7, 9, 10, and 11 and were repeatedly detected via at least two models. Among the 14 QTLs, 9 stable QTLs were detected in at least 2 years and via two models, and those QTLs were further analyzed (Table 2). For PL, *qPL1*, *qPL6*, *qPL9.1* and *qPL9.2* were identified on chromosomes 1, 6, and 9, and these QTLs

were detected in 3 years and via two models. For SSR, *qSRR1* was identified on chromosome 1 and was detected in 2 years and via two models. For TGP, FGP, and GWP, the Chr4_28,917,371 (*qTGP4*, *qFGP4*, and *qGWP4*) locus on chromosome 4 was found to be associated with TGP, FGP, and GWP simultaneously in at least 2 years and via two models. *qTGP4*, *qFGP4*, and *qGWP4* were the same QTL; hereafter, this QTL is referred to as *qTGP4*. In addition, *qFGP7* was identified on chromosome 7 and was detected in both models for two consecutive years (Table 2).

3.4 GWAS results of single environment method of PL, TGP, FGP, SSR, and GWP traits

The MLM with correction of kinship bias was mainly used in this study. We conducted a GWAS between panicle traits and SNPs across the 162 rice accessions. The GWAS results showed that a total of 9 SNPs (QTLs) were significantly associated with PL, TGP, FGP, SSR, and GWP (Table 3; Figure 6). For PL, *qPL1*, *qPL6*, *qPL9.1* and *qPL9.2* were identified on chromosomes 1, 6, and 9. For SSR, *qSRR1* was identified on chromosome 1. For TGP, FGP, and GWP, the Chr4_28,917,371 (*qTGP4*, *qFGP4*, and *qGWP4*) locus on chromosome 4 was found to be associated with TGP, FGP, and GWP simultaneously, *qTGP4*, *qFGP4*, and *qGWP4* were the same QTL; hereafter, this QTL is referred to as *qTGP4*. In addition, *qFGP7* was identified on chromosome 7. In summary, the results of single-environment GWAS are consistent with the results of multi-environment GWAS.

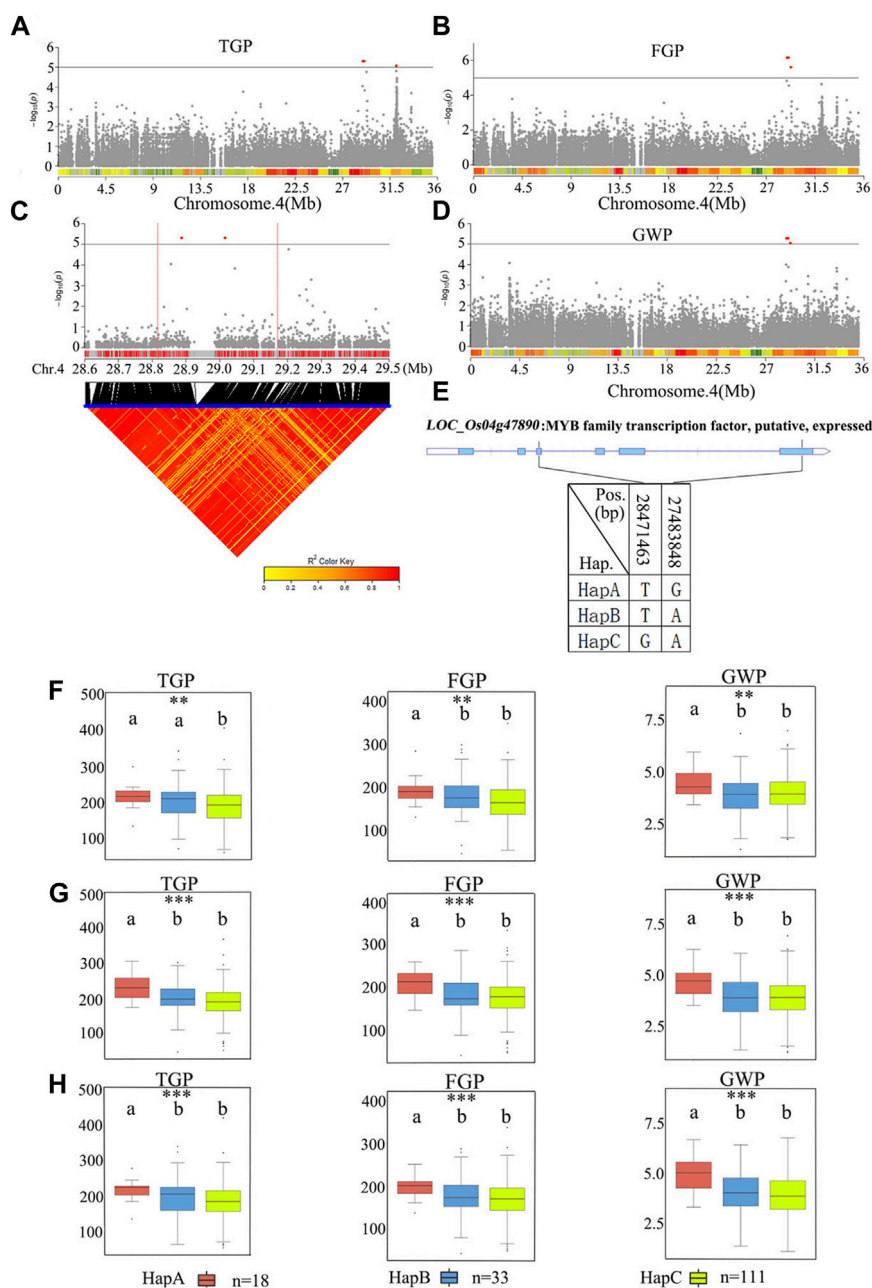


FIGURE 9

The results of haplotype analysis of candidate genes are as follows: (A) Manhattan result chart of TGP trait; (B) Manhattan result chart of FGP trait; (C) Local Manhattan result (top part) and LD heatmap (bottom part); (D) Manhattan plot for GWP. The dashed lines represent significance thresholds; (E) The structure diagram for TGP, FGP, and GWP traits for the three haplotypes in 2019; (F) Box line diagram for TGP, FGP, and GWP traits for the three haplotypes in 2020; (G) Box plots for the TGP, FGP, and GWP traits for the three haplotypes in 2021. The central lines indicate the median values, and the box edges represent the upper and lower quartiles (**, $p < 0.01$; ***, $p < 0.001$, ANOVA).

3.5 Identification of candidate genes for *qPL1*

There were 38 candidate genes associated with significant SNPs in *qPL1* (chr1_24,700,283) in the range of 24.9–25.1 Mb (Figure 7; Supplementary Table S4). In this region, 10 of 38 genes contained nonsynonymous SNPs (Supplementary Table S4). Only one nonsynonymous SNP in *LOC_Os01g43700* was

found to be significantly associated with PL. This gene encodes a cytochrome P450 protein. Cytochrome P450 plays vital roles in promoting plant growth and development (Distefano et al., 2021). According to the SNP site in the cDNA sequence of the *LOC_Os01g43700* gene, it was divided into two haplotypes (Figure 7B). The average PL values of 120 accessions carrying the *LOC_Os01g43700* allele were 27.0 ± 4.0 cm. The average PL 42 accessions carrying the allele

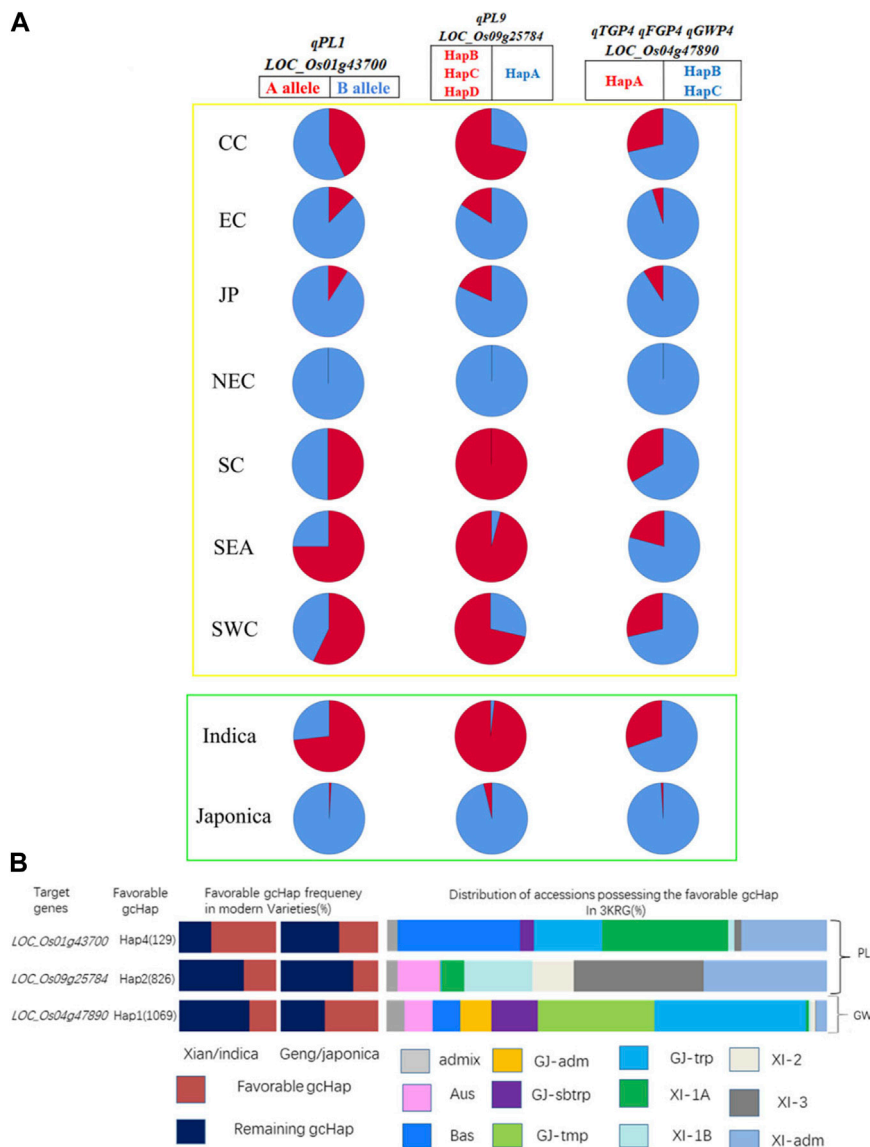


FIGURE 10

Haplotype distribution of candidate genes. (A) We mapped the haplotypes of candidate genes in 2 subgroups and 7 geographical populations. Red and blue were used to represent favorable and unfavorable haplotypes, respectively. The yellow box represents the areas, and the green box represents the subgroups. SC, South China; CC, Central China; EC, East China; NEC, Northeast China; SWC, Southwest China; JP, Japan; SEA, Southeast Asia; (B) Frequencies of the favorable g-Hap of 3 important candidate genes affecting four panicle traits (PL, TGP, FGP and GWP) in XI and GJ modern varieties and different rice subpopulations.

LOC_Os01g43700 were 21.7 ± 4.3 cm. Analysis of variance showed that the differences in PL values of the two alleles were highly significant, and the B allele was significantly associated with lower PL than the A allele. (Figures 7C–E). Therefore, *LOC_Os01g43700* may be a candidate gene for PL.

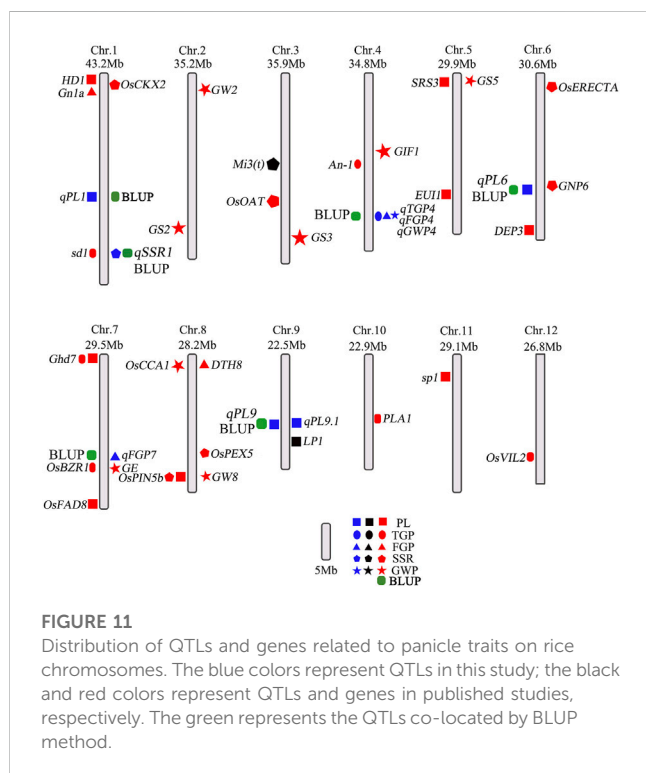
3.6 Identification of candidate genes for *qPL9*

A stabilized QTL, *qPL9*, was found to include the previously reported *LPI* QTL that affects PL through a new regulatory pathway

(Liu et al., 2016). For *qPL9* (chr9_15,393,518) in the 15.3–15.5 Mb region, there were 20 candidate genes related to remarkable SNPs (Figure 8; Supplementary Table S5). Among the 20 candidate genes, 12 genes had nonsynonymous mutations (Supplementary Table S5). Five nonsynonymous SNPs in *LOC_Os09g25784* were found to be significantly associated with PL. *LOC_Os09g25784* encodes an auxin-induced protein, 5NG4. A previous study demonstrated that auxins are essential molecules that control almost every aspect of plant development (Paque and Weijers, 2016). For *LOC_Os09g25784*, all materials were divided into four haplotypes according to the SNP on the cDNA of this locus (Figure 8B). The average PL values of the four haplotypes were 20.9 ± 3.7 cm, $26.3 \pm$

TABLE 4 Predicting excellent parents for improving panicle traits.

Best predicted parents	Predicted PL improvement (cm)		Predicted TGP improvement	Predicted FGP improvement	Predicted GWP improvement (g)
	<i>LOC_Os01g43700</i>	<i>LOC_Os09g25784</i>			
Baikenuo (<i>indica</i>)	HapA (4.03)	HapB (3.30)	HapA (27.00)	HapA (25.00)	HapA (0.72)
Yuedao 61 (<i>indica</i>)	HapA (4.03)	HapC (4.83)	HapA (27.00)	HapA (25.00)	HapA (0.72)
Yuedao 109 (<i>indica</i>)	HapA (4.03)	HapD (2.92)	HapB (9.00)	HapB (5.00)	HapB (0.02)
Yuedao 62 (<i>indica</i>)	HapA (4.03)	HapC (4.83)	HapA (27.00)	HapA (25.00)	HapA (0.72)
Nongxiang 26 (<i>indica</i>)	HapA (4.03)	HapC (4.83)	HapA (27.00)	HapA (25.00)	HapA (0.72)
Yuedao 48 (<i>indica</i>)	HapA (4.03)	HapC (4.83)	HapB (9.00)	HapB (5.00)	HapB (0.02)
Zhengdao 10 Hao (<i>indica</i>)	HapA (4.03)	HapB (3.30)	HapB (9.00)	HapB (5.00)	HapB (0.02)
Chuan 5 Xian (<i>indica</i>)	HapA (4.03)	HapB (3.30)	HapA (27.00)	HapA (25.00)	HapA (0.72)
QimiaoXiang 2 Hao (<i>indica</i>)	HapA (4.03)	HapD (2.92)	HapB (9.00)	HapB (5.00)	HapB (0.02)
Shengtangqing (<i>Japonica</i>)	HapA (4.03)	HapD (2.92)	HapA (27.00)	HapA (25.00)	HapA (0.72)



4.4 cm, 27.9 ± 4.6 cm, and 26.0 ± 3.6 cm, respectively. Haplotype analysis of the whole population showed that there was no significant difference in PL values among Hap B, Hap C and Hap D, but the PL of Hap B, Hap C and Hap D were significantly longer than Hap A (Figures 8C–E). Among the four haplotypes, HapA had the shortest panicle length. Therefore, we further studied the gene *LOC_Os09g25784*.

3.7 Identification of candidate genes for *qTGP4*

qTGP4 was found to be a candidate region for TGP, FGP, and GWP. The 28.8–29.2 Mb regions on chromosome 4 contained 81 genes (Figure 9; Supplementary Table S6). Among the 81 candidate genes, 29 genes had nonsynonymous mutations (Supplementary Table S6). A previous study has shown that the functions of MYB proteins in plants include regulation of secondary metabolism, control of cellular morphogenesis and regulation of meristem formation and the cell cycle (Jin and Martin, 1999). Therefore, *LOC_Os04g47890* (which harbors a gene encoding a MYB family transcription factor) was predicted as a candidate gene QTL. According to the SNP site in the cDNA sequence of the *LOC_Os04g47890* gene, it was divided into three haplotypes (Figure 9E). The average TGP and FGP of the three haplotypes ranged from 187.0 ± 53.2 to 221.0 ± 28.2 and from 170.0 ± 48.4 to 200.0 ± 27.9 , respectively. The average GWP ranged from 3.9 ± 1.0 g to 4.8 ± 0.8 g. For TGP, FGP, and GWP, haplotype analysis for populations showed that HapA was associated with significantly greater TGP, FGP, and GWP than HapB and HapC, while there were no significant differences in these traits between HapB and HapC (Figures 9F–H). Therefore, *LOC_Os04g47890* was chosen for further research.

3.8 Haplotype distribution of candidate genes

We summarized the haplotypes of candidate genes based on geographical regions and subgroups (Figure 10A). The favorable haplotypes HapB, HapC and HapD of favorable alleles *LOC_Os01g43700* and *LOC_Os09g25784* were mainly distributed in *indica* rice subgroups such as Southern China (SC), Southwest

China (SWC) and Southeast Asia (SEA) and low latitudes. In contrast, the nonfavorable HapB allele of *LOC_Os01g43700* and the nonfavorable HapA allele of *LOC_Os09g25784* were mainly distributed in *japonica* rice subgroups and high latitudes, such as Northeast China (NEC). Similarly, the favorable haplotypes of the three candidate genes were mainly distributed in *indica* rice, and Hap A of *LOC_Os04g47890* was mainly distributed in SC, SWC and SEA. Therefore, germplasms with better PLs, more grains and heavier panicles were mainly distributed in subpopulations of *indica* rice and low-latitude regions. We further analyzed the frequencies of the favorable gcHap of 3 important candidate genes affecting two panicle traits (PL and GW) in Xian/*indica* (XI) and Geng/*japonica* (GJ) modern varieties and different rice subpopulations and the results showed that the GW candidate genes were mainly distributed in *japonica* rice, and the favorable haplotypes of the two PL candidate genes in modern varieties were mainly distributed in *indica* rice. In addition, the favorable haplotypes of 3KRG candidate genes were distributed in different subgroups (Supplementary Table S8). Compared with the three haplotypes carrying the *LOC_Os01g43700* gene from 3,000 rice genome databases, the favorable haplotype carrying the A allele in this study was basically the same with Hap1 (Supplementary Table S8). And compared with the three haplotypes of *LOC_Os09g25784* gene from 3,000 rice genome databases, the favorable haplotypes carrying the B, C, and D alleles had basically the same panicle length with the favorable haplotypes in 3,000 rice genome database (Supplementary Table S8).

3.9 Excellent parental combinations predicted for panicle traits

Through further research, we found that the haplotype analysis results of the three genes showed that six haplotypes showed positive effects and the other three haplotypes showed negative effects (Supplementary Table S7). We predicted 10 excellent parents for panicle trait improvement (Table 4). Among the ten predicted excellent parents, nine varieties belonged to *indica* rice, and only one belonged to *japonica* rice. The favorable haplotypes of *LOC_Os01g43700*, *LOC_Os09g25784*, and *LOC_Os04g47890* improved panicle traits. To further verify this hypothesis, we used the *indica* rice variety Yuedao 61 as an example and found that the GWP of the dominant haplotype Hap A containing these three genes could theoretically increase by 0.72 g. Similarly, we also predicted other varieties.

4 Discussion

Panicle traits are complex quantitative traits regulated by multiple genes and controlled by the environment (Bai et al., 2021). Here, we observed and collated the phenotypic data of five panicle traits of 162 rice materials, and identified abundant phenotypic variations. Although only 162 rice germplasm resources were used in this study, these germplasm resources have a wide range of sources, including germplasm resources from China and four other Asian countries. Furthermore, the

162 materials in this study have abundant genetic and phenotypic variations, which are suitable for association analysis. The phenotypic variation of PL was high, and the coefficient of variation (CV) was between 20.9% and 21.2%, which was similar to other studies. For instance, the PL phenotype of 340 rice materials from 3,000 rice genome projects was identified, and the coefficient of variation was between 15% and 16% (Bai et al., 2021). Liu et al. used three mapping populations for PL phenotypic identification and found that the CVs ranged from 15.3% to 19.6% (Liu et al., 2016). In addition, we calculated the Pearson correlation coefficient, and the results showed that among 162 rice materials, TGP, FGP and GWP had similar positive correlation trends (Figure 3), which is consistent with the GWAS results; that is, some QTLs can be detected by two models within 2 years or related to multiple traits (Supplementary Table S2).

Genome-wide association analysis (GWAS) is a powerful method to understand the genetic structure of multiple traits in rice. Generally, regression models are constructed to test whether there is a correlation between markers and phenotypes. Previous studies have shown that MLMs control false positives better than GLMs do, which was strongly reflected by experimental results. Therefore, we used the MLM and GLM for further analysis. In this study, a total of 14 QTLs related to panicle traits were detected, and 9 QTLs were detected in at least 2 years and via two models. The local LD regions of the four QTLs associated with panicle traits overlapped with the flanking regions of the two genes (*OsBZR1* and *sd1*) among the 9 stable QTLs and one QTL (*LPI*) reported previously (Ye et al., 2015; Liu et al., 2016; Qiao et al., 2017). The remaining 6 QTLs were all newly detected (Table 2; Supplementary Figure S1), and the results of the genomic BLUP method were completely consistent (Figure 6), indicating that our results were highly reliable. *LPI* colocalized with *qPL9* and *PL9.1*, which was delimited to a 90 kb region for the first time (Liu et al., 2016). *OsBZR1* colocalized with *qFGP7*, which is involved in plant growth and development. Overexpression of *OsBZR1* results in higher sugar accumulation in developing anthers and seeds and increased seed yield (Qiao et al., 2017). *Sd1*, near *qSSR1*, encodes a gibberellin 20 oxidase enzyme participating in gibberellic acid biosynthesis and is involved in the control of the grain number and SSR of rice (Su et al., 2021). The QTL *qFGP7* was near the cloned gene *GE*, which encodes a CYP78A subfamily P450 monooxygenase (Yang et al., 2013). Therefore, a total of 6 QTLs are novel QTLs reported in our study (Figure 11).

Haplotype analysis of *LOC_Os01g43700* in the significant locus qPL1 revealed two haplotypes (Figure 7B). In the whole population, haplotypes affected the average PL, and HapA guided greater PL (Figure 7C), indicating that *LOC_Os01g43700* may affect PL variation. The difference in haplotype frequency between the two subspecies indicated that the gene *LOC_Os01g43700* was selected in the rice breeding process, and HapA was contained in 70% of *indica* rice materials (Figure 10). In addition, the gene *LOC_Os01g43700* encodes a cytochrome P450 protein. Previous studies have shown that cytochrome P450 plays an important role in plant growth and development (Pandian et al., 2020; Distefano et al., 2021) and is involved in biosynthetic pathways

and detoxification pathways (Xu et al., 2015). The results demonstrated that the gene *LOC_Os01g43700* may be a candidate gene for PL variation. *qPL9* (*LOC_Os09g25784*), which was found to be associated with PL, was detected in all 3 years in this study and this gene was selected as the focus for further analysis. *LOC_Os09g25784* encodes the auxin-induced protein 5NG4, and auxin is an essential molecule that controls almost every aspect of plant development (Paque and Weijers, 2016). Analysis of *LOC_Os09g25784* revealed four haplotypes (Figure 8B). Significant differences in mean PL were found between accessions with different haplotypes throughout the whole panel, with HapB, HapC, and HapD governing the longer panicles, indicating that the gene *LOC_Os09g25784* may be responsible for PL variation. *qTGP4*, *qFGP4*, and *qGWP4* were in the same QTL, and were found to be associated with TGP, FGP, and GWP in all 3 years. Based on the results of functional annotation and haplotype analysis of candidate genes, we proposed that *LOC_Os04g47890* (which encodes a MYB family transcription factor) is a candidate gene in *qTGP4*, *qFGP4*, and *qGWP4*. The MYB transcription factor family is present in all eukaryotes. Compared with other organisms, plants encode many MYB genes (Kranz et al., 2000; Qu and Zhu, 2006; Yanhui et al., 2006). MYB proteins compose a superfamily of transcription factors that play regulatory roles in developmental processes in plants (Yanhui et al., 2006; Wang et al., 2020). Consequently, *LOC_Os04g47890* may be a candidate gene for regulating rice panicle traits.

The results showed that favorable haplotypes could improve the panicle traits of rice. Among the 10 predicted parents, 9 *indica* rice improved panicle traits better than *indica* rice, and these results need to be further verified in production. Overall, further functional analysis of these candidate genes will help us better understand the genetic basis for natural variation in rice panicle traits and show that these genes may be used to improve rice yield. With the development of molecular biology, our research on the mining of excellent alleles of rice panicle traits and the prediction of excellent parents will provide a theoretical basis for molecular design breeding. Therefore, these candidate genes are to be used in breeding programme not only the breeding program discussed here but also other breeding programs.

5 Conclusion

In our study, we dissected the genetic basis underlying rice panicle traits. According to the results of gene function annotation and haplotype analysis, we identified three candidate genes related to panicle traits. *LOC_Os01g43700* and *LOC_Os09g25784* were identified as candidate genes for PL, while *LOC_Os04g47890* was identified as a candidate gene for TGP, FGP, and GWP. These genes are important for further research. At the same time, we infer that these 10 rice materials can be used as excellent parents with favorable alleles of panicle trait genes. Using the favorable alleles detected in this study can provide a theoretical basis for the improvement and breeding of panicle traits.

Data availability statement

The data presented in the study are deposited in the NCBI Sequence Read Archive repository, accession number PRJNA554986.

Author contributions

JX, XL, YC, JS, XC, MX, and ZZ conducted the experiments and collected the data; HS and YZ collated and statistically analyzed the data; ZL and MD drawing Graphics; ZL wrote the paper; EL participated in the design of the experiment provided intellectual guidance and checked the paper. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (grant numbers 32101768 and U21A20214); Natural Science Foundation of Anhui Province (grant number 2108085MC97); Natural Science Foundation of Anhui Universities (grant number KJ 2020A0118); Talent Project of Anhui Agricultural University (grant number rc312002); Natural Science Research Project of Colleges and Universities in Anhui Province (grant number YJS20210250); and Key Research and Development Program of Anhui Province (grant number 202004a06020024).

Acknowledgments

Thank Hong Delin of the State Key Laboratory of Crop Genetics and Germplasm Innovation of Nanjing Agricultural University for the experimental materials provided for this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1239550/full#supplementary-material>

References

- Agata, A., Ando, K., Ota, S., Kojima, M., Hobo, T., Takehara, S. (2020). Diverse panicle architecture results from various combinations of *prl5/ga20ox4* and *pbl6/apo1* alleles. *Commun. Biol.* 3 (1), 302. doi:10.1038/s42003-020-1036-8
- Bai, S., Hong, J., Li, L., Su, S., Li, Z., Wang, W. (2021). Dissection of the genetic basis of rice panicle architecture using a genome-wide association study. *Rice* 14 (1), 77. doi:10.1186/s12284-021-00520-w
- Bai, X. F., Zhao, H., Huang, Y., Xie, W. B., Han, Z. M., Zhang, B. (2016). Genome-wide association analysis reveals different genetic control in panicle architecture between *indica* and *japonica* rice. *Plant Genome* 9 (2). doi:10.3835/plantgenome2015.11.0115
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). Tassel: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23 (19), 2633–2635. doi:10.1093/bioinformatics/btm308
- Crowell, S., Korniliev, P., Falcao, A., Ismail, A., Gregorio, G., Mezey, J. (2016). Genome-wide association and high-resolution phenotyping link *oryza sativa* panicle traits to numerous trait-specific qtl clusters. *Nat. Commun.* 7, 10527. doi:10.1038/ncomms10527
- Distefano, A. M., Setzes, N., Cascallares, M. M., Fiol, D. F., Zabaleta, E. J., and Pagnussat, G. C. (2021). Roles of cytochromes p450 in plant reproductive development. *Int. J. Dev. Biol.* 65 (4-5-6), 187–194. doi:10.1387/ijdb.200100gp
- Du, M. Y., Xiong, M. Y., Chang, Y. P., Liu, Z. B., Wang, R., Lin, X. X. (2022). Mining candidate genes and favorable haplotypes for flag leaf shape in rice (*oryza sativa* L.) based on a genome-wide association study. *Agronomy* 12 (8), 1814. doi:10.3390/agronomy12081814
- Earl, D. A., and Vonholdt, B. M. (2012). Structure harvester: a website and program for visualizing structure output and implementing the evanno method. *Conserv. Genet. Resour.* 4 (2), 359–361. doi:10.1007/s12686-011-9548-7
- Evanno, G. S., Regnaut, S. J., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14 (8), 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
- Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7 (4), 574–578. doi:10.1111/j.1471-8286.2007.01758.x
- Franois, O., and Durand, E. (2010). Spatially explicit bayesian clustering models in population genetics. *Mol. Ecol. Resour.* 10 (5), 773–784. doi:10.1111/j.1755-0998.2010.02868.x
- Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic blup decoded: a look into the black box of genomic prediction. *Genetics* 194, 597–607. doi:10.1534/genetics.113.152207
- Han, K., Jeong, H. J., Yang, H. B., Kang, S. M., Kwon, J. K., Kim, S. (2016). An ultra-high-density bin map facilitates high-throughput QTL mapping of horticultural traits in pepper (*Capsicum annuum*). *DNA Res.* 23 (2), 81–91. doi:10.1093/dnares/dsv038
- Huang, R., Jiang, L., Zheng, J., Wang, T., Hong, Z., Huang, Y. (2012). Genetic bases of rice grain shape: so many genes, so little known. *Trends Plant Sci.* 18 (4), 218–226. doi:10.1016/j.tplants.2012.11.001
- Huang, X., Kurata, N., Wei, X., Han, B., Wang, A., Zhao, Q. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490 (7421), 497–501. doi:10.1038/nature11532
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42 (11), 961–967. doi:10.1038/ng.695
- Jin, H., and Martin, C. (1999). Multifunctionality and diversity within the plant myb-gene family. *Plant Mol. Biol.* 41, 577–585. doi:10.1023/a:1006319732410
- Kobayashi, S., Fukuta, Y., Sato, T., Osaki, M., and Khush, G. S. (2003). Molecular marker dissection of rice (*Oryza sativa* L.) plant architecture under temperate and tropical climates. *Theor. Appl. Genet.* 107 (8), 1350–1356. doi:10.1007/s00122-003-1388-8
- Kranz, H., Scholz, K., and Weisshaar, B. (2000). C-myb oncogene-like genes encoding three myb repeats occur in all major plant lineages. *Plant J.* 21 (2), 231–235. doi:10.1046/j.1365-3113x.2000.00666.x
- Li, M. X., Yeung, J. M., Cherny, S. S., and Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131 (5), 747–756. doi:10.1007/s00439-011-1118-2
- Li, S., Li, W., Huang, B., Cao, X., Zhou, X., Ye, S. (2013). Natural variation in *PTB1* regulates rice seed setting rate by controlling pollen tube growth. *Nat. Commun.* 4, 2793. doi:10.1038/ncomms3793
- Liu, C., Xue, Z., Tang, D., Shen, Y., Shi, W., Ren, L. (2018). Ornithine delta-aminotransferase is critical for floret development and seed setting through mediating nitrogen reutilization in rice. *Plant J.* 96 (4), 842–854. doi:10.1111/tpj.14072
- Liu, E., Liu, Y., Wu, G., Zeng, S., Tran Thi, T. G., Liang, L. (2016). Identification of a candidate gene for panicle length in rice (*oryza sativa* L.) via association and linkage analysis. *Front. Plant Sci.* 7, 596. doi:10.3389/fpls.2016.00596
- Loos, J., Abson, D. J., Chappell, M. J., Hanspach, J., Mikulcak, F., Tichit, M. (2014). Putting meaning back into “sustainable intensification”. *Front. Ecol. Environ.* 12 (6), 356–361. doi:10.1890/130157
- Lu, G., Coneva, V., Casaretto, J. A., Ying, S., Mahmood, K., Liu, F. (2015). *Osp15b* modulates rice (*oryza sativa*) plant architecture and yield by changing auxin homeostasis, transport and distribution. *Plant J.* 83 (5), 913–925. doi:10.1111/tpj.12939
- Mei, H. W., Li, Z. K., Shu, Q. Y., Guo, L. B., Wang, Y. P., Yu, X. Q. (2005). Gene actions of qtls affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. *Theor. Appl. Genet.* 110 (4), 649–659. doi:10.1007/s00122-004-1890-7
- Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V., and Jugulam, M. (2020). Role of cytochrome p450 enzymes in plant stress response. *Antioxidants (Basel)* 9 (5), 454. doi:10.3390/antiox9050454
- Paque, S., and Weijers, D. (2016). Q&A: auxin: the plant molecule that influences almost anything. *BMC Biol.* 14 (1), 67. doi:10.1186/s12915-016-0291-0
- Qiao, S., Sun, S., Wang, L., Wu, Z., Li, C., Li, X. (2017). The *rla1/smos1* transcription factor functions with *osbzr1* to regulate brassinosteroid signaling and rice architecture. *Plant Cell* 29 (2), 292–309. doi:10.1105/tpc.16.00611
- Qu, L. J., and Zhu, Y. X. (2006). Transcription factor families in arabidopsis: major progress and outstanding issues for future research. *Curr. Opin. Plant Biol.* 9 (5), 544–549. doi:10.1016/j.pbi.2006.07.005
- Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q. (2016). *OsSPL13* controls grain size in cultivated rice. *Nat. Genet.* 48 (4), 447–456. doi:10.1038/ng.3518
- Song, X. J., Huang, W., Shi, M., Zhu, M. Z., and Lin, H. X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39 (5), 623–630. doi:10.1038/ng2014
- Su, J., Pang, C., Wei, H., Li, L., Liang, B., Wang, C. (2016). Identification of favorable snp alleles and candidate genes for traits related to early maturity via gwas in upland cotton. *BMC Genomics* 17, 687. doi:10.1186/s12864-016-2875-z
- Su, S., Hong, J., Chen, X., Zhang, C., Chen, M., Luo, Z. (2021). Gibberellins orchestrate panicle architecture mediated by *DELLA-KNOX* signalling in rice. *Plant Biotechnol. J.* 19 (11), 2304–2318. doi:10.1111/pbi.13661
- Thapa, R., Tabien, R. E., and Septiningsih, E. M. (2021). Genome-wide association study to identify chromosomal regions related to panicle architecture in rice (*Oryza sativa*). *Genet. Resour. Crop Evol.* 68 (7), 2849–2865. doi:10.1007/s10722-021-01159-8
- Thomson, M. J., Tai, T. H., Mcclung, A. M., Lai, X. H., Hinga, M. E., Lobos, K. B. (2003). Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *oryza rufipogon* and the *oryza sativa* cultivar jefferson. *Theor. Appl. Genet.* 107 (3), 479–493. doi:10.1007/s00122-003-1270-8
- Tu, B., Tao, Z., Wang, S., Zhou, L., Zheng, L., Zhang, C. (2022). Loss of *Gn1a/OsCKX2* confers heavy-panicle rice with excellent lodging resistance. *J. Integr. Plant Biol.* 64 (1), 23–38. doi:10.1111/jipb.13185
- Wang, F., Han, T., Song, Q., Ye, W., Song, X., Chu, J. (2020). The rice circadian clock regulates tiller growth and panicle development through strigolactone signaling and sugar sensing. *Plant Cell* 32 (10), 3124–3138. doi:10.1105/tpc.20.00289
- Wang, K., Li, M., and Hakonarson, H. (2010). Annovar: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38 (16), e164. doi:10.1093/nar/gkq603
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Leung, H., Wu, Z. (2018). Genomic variation in 3,010 diverse accessions of asian cultivated rice. *Nature* 557 (7703), 43–49. doi:10.1038/s41586-018-0063-9
- Wang, X., Liu, G., Wang, Z., Chen, S., Xiao, Y., and Yu, C. (2019). Marine natural products in the discovery and development of potential pancreatic cancer therapeutics. *Plant Breed.* 138 (3), 299–314. doi:10.1016/bs.acr.2019.05.003
- Xiang, X., Zhang, P., Yu, P., Zhang, Y., Yang, Z., Sun, L. (2019). *LSSR1* facilitates seed setting rate by promoting fertilization in rice. *Rice (N Y)* 12 (1), 31. doi:10.1186/s12284-019-0280-3
- Xing, Y., and Zhang, Q. (2010). Genetic and molecular bases of rice yield. *Annu. Rev. Plant Biol.* 61, 421–442. doi:10.1146/annurev-arplant-042809-112209
- Xing, Z., Tan, F., Hua, P., Sun, L., Xu, G., and Zhang, Q. (2002). Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. *Theor. Appl. Genet.* 105 (2-3), 248–257. doi:10.1007/s00122-002-0952-y
- Xu, J., Wang, X. Y., and Guo, W. Z. (2015). The cytochrome P450 superfamily: key players in plant development and defense. *J. Integr. Agric.* 14 (9), 1673–1686. doi:10.1016/s2095-3119(14)60980-1
- Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L. (2008). Natural variation in *ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* 40 (6), 761–767. doi:10.1038/ng.143

- Yang, W., Gao, M., Yin, X., Liu, J., Xu, Y., Zeng, L. (2013). Control of rice embryo development, shoot apical meristem maintenance, and grain yield by a novel cytochrome p450. *Mol. Plant* 6 (6), 1945–1960. doi:10.1093/mp/sst107
- Yanhui, C., Kun, H., Jigang, L., Zhiqiang, L., Xiaoxiao, W., Yunping, S. (2006). The myb transcription factor superfamily of arabidopsis: expression analysis and phylogenetic comparison with the rice myb family. *Plant Mol. Biol.* 60 (1), 107–124. doi:10.1007/s11103-005-2910-y
- Ye, H., Feng, J., Zhang, L., Zhang, J., Mispan, M. S., Cao, Z. (2015). Map-based cloning of seed dormancy1-2 identified a gibberellin synthesis gene regulating the development of endosperm-imposed dormancy in rice. *Plant Physiol.* 169 (3), 2152–2165. doi:10.1104/pp.15.01202
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38 (2), 203–208. doi:10.1038/ng1702
- Zhang, B., Ye, W., Ren, D., Tian, P., Peng, Y., Gao, Y. (2015). Genetic analysis of flag leaf size and candidate genes determination of a major QTL for flag leaf width in rice. *Rice (N Y)* 8 (1), 39. doi:10.1186/s12284-014-0039-9
- Zhang, H., Zhu, S., Liu, T., Wang, C., Cheng, Z., Zhang, X. (2019). DELAYED HEADING DATE1 interacts with OsHAP5C/D, delays flowering time and enhances yield in rice. *Plant Biotechnol. J.* 17 (2), 531–539. doi:10.1111/pbi.12996
- Zhang, L., Wang, J., Wang, J., Wang, L., Ma, B., Zeng, L. (2015). Quantitative trait locus analysis and fine mapping of the qpl6 locus for panicle length in rice. *Theor. Appl. Genet.* 128 (6), 1151–1161. doi:10.1007/s00122-015-2496-y
- Zhang, L., Zou, Y., Bian, Z., Xie, D., Yeilaghi, H., Fan, X. (2021). Fine mapping and candidate gene prediction of the quantitative trait locus qpl8 for panicle length in rice. *Phyt. J. Exp. Bot.* 003 (090), 789–802. doi:10.32604/phyton.2021.014880
- Zhang, Y. F., Ma, Y., Chen, Z., Zhou, J., Chen, T., Li, Q. (2015). Genome-wide association studies reveal new genetic targets for five panicle traits of international rice varieties. *Rice Sci.* 22 (5), 217–226. doi:10.1016/j.rsci.2015.07.001
- Zhang, Y., Zhu, Y., Peng, Y., Yan, D., Li, Q., Wang, J. (2008). Gibberellin homeostasis and plant height control by EUI and a role for gibberellin in root gravity responses in rice. *Cell Res.* 18 (3), 412–421. doi:10.1038/cr.2008.28
- Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS Genet.* 3 (1), e4. doi:10.1371/journal.pgen.0030004
- Zhong, H., Liu, S., Meng, X., Sun, T., Deng, Y., Kong, W. (2021). Uncovering the genetic mechanisms regulating panicle architecture in rice with GPWAS and GWAS. *BMC Genomics* 22 (1), 86. doi:10.1186/s12864-021-07391-x