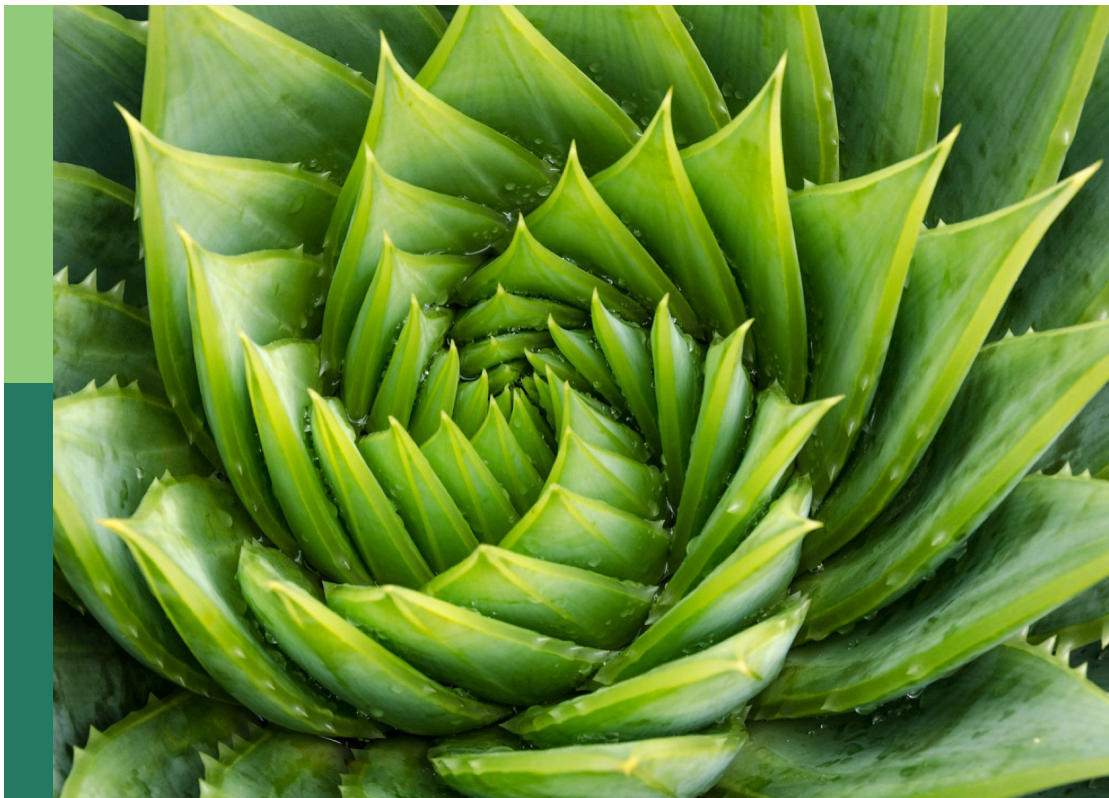
frontiers | Research Topics

Genetics, genomics and breeding of plant architecture, biomass, grain quality and grain yield traits in rice and wheat

Edited by
Yue Feng, Yong Zhou, Rongrong Zhai, Gaoneng Shao, Reyazul Rouf Mir and Shouvik Das

Coordinated by
Kai Wang and Yan He

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-4521-8
DOI 10.3389/978-2-8325-4521-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Genetics, genomics and breeding of plant architecture, biomass, grain quality and grain yield traits in rice and wheat

Topic editors

Yue Feng — China National Rice Research Institute, Chinese Academy of Agricultural Sciences, China
Yong Zhou — Yangzhou University, China
Rongrong Zhai — Zhejiang Academy of Agricultural Sciences, China
Gaoneng Shao — Zhejiang University, China
Reyazul Rouf Mir — Sher-e-Kashmir University of Agricultural Sciences and Technology, India
Shouvik Das — Regional Centre for Biotechnology (RCB), India

Topic Coordinators

Kai Wang — Yuan Longping High-tech Agriculture Co., Ltd., China
Yan He — Seoul National University, Republic of Korea

Citation

Feng, Y., Zhou, Y., Zhai, R., Shao, G., Mir, R. R., Das, S., Wang, K., He, Y., eds. (2024). *Genetics, genomics and breeding of plant architecture, biomass, grain quality and grain yield traits in rice and wheat*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-4521-8

Table of contents

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fpls.2022.1098560

OPEN ACCESS

EDITED BY
Yong Zhou,
Yangzhou University, China

REVIEWED BY
Jindong Liu,
Chinese Academy of Agricultural Sciences, China
Caixia Lan,
Huazhong Agricultural University, China

*CORRESPONDENCE
Hui Li
zwslihui@163.com
Yingjun Zhang
zhangyingjun1977@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 15 November 2022
ACCEPTED 13 December 2022
PUBLISHED 04 January 2023

CITATION
Zhao J, Sun L, Gao H, Hu M, Mu L, Cheng X, Wang J, Zhao Y, Li Q, Wang P, Li H and Zhang Y (2023) Genome-wide association study of yield-related traits in common wheat (*Triticum aestivum* L.) under normal and drought treatment conditions. *Front. Plant Sci.* 13:1098560. doi: 10.3389/fpls.2022.1098560

Genome-wide association study of yield-related traits in common wheat (*Triticum aestivum* L.) under normal and drought treatment conditions

Jie Zhao[1†], Lijing Sun[1†], Huimin Gao[2†], Mengyun Hu[1], Liming Mu[3], Xiaohu Cheng[3], Jianbing Wang[3], Yun Zhao[1], Qianying Li[1], Peinan Wang[1], Hui Li[1*] and Yingjun Zhang[1*]

[1]Institute of Cereal and Oil Crops, Laboratory of Crop Genetics and Breeding of Hebei, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, China, [2]Institute of Cash Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, China, [3]Institute of Cereal Crops, Dingxi Academy of Agricultural Sciences, Dingxi, China

The primary goal of modern wheat breeding is to develop new high-yielding and widely adaptable varieties. We analyzed four yield-related agronomic traits in 502 wheat accessions under normal conditions (NC) and drought treatment (DT) conditions over three years. The genome-wide association analysis identified 51 yield-related and nine drought-resistance-related QTL, including 13 for the thousand-grain weight (TGW), 30 for grain length (GL), three for grain width (GW), five for spike length (SL) and nine for stress tolerance index (STI) QTL in wheat. These QTL, containing 72 single nucleotide polymorphisms (SNPs), explained 2.23 – 7.35% of the phenotypic variation across multiple environments. Eight stable SNPs on chromosomes 2A, 2D, 3B, 4A, 5B, 5D, and 7D were associated with phenotypic stability under NC and DT conditions. Two of these stable SNPs had association with TGW and STI. Several novel QTL for TGW, GL and SL were identified on different chromosomes. Three linked SNPs were transformed into kompetitive allele-specific PCR (KASP) markers. These results will facilitate the discovery of promising SNPs for yield-related traits and/or drought stress tolerance and will accelerate the development of new wheat varieties with desirable alleles.

KEYWORDS

common wheat, yield, GWAS, drought, KASP markers

Introduction

Wheat (*Triticum aestivum* L.), one of the most widely cultivated and most important cereal crops, provides approximately 20% of the dietary calories and proteins for human nutrition (Wang et al., 2021; Li et al., 2022a). However, climate stress and depleting fresh water for agricultural irrigation have severely affected wheat production, and drought stress is a major threat to wheat yield (Ahmed et al., 2022). Wheat is particularly susceptible to drought-induced stress throughout the growth period, and therefore, mining drought-stable QTL is vital for increasing wheat yield.

Wheat yield is a complex trait affected by grain and spike agronomic characteristics, including thousand-grain weight (TGW), grain length (GL), grain width (GW), and spike length (SL) (Li et al., 2019). In the present years, numerous QTL for wheat yield were mapped on 21 wheat chromosomes (Tshikunde et al., 2019; Cao et al., 2020). Several genes associated with wheat yield have been cloned in rice or *Arabidopsis*. These homologous genes include *TaCwi-A1* (Ma et al., 2012), *TaCKX6-D1* (Zhang et al., 2012), *TaSAP1* (Chang et al., 2013), *TaGASR7-A1* (Dong et al., 2014), *TaGW2A* (Hong et al., 2014), *TaGS-D1* (Zhang et al., 2014b), *TaCWI-4A* (Jiang et al., 2015), *TaCWI-5D* (Jiang et al., 2015), *TaCYP78A3* (Ma et al., 2015), *TaTGW6-A1* (Hanif et al., 2016), *TaTGW-7A* (Hu et al., 2016), *TaGS5-3A* (Ma et al., 2016), *TaGS5-A1* (Wang et al., 2016), *TaFlo2-A1* (Sajjad et al., 2017), *TaBT1* (Wang et al., 2019), *TaDA1* (Liu et al., 2020), *TaPRR1* (Sun et al., 2020), *TaSPL14* (Cao et al., 2021), *TaPGS1* (Guo et al., 2022), *TaIAA21* (Jia et al., 2021), and *TaPIN1s* (Yao et al., 2021). Many yield-related QTL/genes have been detected, but limited research has investigated their consistency in stressful and optimal environments. Thus, elucidating the genetic basis of grain and spike morphometric traits under diverse conditions is crucial for improving wheat yield and yield stability.

The rapid development of sequencing and high-density marker genotyping technologies has made GWAS (genome-wide association study) an effective method of identifying the quantitative variation of complex crop traits (Nordborg and Weigel, 2008; Hayes, 2013). The GWAS technique has high resolution and accuracy and has been widely used to study agronomic traits (Pang et al., 2020) and biotic and abiotic resistance (Kang et al., 2020; Shan et al., 2022) in wheat. In addition, GWAS performs genome-scale sequencing of numerous cultivars with diverse genetic backgrounds, thus, accelerating the genetic dissection of complex traits in crops (Juliana et al., 2022; Shan et al., 2022).

In this study, 502 Chinese wheat accessions were genotyped using the wheat 15K SNP array. The TGW, GL, GW, SL and STI (stress tolerance index) phenotypes were analyzed under normal (NC) and drought treatment (DT) conditions. Next, GWAS was performed using multi-year experimental data to identify high confidence loci associated with yield-related traits in a hexaploidy wheat genome under the two conditions. The findings will improve the understanding of the genetic mechanism of wheat yield-related and drought resistance traits and provide new genetic loci for breeding new high-yielding and stress-resistant wheat varieties.

Materials and methods

Materials and field experiment

Five hundred and two hexaploidy wheat collections in China were used in this experiment (Supplementary Table 1). The wheat collections were mainly from the Winter Wheat Region of the Yellow and Huai Valleys, which accounted for approximately 75% of the wheat produced in China. This study was established in the Dishang Experimental Station of the Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, Hebei Province, China, located at 38°N and 114°83′E. The field experiments were conducted over three consecutive years (2018 – 2019, 2019 – 2020, 2020 – 2021). The average rainfalls in the October – May season of 2018 – 2019, 2019 – 2020, and 2020 – 2021 were 101, 131, and 75 mm, respectively (Supplementary Figure 1). Plants under normal conditions (NC) were irrigated twice at the jointing and booting stages in 2018 – 2019 (NC19), 2019 – 2020 (NC20), and 2020 – 2021 (NC21). In contrast, drought-treated plants were not irrigated throughout the growth period in 2018 – 2019 (DT19), 2019 – 2020 (DT20), and 2020 – 2021 (DT21). All plants were grown in rows 3 by 0.22 m in length and width at a sowing rate of 84 seeds per row in randomized complete blocks. Trials were fertilized and maintained free from weeds, insects, and diseases.

Phenotyping and statistical analysis

Five randomly selected spikes were sampled for manual measurement, and then to generate one average value. The grain traits (TGW, GL, and GW) were recorded using a scaled-camera-assisted phenotyping system (Wanshen Detection Technology Co., Ltd., Hangzhou, China). The stress tolerance index of each accession was calculated using the formula: STI = (TGW.DT × TGW.NC)/($X.NC^2$), where TGW.NC and TGW.DT were the TGW for each accession under normal and drought treatment conditions, respectively; and X.NC was the average TGW of all accessions under normal conditions (Fernandez, 1992).

All statistical analysis was conducted in R v4.1.0 (Yu et al., 2019) and SPSS v16.0 software (SPSS Inc, Chicago). Pearson's correlation coefficient was determined for all traits in the

different environments. The formula $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{gy}^2/y + \sigma_{er}^2/yr)$ was used to calculate broad sense heritability (h^2), where σ_g^2, σ_{gy}^2, and σ_{er}^2 represent the genotypic variance, genotype by year effect, and the residual error, respectively. y and r are the number of years and replications, respectively.

SNP genotyping

The wheat iSelect 15K SNP array from China Golden Maker Biotechnology Co., Ltd., Beijing (http://www.cgmb.com.cn/) was used for genotyping the 502 accessions. The array was identified with specific genetic positions based on the consensus map (Li et al., 2021). The 13,947 SNPs were filtered using PLINK v1.9 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al., 2007) to delete loci with minor allele frequency (MAF) (< 5%) and low-quality SNPs (with > 20% missing data). After filtering, 13,705 SNPs remained and were used for GWAS. The physical maps of all polymorphic SNPs were obtained by searching the reference sequence IWGSC RefSeq v1.0 2018 using URGI BLAST (http://wheat-urgi.versailles.inra.fr/Tools/BLAST).

Population structure, kinship, and linkage disequilibrium

The phylogenetic tree was generated as previously described (Zhao et al., 2020). The principal component analysis (PCA) was constructed using PLINK v1.9, and the percentage of variance explained by the top 10 PCs was plotted to determine the optimal number of PCs. Furthermore, the relative kinship matrix (K-matrix) between accessions was calculated using TASSEL v5.2.73 (Bradbury et al., 2007). Linkage disequilibrium (LD) was estimated as the squared allele frequency correlation (r^2) using PLINK v1.9. The LD decay distance was obtained by constructing a scatterplot of r^2 values against SNP pairs distance and fitting the points to a smooth curve using the R package "ggplot2".

Genome-wide association mapping

The mixed linear model (MLM) (PCA+K) was used for GWAS using the GAPIT package (Wang and Zhang, 2021) in R software. The first five principal components were included in the GWAS model. The mixed linear model accounts for false positives caused by the population structure and relative kinship. For this reason, the P values with cutoff set at 0.05 was highly restrictive. Considering the potential risk of type II error, and combining the GWAS results of the three years, the $-\log_{10}$ (P value) ≥ 3.0 were regarded as significant marker-trait associations (MTAs), as shown in Manhattan plots (Wang et al., 2021). Important P value distributions (observed P values plotted against expected P values) are shown in Q-Q plots.

KASP maker development

The SNP makers highly associated with grain and spike traits were selected and converted to KASP (kompetitive allele-specific PCR) markers. The KASP markers were used for PCR amplification of wheat DNA in 3 μL final volumes containing 40 –50 ng of genomic DNA, 0.75 μL of 2 × KASP Master Mix (LGC Genomics, product code: KBS-1050-112), and 0.0417 μL of assay primer mix (12 mM of each allele-specific primer and 30 mM of the common primer). The cycling conditions followed: 94°C for 15 min, ten cycles of 94°C for 20 s, touchdown starting at 61°C for 20 s (decreasing 0.6°C per cycle), 26 cycles of 94°C for 20 s, and 55°C for 1 min. The end-point fluorescence data were visualized with a microplate reader (Omega, BMG LABTECH, Germany) and analyzed by the Kluster-Caller software (LGC Genomics). The Student's t-tests were used to compare allelic effects of the phenotypic data and the effectiveness of the associated SNPs using KASP markers. The primers used in this research were listed in Supplementary Table 2.

Results

Phenotype assessments

The traits of 502 wheat accessions were characterized during three crop seasons (2018 – 2019, 2019 – 2020, and 2020 – 2021) under normal and drought treatment conditions (Figure 1, Supplementary Figure 2, and Table 1). The correlations were analyzed between the three crop seasons for TGW, GL, GW, and SL (Supplementary Figure 3). For all the detected phenotypes, positive correlations were found among different crop seasons under both normal and drought treatment conditions. Therefore, we compared the phenotypes under the different conditions based on the average values. All genotypes exhibited significant differences for each trait, with the coefficients of variation (CV) ranging from 3.83% to 16.39%. Except for GL, the normal and drought treatment conditions were significantly different ($P < 0.001$) for TGW, GW, and SL. The average TGW under the normal conditions was 48.62 g (ranging from 31.31 to 65.76 g), whereas the average TGW under the drought treatment conditions was 46.68 g (ranging from 32.40 to 63.12 g). Correspondingly, the average STI was 0.97 (ranging from 0.54 to 1.45). Traits GW and SL decreased from 3.62 to 3.55 mm and 8.22 to 7.75 cm under the DT conditions compared to the normal conditions. In contrast, the average GL differed slightly (0.03 cm) between the two

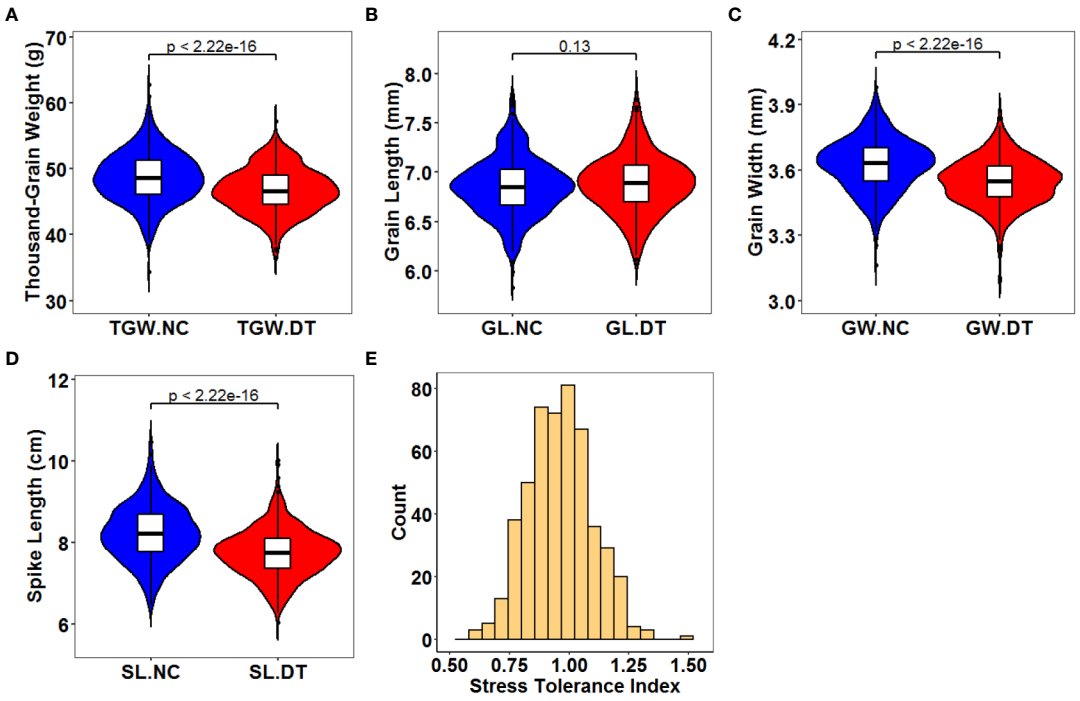


FIGURE 1
Distribution of the average phenotype for the 502 wheat accessions across all environments. **(A)** Thousand-grain weight under normal (TGW.NC) and drought treatment conditions (TGW.DT); **(B)** Grain length under normal (GL.NC) and drought treatment conditions (GL.DT); **(C)** Grain width under normal (GW.NC) and drought treatment conditions (GW.DT); **(D)** Spike length under normal (SL.NC) and drought treatment conditions (SL.DT); **(E)** Stress tolerance index. The center and edges of the boxplot represent the median and the 25th or 75th percentile respectively, while the whiskers represent the median ± 1.5 × IQR (interquartile interval).

conditions (Table 1). Therefore, drought treatment environmentally stressed the plants, reducing their spike and grain yield. All detected traits showed high broad sense heritability (h^2, 76.16 – 95.66%) under the two conditions.

We analyzed the correlations between the four traits (Figure 2). Under the normal conditions, any two of TGW, GL, and GW were positively correlated (r = 0.61*** between TGW and GL, r = 0.78*** between TGW and GW, and r =

TABLE 1 Statistics for thousand-grain weight (TGW), grain length (GL), grain width (GW), spike length (SL) and stress tolerance index (STI) in the association population under normal (NC) and drought treatment (DT) conditions.

Traits	Mean ± SD[a]	Range	CV (%)[b]	h^2 (%)[c]
TGW.NC	48.62 ± 5.25	31.31 – 65.76	10.79	84.01
GL.NC	6.86 ± 0.34	5.72 – 7.97	4.9	94.55
GW.NC	3.62 ± 0.16	3.06 – 4.18	4.34	87.02
SL.NC	8.22 ± 0.92	5.20 – 11.60	11.19	88.08
TGW.DT	46.68 ± 4.34	32.40 – 63.12	9.29	88.23
GL.DT	6.89 ± 0.33	5.79 – 8.05	4.73	95.66
GW.DT	3.55 ± 0.14	2.86 – 4.02	3.83	89.42
SL.DT	7.75 ± 1.00	5.30 – 11.70	12.84	76.16
STI	0.97 ± 0.16	0.54 – 1.45	16.39	91.84

[a]SD, standard deviations; [b]CV, coefficient of variation; [c]h^2, broad sense heritability.

0.35*** between GL and GW). Moreover, SL was positively correlated with TGW and GL (0.098* and 0.18***) respectively and uncorrelated with GW (Figure 2A). The correlations between TGW, GL, and GW were similar under NC and DT conditions. Additionally, SL was weakly correlated with GL (r = 0.085*) but uncorrelated with TGW and GW under DT (Figure 2B).

Statistical analysis of genotypes

All 502 accessions were genotyped using the wheat iSelect 15K array. After filtering (missing rate > 20% and minor allele frequency < 0.05), 13,705 high-quality SNPs were retained and used for genetic analysis. These SNPs were distributed on all 21 wheat chromosomes, with 4916, 5326, and 3463 SNPs in the A, B, and D sub-genomes. Chromosome 3B had the most SNPs (1024) but the shortest average distance between SNPs (833 kb). However, 4D had the least (264 SNPs and 2056 kb) (Supplementary Figure 4).

Population structure and LD analysis

We used different approaches to cluster the accessions into different groups and assessed the similarity of the results. The neighbor-joining tree divided the panel into three major groups (Figure 3A). Similarly, PCA classified the panel into three groups despite the numerous admixtures. These results revealed a weak population structure since the first three principal components collectively explained only approximately 19.31% of the total variance, i.e., 9.73, 5.37, and 4.21% for PC1, PC2, and PC3, respectively (Figure 3B). Overall, the above methods classify the populations into three subgroups.

The filtered SNP data were used to calculate LD. The average r^2 values for the A, B, and D sub-genomes and the whole genome gradually decreased with increasing pairwise distance (Figure 3C). When the cutoff threshold for r^2 was half of its maximum value, the A sub-genome had the longest LD decay distance (approximately 12 Mb) compared to the B and D sub-genomes (> 10 Mb). The LD decay distance for the entire genome was approximately 12 Mb.

GWAS of traits under two conditions

We performed GWAS for all the traits with the MLM in different environments under both conditions. We defined the significant and repetitive SNPs in at least two environments as true and reliable association loci. The purpose was to eliminate the influence of the environmental background and explain the true components of genetic variation. The Q-Q and Manhattan plots based on the average values of the individual traits are shown in Figures 4, 5. The GWAS identified 666 and 821 SNPs in different environments under NC and DT conditions, respectively, including 67 significant SNPs in 51 QTL from the multiple environments (Tables 2, 3 and Supplementary Table 3). In addition, 311 SNPs were detected for STI in different environments. Among them, 18 high confidence SNPs in nine QTL for STI were detected under more than one environment (Table 4 and Supplementary Table 4).

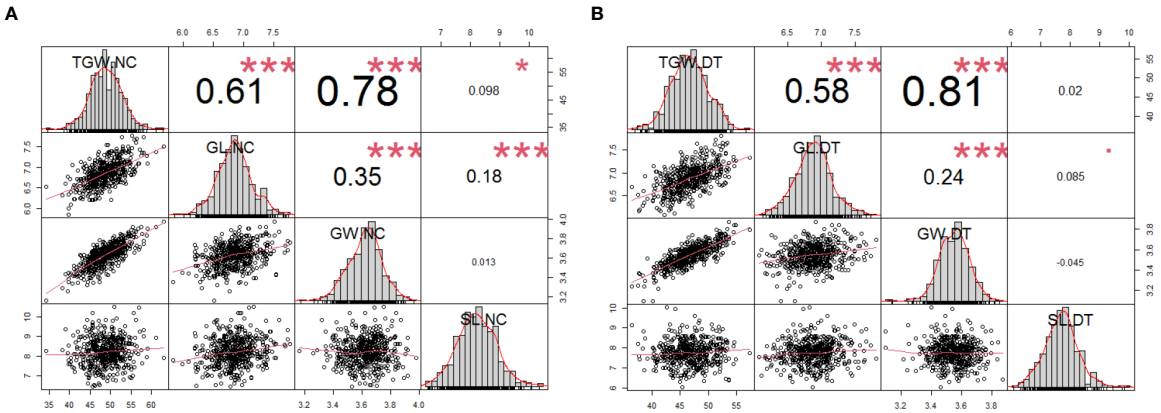


FIGURE 2
Phenotypic performances, distribution, and correlation coefficients for the 502 wheat accessions using the average phenotypic data under normal **(A)** and drought treatment conditions **(B)**. The frequency distribution of the phenotypic data for each trait is shown in the histograms. The *X-Y* scatter plot with the adjusted Pearson's coefficients. *P < 0.05, ***P < 0.001. TGW.NC, thousand-grain weight under normal conditions; GL.NC, grain length under normal conditions; GW.NC, grain width under normal conditions; SL.NC, spike length under normal conditions; TGW.DT, thousand-grain weight under drought treatment conditions; GL.DT, grain length under drought treatment conditions; GW.DT, grain width under drought treatment conditions; SL.DT, spike length under drought treatment conditions.

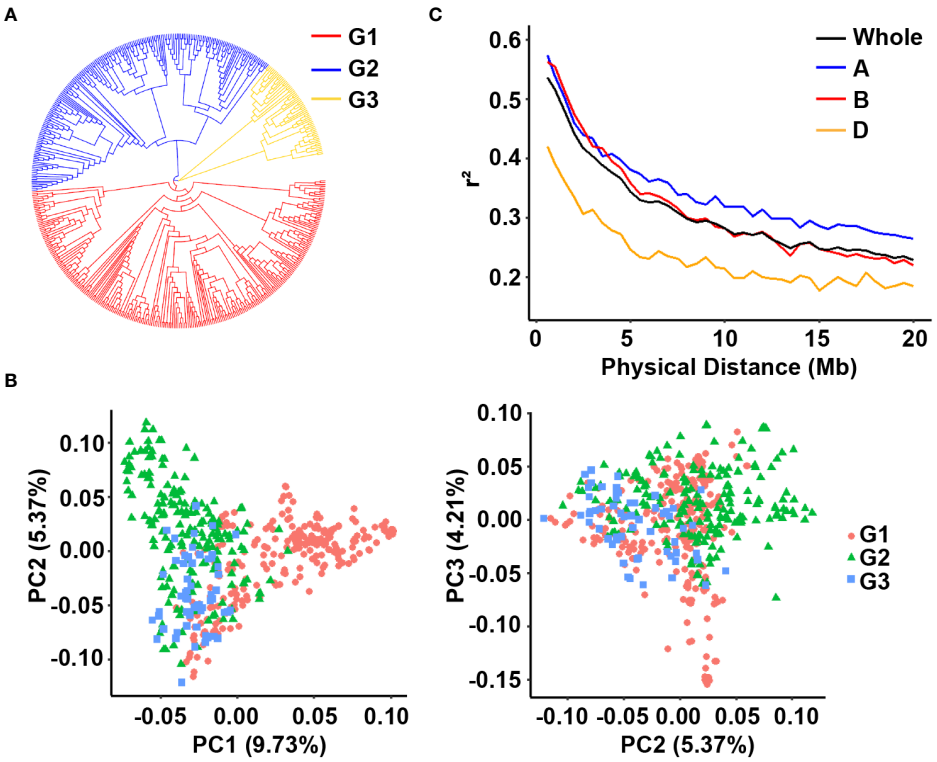


FIGURE 3
Analysis of the population structure of the 502 wheat accessions. **(A)** Neighbor-joining tree of all wheat accessions; **(B)** Principal component analysis showing the population structure in the diversity panel; **(C)** LD decay value in the whole genome, A sub-genome, B sub-genome and D sub-genome of wheat.

Thousand-grain weight

Eight QTL, represented by 18 SNPs, were detected on chromosomes 1B, 2A, 2B, 4A, 7B, and 7D for TGW under normal conditions (Figure 4A, Supplementary Figures 5A–C, Table 2, and Supplementary Table 3). One important QTL, *qNTGW-2B*, had five significant SNPs in multiple environments. *qNTGW-2B* was mapped to the physical region of 666.47 – 671.74 Mb on chromosome 2B, accounting for 2.84 – 3.40% of the phenotypic variation. Two QTL associated with TGW were detected on chromosome 4A. The first, *qNTGW-4A.1*, had four SNPs mapped to the physical position of 605.75 – 605.78 Mb and explained 2.54 – 3.44% of the phenotypic variation. The second *qNTGW-4A.2*, located at position of 633.30 – 635.76 Mb, was also represented by four SNPs and explained approximately 2.72 – 3.03% of the phenotypic variation. Besides, chromosome 7D also had two QTL (*qNTGW-7D.1* and *qNTGW-7D.2*), marked by SNPs *AX-108758043* and *AX-108770812* and explaining 2.96 and 3.23% of the phenotypic variation, respectively. In addition, three QTL, *qNTGW-1B* (430.15 Mb), *qNTGW-2A* (608.86 Mb), and *qNTGW-7B* (701.33 Mb), were represented by SNPs *AX-111169510*, *AX-109290429*, and *AX-89432708*, respectively, and collectively explained 14.70% of the phenotypic variation. Under drought treatment conditions, five QTL were associated with TGW (Figure 4B, Supplementary Figures 5D–F, Table 3, and Supplementary Table 3). Among these, *qDTGW-5A* physically was mapped to the physical region of 546.52 – 547.33 Mb on chromosome 5A. This QTL was represented by two SNPs and explained 2.33 – 3.72% of the phenotypic variation under drought treatment conditions. Another QTL, *qDTGW-5D*, at 292.94 Mb on chromosome 5D and *qDTGW-6A* at 328.13 Mb on chromosome 6A explained 2.44 and 3.08% of the phenotypic variation, respectively. Furthermore, two QTL represented by SNPs *AX-109375057* and *AX-108770812* were located on chromosomes 4A and 7D under both conditions. *qDTGW-4A* on chromosome 4A explained 3.26% of the phenotypic variation under the DT conditions, whereas *qDTGW-7D* on chromosome 7D explained 2.69% of the phenotypic variation.

Grain length

Under the normal conditions, 15 QTL for GL, represented by 17 SNPs, were detected on chromosomes 1B, 2A, 2D, 3B, 3D, 4B, 5B, and 5D (Figure 4C, Supplementary Figures 6A–C,

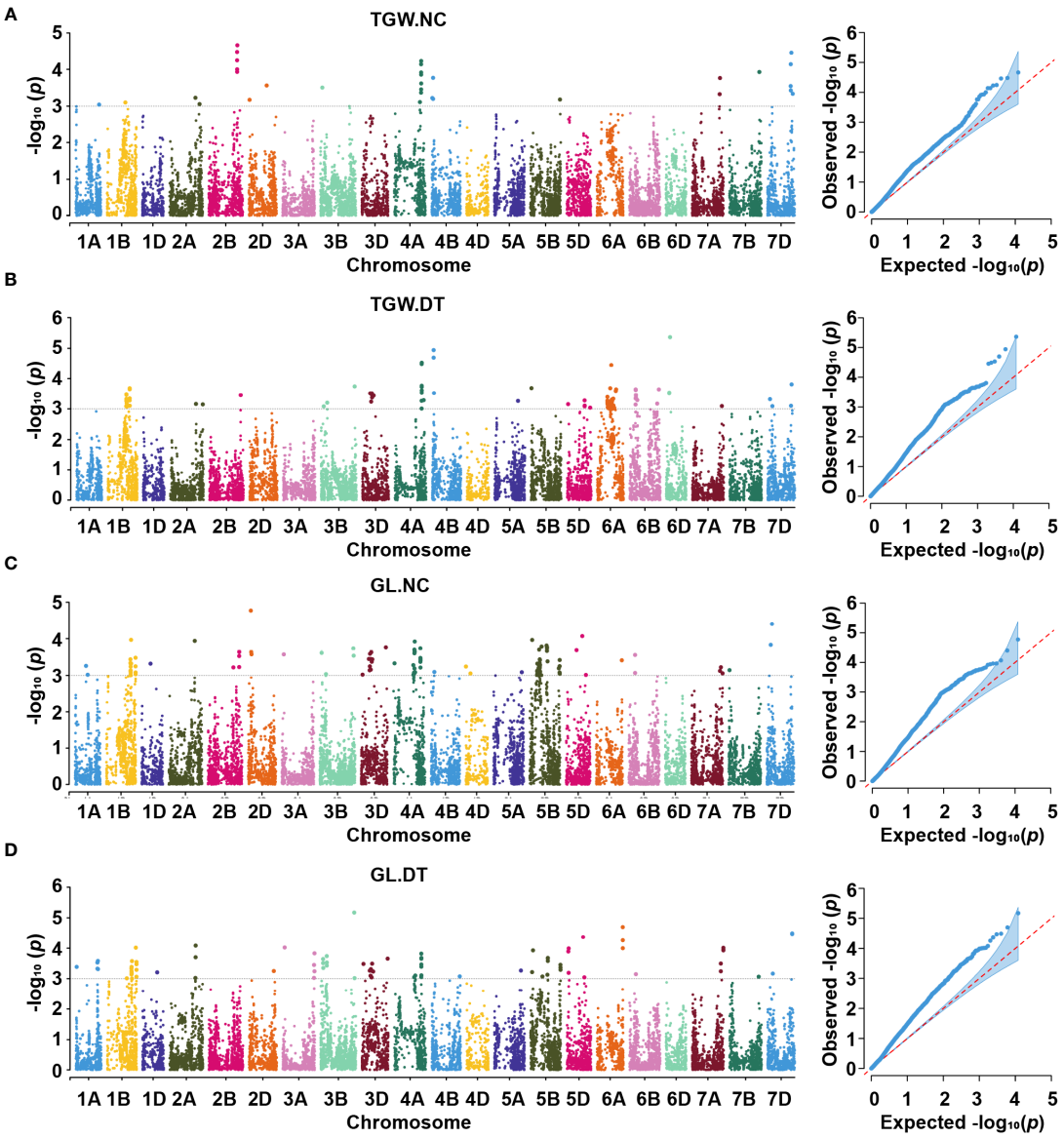


FIGURE 4
Manhattan and Q-Q plots for thousand-grain weight **(A, B)** and grain length **(C, D)** under two conditions. TGW.NC, thousand-grain weight under normal conditions; TGW.DT, thousand-grain weight under drought treatment conditions; GL.NC, grain length under normal conditions; GL.DT, grain length under drought treatment conditions.

Table 2, and Supplementary Table 3). *qNGL-1B.1* (565.51 Mb), *qNGL-1B.2* (580.07 Mb), and *qNGL-1B.3* (685.28 Mb) were detected on chromosome 1B, accounting for 2.30 – 2.82% of the phenotypic variation. Three QTL (*qNGL-2A*, *qNGL-3B.1*, and *qNGL-3D*) marked by *AX-109290429*, *AX-111534973*, and *AX-109392999*, were detected on chromosomes 2A, 3B, and 3D, respectively. Two QTL (*qNGL-2D.1*, *qNGL-2D.2*) were mapped to 63.56 Mb and 79.09 Mb on chromosome 2D. Besides, *qNGL-3B.2* and *qNGL-4B*, explaining 2.63 – 2.82% of the phenotypic variation in GL, were represented by two SNPs, respectively. Finally, five QTL (*qNGL-5B.1*, *qNGL-5B.2*, *qNGL-5B.3*, *qNGL-5D.1*, and *qNGL-5D.2*) located on chromosomes 5B and 5D were represented by five SNPs and accounted for 2.59 – 3.85% of the phenotypic variation.

Under drought treatment conditions, fifteen QTL associated with GL were represented by 26 significant SNPs on chromosomes 1A, 1B, 2A, 2D, 3A, 3B, 3D, 5B, 5D, 6A, 7A, and 7D (Figure 4D, Supplementary Figures 6D–F, Table 3, and Supplementary Table 3). One of the 15, *qDGL-1A*, was detected at the physical position of 503.44 – 504.00 Mb on chromosome

TABLE 2 The QTL and significant SNPs identified by GWAS for the yield-related traits under normal (NC) conditions.

Trait	QTL	SNP	Chr	Position (bp)[a]	Environment	P value	PVE (%)[b]	Reference[c]
TGW.NC	*qNTGW-1B*	*AX-111169510*	1B	430,155,688	NC20, NC21, NAVE	2.90E-04	7.35	
	qNTGW-2A	*AX-109290429*	2A	608,865,127	NC19, NC21, NAVE	1.27E-04	2.91	Mccartney et al, 2005
	qNTGW-2B	*AX-109505207*	2B	666,474,502	NC19, NC21, NAVE	3.29E-05	3.25	Groos et al., 2003
		AX-111661669	2B	666,653,511	NC19, NC21, NAVE	5.68E-05	3.05	
		AX-111728700	2B	667,415,257	NC19, NC21, NAVE	9.98E-05	2.84	
		AX-94979731	2B	667,805,353	NC19, NC21, NAVE	5.87E-05	3.20	
		AX-95219673	2B	671,741,110	NC19, NC21, NAVE	2.16E-05	3.40	
	qNTGW-4A.1	*AX-110171894*	4A	605,753,144	NC19, NC21	2.04E-04	3.44	Zheng et al., 2010;
		AX-109626103	4A	605,764,861	NC19, NC21	3.00E-04	3.26	Jiang et al., 2015
		AX-108777965	4A	605,767,873	NC19, NC21	3.38E-04	2.54	
		AX-110961937	4A	605,785,021	NC19, NC21, NAVE	2.08E-04	2.72	
	qNTGW-4A.2	*AX-109375057*	4A	633,308,367	NC20, NC21, NAVE	7.27E-05	2.96	Cui et al., 2016;
		AX-110468999	4A	634,108,610	NC20, NC21, NAVE	1.39E-04	2.72	Gao et al., 2015;
		AX-111453918	4A	634,833,572	NC20, NC21, NAVE	1.22E-04	2.77	Guan et al., 2018
		AX-108950546	4A	635,764,214	NC20, NC21, NAVE	5.93E-05	3.03	
	qNTGW-7B	*AX-89432708*	7B	701,338,884	NC20, NC21, NAVE	6.98E-06	4.44	Zhang et al., 2014a; Gao et al., 2015; Cao et al., 2020
	qNTGW-7D.1	*AX-108758043*	7D	547,277,450	NC20, NC21, NAVE	7.26E-05	2.96	
	qNTGW-7D.2	*AX-108770812*	7D	561,926,335	NC20, NC21, NAVE	3.44E-05	3.23	Li et al., 2007
GL.NC	*qNGL-1B.1*	*AX-111130381*	1B	565,517,030	NC20, NC21, NAVE	4.60E-04	2.30	
	qNGL-1B.2	*AX-110402488*	1B	580,075,805	NC20, NC21, NAVE	1.08E-04	2.82	
	qNGL-1B.3	*AX-109353011*	1B	685,283,930	NC20, NC21, NAVE	3.31E-04	2.42	
	qNGL-2A	*AX-109290429*	2A	608,865,127	NC19, NC21, NAVE	7.90E-05	3.98	
	qNGL-2D.1	*AX-109013500*	2D	63,568,198	NC20, NC21, NAVE	7.59E-06	3.98	

(Continued)

TABLE 2 Continued

Trait	QTL	SNP	Chr	Position (bp)[a]	Environment	P value	PVE (%)[b]	Reference[c]
	qNGL-2D.2	*AX-111559873*	2D	79,095,877	NC20, NC21, NAVE	5.09E-05	3.25	
	qNGL-3B.1	*AX-111534973*	3B	41,633,676	NC19, NC20, NAVE	2.43E-04	2.53	
	qNGL-3B.2	*AX-110418888*	3B	785,432,286	NC19, NC20, NC21, NAVE	1.85E-04	2.63	Kumar et al., 2016
		AX-109881148	3B	793,075,912	NC19, NC20, NAVE	1.95E-04	2.76	
	qNGL-3D	*AX-109392999*	3D	236,169,118	NC19, NC20, NAVE	2.34E-04	2.54	
	qNGL-4B	*AX-110436979*	4B	73,152,359	NC19, NC21	1.60E-04	2.82	
		AX-110361956	4B	75,741,547	NC19, NC21, NAVE	1.89E-04	2.75	
	qNGL-5B.1	*AX-110398218*	5B	57,493,343	NC19, NC21, NAVE	1.09E-04	2.82	Yang et al., 2020
	qNGL-5B.2	*AX-110516200*	5B	277,898,057	NC20, NC21, NAVE	1.65E-04	2.67	Li et al., 2018
	qNGL-5B.3	*AX-108733256*	5B	702,228,057	NC19, NC20, NAVE	1.03E-04	3.85	Li et al., 2015
	qNGL-5D.1	*AX-111739361*	5D	244,617,835	NC20, NC21, NAVE	2.05E-04	2.59	
	qNGL-5D.2	*AX-110503408*	5D	379,836,128	NC20, NC21, NAVE	8.53E-05	2.91	
GW.NC	*qNGW-4A*	*AX-95629274*	4A	38,367,204	NC19, NC21, NAVE	1.08E-04	2.92	Li et al., 2018
SL.NC	*qNSL-6B*	*AX-109337721*	6B	528,705,208	NC20, NC21, NAVE	7.17E-05	3.15	

[a]Physical position of the SNP in the reference genome (IWGSC RefSeq v1.0); [b]Phenotypic variation explained by the identified SNP; [c]The reference for the near loci previously reported; NC19, NC20, NC21 and NAVE represented 2018 – 2019, 2019 – 2020, 2020 – 2021 and the mean value of three crop seasons for the yield-related traits under NC conditions, respectively.

1A. Two SNPs represented this QTL, collectively explaining 4.69% of the phenotypic variation. Seven of these QTL, including *qDGL-2A.2* (693.33 Mb), *qDGL-2D* (63.56 Mb), *qDGL-3B* (785.43 Mb), *qDGL-3D* (599.68 Mb), *qDGL-5B.1* (57.49 Mb), *qDGL-5D.1* (30.85 Mb) and *qDGL-5D.2* (379.83 Mb) were represented by various SNPs, which accounted for 2.33 – 3.97% of the phenotypic variation. *qDGL-7A* and *qDGL-7D* are located on chromosomes 7A and 7D, represented by five SNPs, accounting for 2.75 – 5.50% of the phenotypic variation. *qDGL-1B* was located at the physical position of 686.75 – 686.78 Mb on chromosome 1B, and *qDGL-5B.2* was at the physical position of 699.99 – 702.22 Mb on chromosome 5B. These two QTL, which explained 2.23 – 3.17% of the phenotypic variation, were represented by three SNPs. Three QTL (*qDGL-2A.1*, *qDGL-3A*, and *qDGL-6A*) represented by two SNPs were mapped on chromosomes 2A, 3A, and 6A, respectively, explaining 2.73 – 4.14% of the phenotypic variation. Additionally, six SNPs, including *AX-109290429* (on chromosome 2A), *AX-109013500* (on chromosome 2D), *AX-110418888* (on chromosome 3B), *AX-110398218* (on chromosome 5B), *AX-108733256* (on chromosome 5B), and *AX-110503408* (on chromosome 5D) were detected under NC and DT conditions. Most notably, *AX-109290429* was associated with TGW.NC, GL.NC, and GL.DT.

Grain width

Under the normal conditions, one QTL, *qNGW-4A*, was detected at 38.36 Mb on chromosome 4A and accounted for

10.3389/fpls.2022.1098560

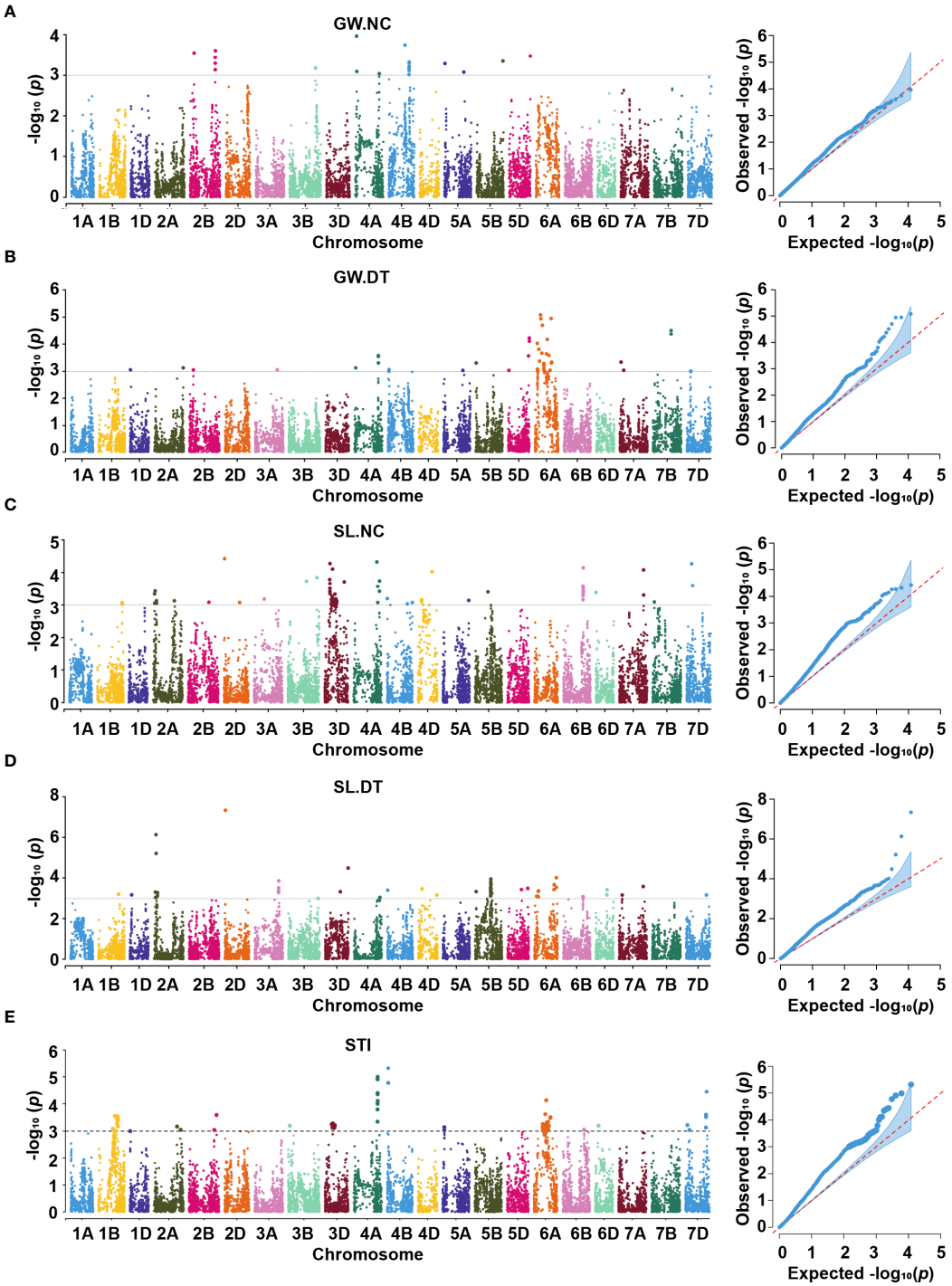


FIGURE 5
Manhattan and Q-Q plots for grain width **(A, B)**, spike length **(C, D)** and stress tolerance index **(E)** under two conditions. GW.NC, grain width under normal conditions; GW.DT, grain width under drought treatment conditions; SL.NC, spike length under normal conditions; SL.DT, spike length under drought treatment conditions; STI, stress tolerance index.

10.3389/fpls.2022.1098560

TABLE 3 The QTL and significant SNPs identified by GWAS for the yield-related traits under drought treatment (DT) conditions.

Trait	QTL	SNP	Chr	Position (bp)[a]	Environment	P value	PVE (%)[b]	Reference[c]
TGW.DT	*qDTGW-4A*	*AX-109375057*	4A	633,308,367	DT20, DT21, DAVE	3.23E-05	3.26	Gao et al., 2015; Cui et al., 2016; Guan et al., 2018
	qDTGW-5A	*AX-109369427*	5A	546,521,932	DT19, DT20	9.24E-05	3.72	Cao et al., 2020;
		AX-110016633	5A	547,333,614	DT19, DT20, DAVE	5.17E-04	2.33	Li et al., 2018
	qDTGW-5D	*AX-110758473*	5D	292,945,585	DT19, DT21	4.27E-04	2.44	
	qDTGW-6A	*AX-109334618*	6A	328,130,345	DT19, DT20, DAVE	4.94E-05	3.08	Wang et al., 2011
	qDTGW-7D	*AX-108770812*	7D	561,926,335	DT19. DT21, DAVE	1.55E-04	2.69	Li et al., 2007
GL.DT	*qDGL-1A*	*AX-110046566*	1A	503,449,799	DT19, DT20, DT21, DAVE	2.95E-04	2.43	
		AX-111802924	1A	504,003,811	DT19, DT20, DT21, DAVE	4.70E-04	2.26	
	qDGL-1B	*AX-108737720*	1B	686,751,506	DT19, DT20, DAVE	2.85E-04	2.44	
		AX-111190944	1B	686,755,725	DT19, DT20, DAVE	5.12E-04	2.23	
		AX-109847715	1B	686,786,607	DT19, DT20, DAVE	3.36E-04	2.38	
	qDGL-2A.1	*AX-110451187*	2A	607,768,697	DT19, DT21, DAVE	2.52E-05	3.42	
		AX-109290429	2A	608,865,127	DT19, DT21, DAVE	1.29E-05	3.67	
	qDGL-2A.2	*AX-109425314*	2A	693,336,312	DT19, DT21	2.91E-04	3.31	
	qDGL-2D	*AX-109013500*	2D	63,568,198	DT20, DT21	4.85E-04	2.33	
	qDGL-3A	*AX-110707258*	3A	733,432,801	DT20, DT21, DAVE	1.47E-04	2.73	
		AX-109281734	3A	744,244,068	DT19, DT20, DAVE	5.23E-05	4.14	
	qDGL-3B	*AX-110418888*	3B	785,432,286	DT19, DT20, DT21, DAVE	6.71E-06	3.78	Kumar et al., 2016
	qDGL-3D	*AX-109500294*	3D	599,688,383	DT19, DT21, DAVE	1.25E-04	2.83	
	qDGL-5B.1	*AX-110398218*	5B	57,493,343	DT19, DT21, DAVE	1.19E-04	2.75	Yang et al., 2020
	qDGL-5B.2	*AX-109879806*	5B	699,999,185	DT20, DT21, DAVE	2.08E-04	2.64	Li et al., 2015;
		AX-108757760	5B	701,544,672	DT19, DT20, DAVE	2.56E-04	2.53	Kumar et al., 2016
		AX-108733256	5B	702,228,057	DT19, DT20, DAVE	3.86E-04	3.17	

(Continued)

TABLE 3 Continued

Trait	QTL	SNP	Chr	Position (bp)[a]	Environment	P value	PVE (%)[b]	Reference[c]
	qDGL-5D.1	*AX-110521341*	5D	30,854,436	DT19, DT21, DAVE	7.43E-05	3.97	
	qDGL-5D.2	*AX-110503408*	5D	379,836,128	DT19, DT21, DAVE	4.30E-05	3.11	
	qDGL-6A	*AX-111034352*	6A	603,563,082	DT19, DT21, DAVE	2.01E-05	3.38	
		AX-94788907	6A	604,107,719	DT19, DT21, DAVE	5.52E-05	3.02	
	qDGL-7A	*AX-108748484*	7A	730,508,297	DT20, DT21, DAVE	9.05E-05	2.91	Gill et al., 2022
		AX-110007752	7A	730,855,963	DT20, DT21, DAVE	1.18E-04	2.75	
		AX-109272527	7A	731,240,127	DT20, DT21, DAVE	9.80E-05	2.82	
	qDGL-7D	*AX-108815963*	7D	577,348,898	DT19, DT21, DAVE	3.37E-06	5.50	
		AX-109867095	7D	577,937,328	DT19, DT21, DAVE	3.36E-06	5.50	
GW.DT	*qDGW-6A.1*	*AX-111041695*	6A	148,524,505	DT19, DT21, NAVE	8.24E-06	3.82	Li et al., 2022b
		AX-110936625	6A	158,068,188	DT19, DT21, NAVE	1.14E-05	3.70	
	qDGW-6A.2	*AX-110942969*	6A	428,454,560	DT19, DT21, NAVE	1.12E-05	3.71	Li et al., 2022b
SL.DT	*qDSL-2A*	*AX-108747720*	2A	58,539,285	DT19, DT20, DAVE	6.08E-06	4.01	
	qDSL-2D	*AX-110647062*	2D	23,025,488	DT19, DT20, DAVE	1.45E-08	6.31	Ma et al., 2007; Heidari et al., 2011; Wang et al., 2011; Wu et al., 2014; Deng et al., 2017; Chai et al., 2019; Li et al, 2022a;
	qDSL-3D	*AX-111027124*	3D	400,781,396	DT19, DT20, DAVE	1.79E-04	2.70	Yu et al., 2014
	qDSL-5B	*AX-111124661*	5B	437,564,786	DT19, DT20	2.44E-04	2.63	Gill et al., 2022;

[a]Physical position of the SNP in the reference genome (IWGSC RefSeq v1.0); [b]Phenotypic variation explained by the identified SNP; [c]The reference for the near loci previously reported; DT19, DT20, DT21 and DAVE represented 2018 – 2019, 2019 – 2020, 2020 – 2021 and the mean value of three crop seasons for the yield-related traits under DT conditions, respectively.

2.92% of the phenotypic variation (Figure 5A, Supplementary Figures 7A–C, Table 2, and Supplementary Table 3). By contrast, the DT conditions revealed two QTL represented by three SNPs on chromosome 6A (Figure 5B, Supplementary Figures 7D–F, Table 3, and Supplementary Table 3). The first QTL *qDGW-6A.1*, located at 148.52 – 158.06 Mb, was represented by two SNPs and explained 3.70 – 3.82% of the total phenotypic variation. The second QTL, *qDGW-6A.2*, represented by SNP *AX-110942969*, was mapped to the physical position of 428.45 Mb and explained 3.71% of the phenotypic variation.

Spike length

For SL, one QTL, *qNSL-6B*, represented by *AX-109337721*, was mapped to 528.70 Mb on chromosome 6B and explained 3.15% of the total phenotypic variation under the normal conditions (Figure 5C, Supplementary Figures 8A–C,Table 2, and Supplementary Table 3). In contrast, we detected four QTL represented by four SNPs on chromosomes 2A, 2D, 3D, and 5B under the DT conditions (Figure 5D, Supplementary Figures 8D–F, Table 3, and Supplementary Table 3). The first

TABLE 4 The QTL and significant SNPs identified by GWAS for stress tolerance index (STI).

QTL	SNP	Chr	Position (bp)[a]	Environment	*P* value	PVE (%)[b]
qSTI-1B.1	*AX-111503092*	1B	380,647,070	STI20, STI21, STIA	5.72E-05	3.48
	AX-110041061	1B	390,870,930	STI20, STI21	2.42E-04	2.88
qSTI-1B.2	*AX-111169510*	1B	430,155,688	STI20, STI21, STIA	1.50E-04	2.85
qSTI-2A	*AX-109290429*	2A	608,865,127	STI19, STI21, STIA	5.11E-04	2.39
qSTI-4A.1	*AX-110171894*	4A	605,753,144	STI19, STI21	2.20E-04	3.37
	AX-108777965	4A	605,767,873	STI19, STI21	2.56E-04	3.30
	AX-110961937	4A	605,785,021	STI19, STI21	2.13E-04	3.38
qSTI-4A.2	*AX-109375057*	4A	633,308,367	STI20, STI21, STIA	9.76E-06	4.22
	AX-110468999	4A	634,108,610	STI20, STI21, STIA	2.77E-05	3.78
	AX-111140650	4A	634,368,256	STI20, STI21, STIA	4.82E-05	3.55
	AX-111453918	4A	634,833,572	STI20, STI21, STIA	2.19E-05	3.88
	AX-110440161	4A	635,292,292	STI20, STI21, STIA	2.20E-05	3.88
	AX-108950546	4A	635,764,214	STI20, STI21, STIA	6.19E-06	4.41
qSTI-4B	*AX-111150060*	4B	660,589,159	STI19, STI21	5.86E-04	2.91
qSTI-6A	*AX-109334618*	6A	328,130,345	STI19, STI21, STIA	7.48E-05	2.92
qSTI-7D.1	*AX-108758043*	7D	547,277,450	STI20, STI21, STIA	1.52E-04	3.08
	AX-110949705	7D	548,055,156	STI19, STI20, STI21, STIA	1.62E-04	3.05
qSTI-7D.2	*AX-108770812*	7D	561,926,335	STI19, STI21, STIA	3.57E-05	3.19

[a]Physical position of the SNP in the reference genome (IWGSC RefSeq v1.0); [b]Phenotypic variation explained by the identified SNP; STI19, STI20, STI21 and STIA represented 2018 – 2019, 2019 – 2020, 2020 – 2021 and the mean value of three crop seasons for STI, respectively.

QTL, *qDSL-2A*, represented by *AX-108747720*, was marked to chromosome 2A and explained 4.01% of the phenotypic variation. The second QTL, *qDSL-2D*, detected on chromosome 2D (at 23.02 Mb), was represented by *AX-110647062* and explained 6.31% of phenotypic variation in SL. The last two QTL (*qDSL-3D* and *qDSL-5B*) explained 2.70 and 2.63% of the phenotypic variation, respectively.

Stress tolerance index

For STI, nine high confidence QTL were detected under more than one environment. These QTL involved 18 SNPs distributed on six chromosomes (including 1B, 2A, 4A, 4B, 6A and 7D) (Figure 5E, Supplementary Figure 9, Table 4, and Supplementary Table 4). Five of these QTL, including *qSTI-1B.2* (430.15 Mb), *qSTI-2A* (608.86 Mb), *qSTI-4B* (660.58 Mb), *qSTI-6A* (328.13 Mb) and *qSTI-7D.2* (561.92 Mb) were represented by various SNPs, which accounted for 2.39 – 3.19% of the phenotypic variation. In addition, *qSTI-1B.1* was located at 380.64 – 390.87 Mb on chromosome 1B, and *qSTI-7D.1* was detected at 547.27 – 548.05 Mb on chromosome 7D. These two QTL, which explained 2.88 – 3.48% of the phenotypic variation, were represented by two SNPs, respectively. Another two QTL (*qSTI-4A.1* and *qSTI-4A.2*) represented by nine SNPs on chromosome 4A totally explained 33.76% of the phenotypic variation. Notably, twelve SNPs (including *AX-111169510*, *AX-109290429*, *AX-110171894*, *AX-108777965*, *AX-110961937*,*AX-109375057*, *AX-110468999*, *AX-111453918*, *AX-108950546*, *AX-109334618*, *AX-108758043* and *AX-108770812*) had association with STI as well as TGW.

Additive effects of the superior alleles on phenotypes

The marker alleles associated with high TGW, GL, GW, and SL were considered "superior alleles", and the marker alleles with the opposite effects as "inferior alleles" to ease the description of allelic effects. The allelic effect was verified using the highest $-\log_{10}(P)$ SNPs within the detected QTL in the association population by examining whether the differences in phenotypic values (grouped by polymorphism) reached a significant level (Figure 6). The number of superior alleles in each wheat line was calculated based on representative SNPs for the loci detected using the average values for grain and spike traits across all environments. The patterns of relationship were similar for the detected traits,

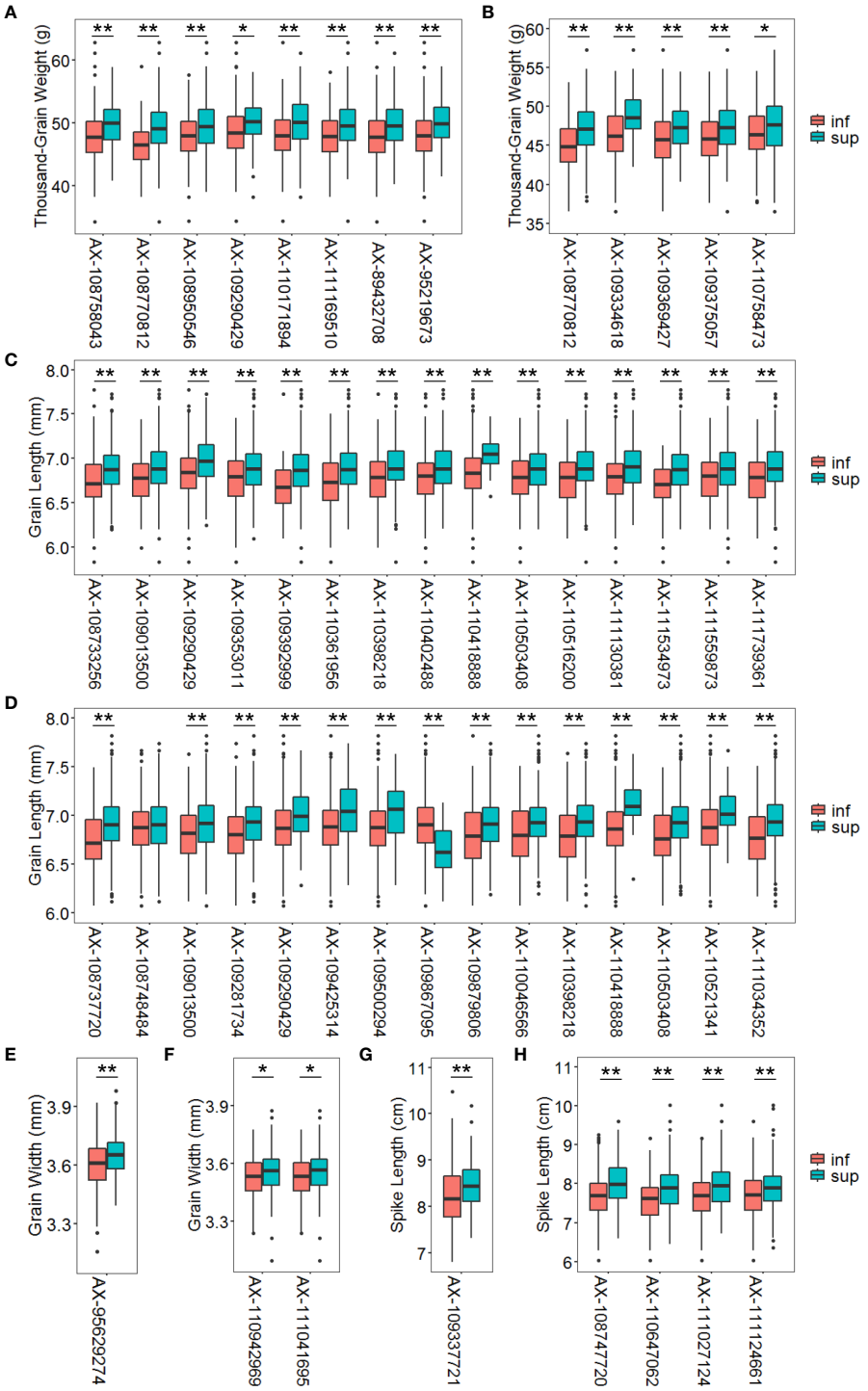


FIGURE 6
Average phenotypic values of accessions with different alleles in multi-environment significant loci under normal and drought treatment conditions for thousand-grain weight **(A, B)**, grain length **(C, D)**, grain width **(E, F)** and spike length **(G, H)**. The *P* value was determined by the two tailed Student's *t*-test. **P* < 0.05, ***P* < 0.01. Inf represents the inferior allele and sup represents the superior allele.

where superior alleles additively increased TGW.NC (R^2 = 0.2015), TGW.DT (R^2 = 0.1581), GL.NC (R^2 = 0.2256), GL.DT (R^2 = 0.2853), GW.DT (R^2 = 0.0098) and SL.DT (R^2 = 0.1230) (Figure 7).

KASP development

Three significant SNP markers (*AX-109290429*, *AX-110418888*, and *AX-109369427*) on chromosomes 2A, 3B, and 5A were identified in multiple environments were converted into KASP markers. The KASP markers could accurately classify different alleles. The results showed that different alleles carrying FAM and HEX fluorescence were distributed around the X and Y axes, respectively (Supplementary Figure 10). These KASP marker-based allelic typing results were consistent with those on the 15K SNP array.

Discussion

The drought stress reduces wheat yields

Wheat is particularly vulnerable to drought stress during growth. This study investigated the effects of drought stress on yield-related traits of wheat. Under the drought treatment conditions, TGW, GW, and SL were reduced by 4, 2, and 6% compared to the normal conditions (Figure 1 and Table 1). However, the broad sense heritability of TGW, GL, GW, and SL was relatively high (76.16 – 95.66%) under both conditions, consistent with previous studies (Sukumaran et al., 2018; Wolde et al., 2019; Ji et al., 2021; Wang et al., 2021; Wang et al., 2022). Furthermore, the association panel showed significant phenotypic variation in TGW, GL, GW, and SL under normal and drought conditions. These results highlight the potential of

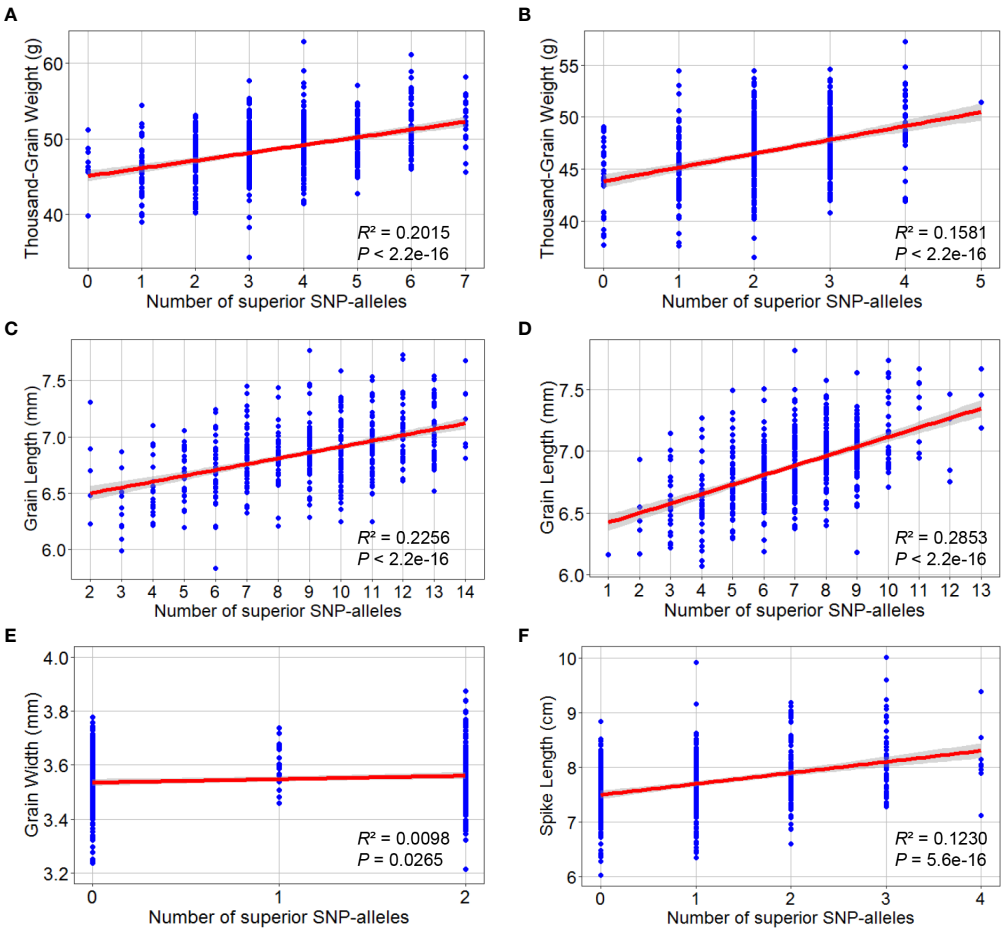


FIGURE 7
Superior allelic effects of the highest -$\log_{10}$ (*P*) SNPs by GWAS for thousand-grain weight, grain length, grain width and spike length. Thousand-grain weight under normal **(A)** and drought treatment **(B)** conditions; Grain length under normal **(C)** and drought treatment **(D)** conditions; **(E)** Grain width under drought treatment condition; **(F)** Spike length under drought treatment condition. At the bottom of the box plot corresponding to each SNP allele.

these germplasms to identify the alleles associated with plant adaption to harsh environments for improving elite wheat yield.

Identifying and comparing SNPs with previous findings

The association results were validated using MLM, which considers population structure and kinship to define the significant SNPs in over two environments under normal and drought conditions. Therefore, 37 and 39 significant SNPs (in 51 QTL) were considered reliable MTAs in multiple environments associated with the yield-related traits under NC and DT conditions (Tables 2, 3). Eighteen MTAs (in nine QTL) associated with STI were identified in more than one environment (Table 4). Furthermore, we identified the QTL for the stability of grain and spike traits by comparing the MTAs detected under normal and drought treatment conditions simultaneously with those detected only under normal conditions.

Normal conditions

We observed 18 credible SNPs associated with TGW, representing eight loci mapped onto chromosomes 1BL, 2AL, 2BL, 4AL, 7BL, and 7DL under normal conditions. One SNP cluster (consisting of five SNPs) was identified on chromosome 2BL, spanning the 5.27 Mb (2B: 666.47 – 671.74 Mb) physical position. This locus was located on the same genomic region containing previously reported QTL for TGW (Groos et al., 2003). In addition, two adjacent major QTL represented by eight SNPs were detected on chromosome 4AL, explaining 23.44% of the TGW variance. One QTL, *qNTGW-4A.1*, accounted for 11.96% of the phenotypic variation, consistent with a previous study (Zheng et al., 2010). Moreover, the QTL was adjacent to the yield-associated gene, *TaCWI-4A*, in the Jiang consensus map (Jiang et al., 2015). Another QTL, *qNTGW-4A.2*, located at the 633.30 – 635.76 Mb interval on chromosome 4A, was adjacent to other QTL for TGW, as previously indicated (Gao et al., 2015; Cui et al., 2016; Guan et al., 2018). In addition, *qNTGW-2A*, *qNTGW-7B*, and *qNTGW-7D.2* associated with TGW were detected on chromosomes 2AL, 7BL, and 7DL, respectively (Mccartney et al., 2005; Li et al., 2007; Zhang et al., 2014a; Gao et al., 2015; Cao et al., 2020). Although previous studies also identified several QTL controlling TGW on chromosome 1B (Zheng et al., 2010; Wang et al., 2012; Cui et al., 2014; Sukumaran et al., 2018; Qu et al., 2021), the QTL were inconsistent with *qNTGW-1B* detected in this study, suggesting that *qNTGW-1B*, reported for the first time in the current study was a novel QTL.

A total of 15 QTL containing 17 SNPs associated with GL were detected on chromosomes 1B, 2A, 2D, 3B, 3D, 4B, 5B, and 5D under normal conditions. Among these loci, *qNGL-3B.2*, associated with GL, was marked at 785.43 – 793.05 Mb on chromosome 3BL. Previous studies have also identified QTL that controls GL at adjacent positions (Kumar et al., 2016). We identified three QTL on chromosome 5B, which explained 9.34% of phenotypic variation in GL. The first QTL, *qNGL-5B.1*, on chromosome 5BS was mapped at a similar position with *Qgl.caas-5BS*, associated with GL (Yang et al., 2020). The second QTL was located at the neighboring chromosomal region of the grain-length-related QTL, *QKL.caas-5BS*, detected in the Chinese bread wheat recombinant inbred population (Li et al., 2018). The third QTL, *qNGL-5B.3*, on chromosome 5BL shared a similar region with *QKl5B.2-16* for GL (Li et al., 2015). Notably, the SNP *AX-109290429* on chromosome 2AL was significantly associated with GL, TGW and STI, suggesting that this locus is highly valuable for these traits. Position-based comparisons of the SNPs identified in this study with previous reports identified another ten new and preliminary loci associated with GL on chromosomes 1B, 2D, 3B, 3D, 4B, and 5D.

Several previous studies also identified QTL for GW, such as *qKW-4A* (Su et al., 2018), *QKW.ndsu.4A* (Kumar et al., 2016), *QKw-4A.1* (Cui et al., 2016), *QKW.caas-4AS* (Li et al., 2018), and *QKW.caas-4AL* (Li et al., 2018) on chromosome 4A. This study also detected *qNGW-4A* for GW on chromosome 4AS under normal conditions, similar with the position identified by Li et al. (Li et al., 2018).

Moreover, there were several QTL for SL detected on chromosome 6B in the previous studies (Liu et al., 2014; Gao et al., 2015; Deng et al., 2017; Li et al., 2018; Cao et al., 2019). We also detected a credible QTL at the physical position of 528.70 Mb on chromosome 6B. However, this QTL for SL was not previously reported, suggesting that it may be new.

Drought treatment conditions

The SNPs detected under DT differed from those under the normal conditions (Table 3 and Supplementary Table 3) owing to the effects of the drought stress. We also detected five QTL represented by six SNPs for TGW under DT. Most notably, two of these SNPs (*AX-109375057* and *AX-108770812*) were associated with TGW.NC, TGW.DT and STI, suggesting that these SNPs' effects on TGW and STI were extremely important and unaffected by environmental changes (Tables 2–4 and Supplementary Tables 3, 4). Moreover, the QTL *qDTGW-5A* for TGW was located at adjacent chromosomal regions (Li et al., 2018; Cao et al., 2020). Furthermore, *qDTGW-6A* was marked at an approximately equivalent chromosomal region in a previous study (Wang et al., 2011). Comparisons between *qDTGW-5D* discovered in this research with other QTL identified in previous studies showed that *qDTGW-5D*, located on chromosome 5D, was a newly detected QTL regulating TGW.

For GL, nine QTL were detected under DT conditions, and six others under NC and DT conditions. Previously, *qDGL-7A*

for GL was only detected under DT conditions in the region near SNP S7A_717859384, associated with GL (Gill et al., 2022). For GW, we detected two QTL on chromosome 6A under DT. These QTL were marked at a near location similar with a previously detected QTL for GW (Li et al., 2022b). We discovered four QTL for SL on chromosome 2AS, 2DS, 3DL, and 5BL under DT. Of these, *qDSL-2D, qDSL-3D*, and *qDSL-5B* were located at similar regions with QTL for SL from previous studies (Ma et al., 2007; Heidari et al., 2011; Wang et al., 2011; Wu et al., 2014; Yu et al., 2014; Deng et al., 2017; Chai et al., 2019; Gill et al., 2022; Li et al., 2022a). Notably, *qDSL-2A* for SL was first discovered in the current study by comparing this study with previous research.

A total of nine high confidence QTL for STI were detected, out of which, seven QTL (*qDRI-1B.2*, *qDRI-2A*, *qDRI-4A.1*, *qDRI-4A.2*, *qDRI-6A*, *qDRI-7D.1*, *qDRI-7D.2*) were co-located with QTL for TGW under NC or DT conditions. Therefore, these seven co-located QTL can be used to synergistically improve the yield and drought resistance of wheat. To the best of our knowledge, only a few QTL for STI, which was calculated from TGW, have been detected in wheat (Negisho et al., 2022). Nevertheless, they were located at different chromosomal regions with these QTL for STI detected in this study.

Many agronomic traits are complex quantitative traits that are susceptible to environmental factors. Therefore, this study detected eight stable MTAs (two for TGW and STI, six for GL) under normal and drought conditions. The MTAs detected are important for both normal and drought treatment conditions, providing insights into the genetic basis of the stability of agronomic traits. These MTAs can be used to develop useful markers for genetic improvement in wheat breeding.

Validation of multi-environment significant SNPs in the association population

This study identified numerous SNPs closely associated with TGW, GL, GW, SL and STI. Among these, three SNP markers were developed as KASP markers, making the genotype identification process more efficient, inexpensive, fast, convenient, and greatly accelerating the process of molecular marker-assisted selection. Under normal conditions, accessions with superior alleles for TGW at eight SNP loci showed TGW values of 49.29 to 50.17 g. In contrast, accessions with inferior alleles at these loci showed TGW values of 46.45 – 48.54 g. Under DT conditions, accessions with superior alleles for TGW at five SNP loci showed TGW values of 47.36 – 49.03 g compared to 44.93 – 46.55 g for accessions with inferior alleles. Nevertheless, the frequency of superior alleles of SNP *AX-109334618* on chromosome 6A was only about 5%, indicating that this SNP has not been fully utilized. Furthermore, this SNP was associated with not only TGW but also STI, suggesting that it is worth considering as a major genetic locus for improving TGW and drought resistance (Figures 6A, B).

Under normal conditions, accessions with superior alleles for GL at 15 SNP loci showed significantly higher GL values (6.87 – 7.07 mm) than those with inferior alleles (6.68 – 6.84 mm). Moreover, under DT conditions, the GL values of accessions with superior alleles for this trait at 15 SNP loci ranged from 6.89 to 7.12 mm compared to 6.64 – 6.87 mm for accessions with inferior alleles at the same loci. However, superior alleles at some loci, such as *AX-109290429*, *AX-109425314*, *AX-110418888*, *AX-109500294*, and *AX-110521341*, were present in < 12% of the accessions, suggesting that these loci require more attention when studying the improvement of GL (Figures 6C, D). Under normal conditions, accessions with a superior allele at *AX-95629274* showed 3.65 mm GW values than the 3.60 mm for accessions with an inferior allele at this locus. Furthermore, accessions with superior alleles at two SNP loci showed higher GW values (~3.53 mm) than accessions with inferior alleles (~3.56 mm) under DT conditions (Figures 6E, F).

However, accessions with the superior allele of SNP *AX-109337721* showed significantly higher SL values (~8.17 cm) than those with the inferior allele (~8.52 cm) under normal conditions. Notably, the superior allele of this SNP was < 14% frequent among accessions, indicating that this SNP is underutilized and a probably significant genetic locus for improving SL. Finally, the SL values of accessions with superior alleles for this trait at the four SNP loci were 7.88 – 8.01 cm compared to 7.55 – 7.70 cm for accessions with inferior alleles under DT conditions. Nevertheless, the frequencies of superior alleles for *AX-108747720*, *AX-111027124*, and *AX-111124661* were < 26%, indicating these loci require more attention when improving SL (Figures 6G, H).

The TGW of the detected varieties from different decades increased with increasing frequency of superior alleles under both conditions. Furthermore, increased TGW was strongly selected for breeding before the 2010s. However, after 2020, the average TGW of the accessions in this study decreased slightly, as did the corresponding number of superior alleles in both conditions (Supplementary Figure 11). Altogether, the elite alleles at these loci detected under two conditions can be integrated into wheat varieties using molecular breeding to improve wheat yield.

Conclusion

In summary, this study identified high confidence loci for yield-related and drought-resistance-related traits in normal and drought-stressed common wheat by GWAS and developed three KASP markers associated with these traits, which is a major step to accelerate the breeding of high-yielding and drought-resistant varieties.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

HL and YJZ conceived and designed the experiments. JZ, LS and HG performed most of the experiments and wrote the manuscript. MH, LM, XC and JW provided technical assistance and conducted the collection and maintenance of wheat germplasm. YZ, QL and PW involved in discussion and participated in field trials. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Natural Science Foundation of China (32101749), Science and Technology Program of Hebei (20326345D, C2022301004), Hebei Modern Agricultural Industrial Technology System (HBCT2018010201), Project for Hebei Scientific and Technological Innovation Team of Modern Wheat Seed Industry (21326318D), HAAFS Science and Technology Innovation Special Project (2022KJCXZX-LYS-1, 2022KJCXZX-LYS-20) and Talents construction project of science and technology innovation, Hebei Academy of Agriculture and Forestry Sciences (C22R0305).

Acknowledgments

The authors thank Dr. Yelun Zhang and Dr. Liangjie Lv, Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, for providing the wheat accessions.

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/fpls.2022.1098560/full#supplementary-material

References

Ahmed, H. G. M., Zeng, Y., Shah, A. N., Yar, M. M., Ullah, A., and Ali, M. (2022). Conferring of drought tolerance in wheat (*Triticum aestivum* L.) genotypes using seedling indices. *Front. Plant Sci.* 13, 961049. doi: 10.3389/fpls.2022.961049

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, 2633–2635. doi: 10.1093/bioinformatics/btm308

Cao, P., Liang, X., Zhao, H., Feng, B., Xu, E., Wang, L., et al. (2019). Identification of the quantitative trait loci controlling spike-related traits in hexaploid wheat (*Triticum aestivum* L.). *Planta* 250, 1967–1981. doi: 10.1007/s00425-019-03278-0

Cao, J., Liu, K., Song, W., Zhang, J., Yao, Y., Xin, M., et al. (2021). Pleiotropic function of the *SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE* gene *TaSPL14* in wheat plant architecture. *Planta* 253, 44. doi: 10.1007/s00425-020-03531-x

Cao, S., Xu, D., Hanif, M., Xia, X., and He, Z. (2020). Genetic architecture underpinning yield component traits in wheat. *Theor. Appl. Genet.* 133, 1811–1823. doi: 10.1007/s00122-020-03562-8

Chai, L., Chen, Z., Bian, R., Zhai, H., Cheng, X., Peng, H., et al. (2019). Dissection of two quantitative trait loci with pleiotropic effects on plant height and spike length linked in coupling phase on the short arm of chromosome 2D of common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 132, 1815–1831. doi: 10.1007/s00122-019-03318-z

Chang, J., Zhang, J., Mao, X., Li, A., Jia, J., and Jing, R. (2013). Polymorphism of *TaSAP1-A1* and its association with agronomic traits in wheat. *Planta* 237, 1495–1508. doi: 10.1007/s00425-013-1860-x

Cui, F., Fan, X., Chen, M., Zhang, N., Zhao, C., Zhang, W., et al. (2016). QTL detection for wheat kernel size and quality and the responses of these traits to low nitrogen stress. *Theor. Appl. Genet.* 129, 469–484. doi: 10.1007/s00122-015-2641-7

Cui, F., Zhao, C., Ding, A., Li, J., Wang, L., Li, X., et al. (2014). Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. *Theor. Appl. Genet.* 127, 659–675. doi: 10.1007/s00122-013-2249-8

Deng, Z., Cui, Y., Han, Q., Fang, W., Li, J., and Tian, J. (2017). Discovery of consistent QTLs of wheat spike-related traits under nitrogen treatment at different development stages. *Front. Plant Sci.* 8, 2120. doi: 10.3389/fpls.2017.02120

Dong, L., Wang, F., Liu, T., Dong, Z., Li, A., Jing, R., et al. (2014). Natural variation of *TaGASR7-A1* affects grain length in common wheat under multiple cultivation conditions. *Mol. Breed.* 34, 937–947. doi: 10.1007/s11032-014-0087-2

Fernandez, G. C. J. (1992). "Effective selection criteria for assessing plant stress tolerance," in *Proceedings of the International Symposium on Adaptation of Vegetables and Other Food Crops in Temperature and Water Stress*. Ed. C. G. Kuo (Tainan, Taiwan).

Gao, F., Wen, W., Liu, J., Rasheed, A., Yin, G., Xia, X., et al. (2015). Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross zhou 8425B/Chinese spring. *Front. Plant Sci.* 6, 1099. doi: 10.3389/fpls.2015.01099

Gill, H. S., Halder, J., Zhang, J., Rana, A., Kleinjan, J., Amand, P. S., et al. (2022). Whole-genome analysis of hard winter wheat germplasm identifies genomic regions associated with spike and kernel traits. *Theor. Appl. Genet.* 135, 2953–2967. doi: 10.1007/s00122-022-04160-6

Groos, C., Robert, N., Bervas, E., and Charmet, G. (2003). Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theor. Appl. Genet.* 106, 1032–1040. doi: 10.1007/s00122-002-1111-1

Guan, P., Lu, L., Jia, L., Kabir, M. R., Zhang, J., Zhao, Y., et al. (2018). Global QTL analysis identifies genomic regions on chromosomes 4A and 4B harboring stable loci for yield-related traits across different environments in wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 9, 529. doi: 10.3389/fpls.2018.00529

Guo, X., Fu, Y., Lee, Y. J., Chern, M., Li, M., Cheng, M., et al. (2022). The PGS1 basic helix-loop-helix (bHLH) protein regulates *Fl3* to impact seed growth and grain yield in cereals. *Plant Biotechnol. J.* 20, 1311–1326. doi: 10.1111/pbi.13809

Hanif, M., Gao, F., Liu, J., Wen, W., Zhang, Y., Rasheed, A., et al. (2016). *TaTGW6-A1*, an ortholog of rice *TGW6*, is associated with grain weight and yield in bread wheat. *Mol. Breed.* 36, 1. doi: 10.1007/s11032-015-0425-z

Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). *Methods Mol. Biol.* 1019, 149–169. doi: 10.1007/978-1-62703-447-0_6

Heidari, B., Sayed-Tabatabaei, B. E., Saeidi, G., Kearsey, M., and Suenaga, K. (2011). Mapping QTL for grain yield, yield components, and spike features in a doubled haploid population of bread wheat. *Genome* 54, 517–527. doi: 10.1139/g11-017

Hong, Y., Chen, L., Du, L. P., Su, Z., Wang, J., Ye, X., et al. (2014). Transcript suppression of *TaGW2* increased grain width and weight in bread wheat. *Funct. Integr. Genomics* 14, 341–349. doi: 10.1007/s10142-014-0380-5

Hu, M. J., Zhang, H. P., Liu, K., Cao, J. J., Wang, S. X., Jiang, H., et al. (2016). Cloning and characterization of *TaTGW-7A* gene associated with grain weight in wheat *via* SLAF-seq-BSA. *Front. Plant Sci.* 7, 1902. doi: 10.3389/fpls.2016.01902

Jia, M., Li, Y., Wang, Z., Tao, S., Sun, G., Kong, X., et al. (2021). TaIAA21 represses TaARF25-mediated expression of *TaERFs* required for grain size and weight development in wheat. *Plant J.* 108, 1754–1767. doi: 10.1111/tpj.15541

Jiang, Y., Jiang, Q., Hao, C., Hou, J., Wang, L., Zhang, H., et al. (2015). A yield-associated gene *TaCWI*, in wheat: Its function, selection and evolution in global breeding revealed by haplotype analysis. *Theor. Appl. Genet.* 128, 131–143. doi: 10.1007/s00122-014-2417-5

Ji, G., Xu, Z., Fan, X., Zhou, Q., Yu, Q., Liu, X., et al. (2021). Identification of a major and stable QTL on chromosome 5A confers spike length in wheat (*Triticum aestivum* L.). *Mol. Breed.* 41, 56. doi: 10.1007/s11032-021-01249-6

Juliana, P., Govindan, V., Crespo-Herrera, L., Mondal, S., Huerta-Espino, J., Shrestha, S., et al. (2022). Genome-wide association mapping identifies key genomic regions for grain zinc and iron biofortification in bread wheat. *Front. Plant Sci.* 13, 903819. doi: 10.3389/fpls.2022.903819

Kang, Y., Barry, K., Cao, F., and Zhou, M. (2020). Genome-wide association mapping for adult resistance to powdery mildew in common wheat. *Mol. Boil. Rep.* 47, 1241–1256. doi: 10.1007/s11033-019-05225-4

Kumar, A., Mantovani, E. E., Seetan, R., Soltani, A., Echeverry-Solarte, M., Jain, S., et al. (2016). Dissection of genetic factors underlying wheat kernel shape and size in an elite × nonadapted cross using a high density SNP linkage map. *Plant Genome* 9, 1–22. doi: 10.3835/plantgenome2015.09.0081

Li, T., Deng, G., Su, Y., Yang, Z., Tang, Y., Wang, J., et al. (2022b). Genetic dissection of quantitative trait loci for grain size and weight by high-resolution genetic mapping in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 135, 257–271. doi: 10.1007/s00122-021-03964-2

Li, A., Hao, C., Wang, Z., Geng, S., Jia, M., Wang, F., et al. (2022a). Wheat breeding history reveals synergistic selection of pleiotropic genomic sites for plant architecture and grain yield. *Mol. Plant* 15, 504–519. doi: 10.1016/j.molp.2022.01.004

Li, S., Jia, J., Wei, X., Zhang, X., Li, L., Chen, H., et al. (2007). A intervarietal genetic map and QTL analysis for yield traits in wheat. *Mol. Breed.* 20, 167–178. doi: 10.1007/s11032-007-9080-3

Liu, G., Jia, L., Lu, L., Qin, D., Zhang, J., Guan, P., et al. (2014). Mapping QTLs of yield-related traits using RIL population derived from common wheat and Tibetan semi-wild wheat. *Theor. Appl. Genet.* 127, 2415–2432. doi: 10.1007/s00122-014-2387-7

Liu, H., Li, H., Hao, C., Wang, K., Wang, Y., Qin, L., et al. (2020). *TaDA1*, a conserved negative regulator of kernel size, has an additive effect with *TaGW2* in common wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 18, 1330–1342. doi: 10.1111/pbi.13298

Li, F., Wen, W., He, Z., Liu, J., Jin, H., Cao, S., et al. (2018). Genome-wide linkage mapping of yield-related traits in three Chinese bread wheat populations using high-density SNP markers. *Theor. Appl. Genet.* 131, 1903–1924. doi: 10.1007/s00122-018-3122-6

Li, F., Wen, W., Liu, J., Zhang, Y., Cao, S., He, Z., et al. (2019). Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* 19, 168. doi: 10.1186/s12870-019-1781-3

Li, Q., Zhang, Y., Liu, T., Wang, F., Liu, K., Chen, J., et al. (2015). Genetic analysis of kernel weight and kernel size in wheat (*Triticum aestivum* L.) using unconditional and conditional QTL mapping. *Mol. Breed.* 35, 194. doi: 10.1007/s11032-015-0384-4

Li, L., Zhang, Y., Zhang, Y., Li, M., Xu, D., Tian, X., et al. (2021). Genome-wide linkage mapping for preharvest sprouting resistance in wheat using 15K single-nucleotide polymorphism arrays. *Front. Plant Sci.* 12, 749206. doi: 10.3389/fpls.2021.749206

Ma, L., Li, T., Hao, C., Wang, Y., Chen, X., and Zhang, X. (2016). *TaGS5-3A*, a grain size gene selected during wheat improvement for larger kernel and yield. *Plant Biotechnol. J.* 14, 1269–1280. doi: 10.1111/pbi.12492

Ma, M., Wang, Q., Li, Z., Cheng, H., Li, Z., Liu, X., et al. (2015). Expression of *TaCYP78A3*, a gene encoding cytochrome P450 CYP78A3 protein in wheat (*Triticum aestivum* L.), affects seed size. *Plant J.* 83, 312–325. doi: 10.1111/tpj.12896

Ma, D., Yan, J., He, Z., Wu, L., and Xia, X. (2012). Characterization of a cell wall invertase gene *TaCwi-A1* on common wheat chromosome 2A and development of functional markers. *Mol. Breed.* 29, 43–52. doi: 10.1007/s11032-010-9524-z

Ma, Z., Zhao, D., Zhang, C., Zhang, Z., Xue, S., Lin, F., et al. (2007). Molecular genetic analysis of five spike-related traits in wheat using RIL and immortalized F_2 populations. *Mol. Genet. Genomics* 277, 31–42. doi: 10.1007/s00438-006-0166-0

Mccartney, C. A., Somers, D. J., Humphreys, D. G., Lukow, O., Ames, N., Noll, J., et al. (2005). Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 × 'AC domain'. *Genome* 48, 870–883. doi: 10.1139/g05-055

Negisho, K., Shibru, S., Matros, A., Pillen, K., Ordon, F., and Wehner, G. (2022). Association mapping of drought tolerance indices in Ethiopian durum wheat (*Triticum turgidum* ssp. durum). *Front. Plant Sci.* 13, 838088. doi: 10.3389/fpls.2022.838088

Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. *Nature* 456, 720–723. doi: 10.1038/nature07629

Pang, Y., Liu, C., Wang, D., St Amand, P., Bernardo, A., Li, W., et al. (2020). High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. *Mol. Plant* 13, 1311–1327. doi: 10.1016/j.molp.2020.07.008

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qu, P., Wang, J., Wen, W., Gao, F., Liu, J., Xia, X., et al. (2021). Construction of consensus genetic map with applications in gene mapping of wheat (*Triticum aestivum* L.) using 90K SNP array. *Front. Plant Sci.* 12, 727077. doi: 10.3389/fpls.2021.727077

Sajjad, M., Ma, X., Habibullah Khan, S., Shoaib, M., Song, Y., Yang, W., et al. (2017). *TaFlo2-A1*, an ortholog of rice *Flo2*, is associated with thousand grain weight in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17, 164. doi: 10.1186/s12870-017-1114-3

Shan, D., Ali, M., Shahid, M., Arif, A., Waheed, M. Q., Xia, X., et al. (2022). Genetic networks underlying salinity tolerance in wheat uncovered with genome-wide analyses and selective sweeps. *Theor. Appl. Genet.* 135, 2925–2941. doi: 10.1007/s00122-022-04153-5

Sukumaran, S., Lopes, M., Dreisigacker, S., and Reynolds, M. (2018). Genetic analysis of multi-environmental spring wheat trials identifies genomic regions for locus-specific trade-offs for grain weight and grain number. *Theor. Appl. Genet.* 131, 985–998. doi: 10.1007/s00122-017-3037-7

Sun, H., Zhang, W., Wu, Y., Gao, L., Cui, F., Zhao, C., et al. (2020). The circadian clock gene, *TaPRR1*, is associated with yield-related traits in wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 11, 285. doi: 10.3389/fpls.2020.00285

Su, Q., Zhang, X., Zhang, W., Zhang, N., Song, L., Liu, L., et al. (2018). QTL detection for kernel size and weight in bread wheat (*Triticum aestivum* L.) using a high-density SNP and SSR-based linkage map. *Front. Plant Sci.* 9, 1484. doi: 10.3389/fpls.2018.01484

Tshikunde, N. M., Mashilo, J., Shimelis, H., and Odindo, A. (2019). Agronomic and physiological traits, and associated quantitative trait loci (QTL) affecting yield response in wheat (*Triticum aestivum* L.): A review. *Front. Plant Sci.* 10, 1428. doi: 10.3389/fpls.2019.01428

Wang, L., Ge, H., Hao, C., Dong, Y., and Zhang, X. (2012). Identifying loci influencing 1,000-kernel weight in wheat by microsatellite screening for evidence of selection during breeding. *PloS One* 7, e29432. doi: 10.1371/journal.pone.0029432

Wang, X., Guan, P., Xin, M., Wang, Y., Chen, X., Zhao, A., et al. (2021). Genome-wide association study identifies QTL for thousand grain weight in winter wheat under normal- and late-sown stressed environments. *Theor. Appl. Genet.* 134, 143–157. doi: 10.1007/s00122-020-03687-w

Wang, Y., Hou, J., Liu, H., Li, T., Wang, K., Hao, C., et al. (2019). *TaBT1*, affecting starch synthesis and thousand kernel weight, underwent strong selection during wheat improvement. *J. Exp. Bot.* 70, 1497–1511. doi: 10.1093/jxb/erz032

Wang, J., Liu, W., Wang, H., Li, L., Wu, J., Yang, X., et al. (2011). QTL mapping of yield-related traits in the wheat germplasm 3228. *Euphytica* 177, 277–292. doi: 10.1007/s10681-010-0267-z

Wang, P., Tian, T., Ma, J., Liu, Y., Zhang, P., Chen, T., et al. (2022). Genome-wide association study of kernel traits using a 35K SNP array in bread wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 13, 905660. doi: 10.3389/fpls.2022.905660

Wang, S., Yan, X., Wang, Y., Liu, H., Cui, D., and Chen, F. (2016). Haplotypes of the *TaGS5-A1* gene are associated with thousand-kernel weight in Chinese bread wheat. *Front. Plant Sci.* 7, 783. doi: 10.3389/fpls.2016.00783

Wang, J., and Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinf.* 19, 629–640. doi: 10.1016/j.gpb.2021.08.005

Wolde, G. M., Trautewig, C., Mascher, M., and Schnurbusch, T. (2019). Genetic insights into morphometric inflorescence traits of wheat. *Theor. Appl. Genet.* 132, 1661–1676. doi: 10.1007/s00122-019-03305-4

Wu, X., Cheng, R., Xue, S., Kong, Z., Wan, H., Li, G., et al. (2014). Precise mapping of a quantitative trait locus interval for spike length and grain weight in bread wheat (*Triticum aestivum* L.). *Mol. Breed.* 33, 129–138. doi: 10.1007/s11032-013-9939-4

Yang, L., Zhao, D., Meng, Z., Xu, K., Yan, J., Xia, X., et al. (2020). QTL mapping for grain yield-related traits in bread wheat *via* SNP-based selective genotyping. *Theor. Appl. Genet.* 133, 857–872. doi: 10.1007/s00122-019-03511-0

Yao, F. Q., Li, X. H., Wang, H., Song, Y. N., Li, Z. Q., Li, X. G., et al. (2021). Down-expression of *TaPIN1s* increases the tiller number and grain yield in wheat. *BMC Plant Biol.* 21, 443. doi: 10.1186/s12870-021-03217-w

Yu, K., Liu, D., Chen, Y., Wang, D., Yang, W., Yang, W., et al. (2019). Unraveling the genetic architecture of grain size in einkorn wheat through linkage and homology mapping and transcriptomic profiling. *J. Exp. Bot.* 70, 4671–4688. doi: 10.1093/jxb/erz247

Yu, M., Mao, S. L., Chen, G. Y., Pu, Z. E., Wei, Y. M., and Zheng, Y. L. (2014). QTLs for uppermost internode and spike length in two wheat RIL populations and their affect upon plant height at an individual QTL level. *Euphytica* 200, 95–108. doi: 10.1007/s10681-014-1156-7

Zhang, X., Deng, Z., Wang, Y., Li, J., and Tian, J. (2014a). Unconditional and conditional QTL analysis of kernel weight related traits in wheat (*Triticum aestivum* L.) in multiple genetic backgrounds. *Genetica* 142, 371–379. doi: 10.1007/s10709-014-9781-6

Zhang, Y., Liu, J., Xia, X., and He, Z. (2014b). *TaGS-D1*, an ortholog of rice *OsGS3*, is associated with grain weight and grain length in common wheat. *Mol. Breed.* 34, 1097–1107. doi: 10.1007/s11032-014-0102-7

Zhang, L., Zhao, Y. L., Gao, L. F., Zhao, G. Y., Zhou, R. H., Zhang, B. S., et al. (2012). *TaCKX6-D1*, the ortholog of rice *OsCKX2*, is associated with grain weight in hexaploid wheat. *New Phytol.* 195, 574–584. doi: 10.1111/j.1469-8137.2012.04194.x

Zhao, J., Wang, S., Qin, J., Sun, C., and Liu, F. (2020). The lipid transfer protein OsLTPL159 is involved in cold tolerance at the early seedling stage in rice. *Plant Biotechnol. J.* 18, 756–769. doi: 10.1111/pbi.13243

Zheng, B. S., Le Gouis, J., Leflon, M., Rong, W. Y., Laperche, A., and Brancourt-Hulmel, M. (2010). Using probe genotypes to dissect QTL × environment interactions for grain yield components in winter wheat. *Theor. Appl. Genet.* 121, 1501–1517. doi: 10.1007/s00122-010-1406-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 23 February 2023
DOI 10.3389/fpls.2023.1134308

OPEN ACCESS

EDITED BY
Yue Feng,
China National Rice Research Institute (CAAS), China

REVIEWED BY
Kuldeep Kumar,
Indian Institute of Pulses Research (ICAR), India
Fan Zhang,
Institute of Crop Sciences (CAAS), China
Qingyun Bu,
CAS, China
Jinsong Bao,
Zhejiang University, China

*CORRESPONDENCE
Junhe Wang
wangjunhe@haas.cn
Hui Jiang
hui@haas.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 30 December 2022
ACCEPTED 10 February 2023
PUBLISHED 23 February 2023

CITATION
Jiang S, Zhang X, Yang X, Liu C, Wang L, Ma B, Miao Y, Hu J, Tan K, Wang Y, Jiang H and Wang J (2023) A chromosome-level genome assembly of an early matured aromatic *Japonica* rice variety Qigeng10 to accelerate rice breeding for high grain quality in Northeast China.
Front. Plant Sci. 14:1134308.
doi: 10.3389/fpls.2023.1134308

A chromosome-level genome assembly of an early matured aromatic *Japonica* rice variety Qigeng10 to accelerate rice breeding for high grain quality in Northeast China

Shukun Jiang[1,2,3†], Xijuan Zhang[2,3,4†], Xianli Yang[2,3,4†], Chuanzeng Liu[1,3†], Lizhi Wang[2,3,4], Bo Ma[1,3], Yi Miao[1,3], Jifang Hu[1,3], Kefei Tan[1,3], Yuxian Wang[1,3], Hui Jiang[5*] and Junhe Wang[2,4*]

[1]Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar, China, [2]Heilongjiang Provincial Key Laboratory of Crop Physiology and Ecology in Cold Region, Heilongjiang Provincial Engineering Technology Research Center of Crop Cold Damage, Harbin, China, [3]Northeast Branch of National Salt-Alkali Tolerant Rice Technology Innovation Center, Harbin, China, [4]Crop Cultivation and Tillage Institute of Heilongjiang Academy of Agricultural Sciences, Harbin, China, [5]Keshan Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar, China

Early-matured aromatic *japonica* rice from the Northeast is the most popular rice commodity in the Chinese market. The Qigeng10 (QG10) was one of the varieties with the largest planting area in this region in recent years. It was an early-matured *japonica* rice variety with a lot of superior traits such as semi-dwarf, lodging resistance, long grain, aromatic and good quality. Therefore, a high-quality assembly of Qigeng10 genome is critical and useful for japonica research and breeding. In this study, we produced a high-precision QG10 chromosome-level genome by using a combination of Nanopore and Hi-C platforms. Finally, we assembled the QG10 genome into 77 contigs with an N50 length of 11.80 Mb in 27 scaffolds with an N50 length of 30.55 Mb. The assembled genome size was 378.31Mb with 65 contigs and constituted approximately 99.59% of the 12 chromosomes. We identified a total of 1,080,819 SNPs and 682,392 InDels between QG10 and Nipponbare. We also annotated 57,599 genes by the Ab initio method, homology-based technique, and RNA-seq. Based on the assembled genome sequence, we detected the sequence variation in a total of 63 cloned genes involved in grain yield, grain size, disease tolerance, lodging resistance, fragrance, and many other important traits. Finally, we identified five elite alleles ($qTGW2^{Nipponbare}$, $qTGW3^{Nanyangzhan}$, $GW5^{IR24}$, $GW6^{Suyunuo}$, and $qGW8^{Basmati385}$) controlling long grain size, four elite alleles ($COLD1^{Nipponbare}$, $bZIP73^{Nipponbare}$, $CTB4a^{Kunmingxiaobaigu}$, and $CTB2^{Kunmingxiaobaigu}$) controlling cold tolerance, three non-functional alleles ($DTH7^{Kitaake}$, $Ghd7^{Hejiang19}$, and $Hd1^{Longgeng31}$) for early heading, two resistant alleles ($Pia^{Akihikari}$ and $Pid4^{Digu}$) for rice blast, a resistant allele $STV11^{Kasalath}$ for rice stripe virus, an $NRT1.1B^{IR24}$ allele for

higher nitrate absorption activity, an elite allele *SCM3*^{Chugoku117} for stronger culms, and the typical aromatic gene *badh2-E2* for fragrance in QG10. These results not only help us to better elucidate the genetic mechanisms underlying excellent agronomic traits in QG10 but also have wide-ranging implications for genomics-assisted breeding in early-matured fragrant *japonica* rice.

KEYWORDS

aromatic *japonica* rice, genome assembly, early-matured, northern limit region, functional genes

Introduction

Rice (*Oryza sativa* L.) is a safe and staple food source for more than half of the world's population and serves as a model plant for cereal genetic studies (Gross and Zhao, 2014). Novo sequencing and genomic technologies have been widely applied in rice to promote the shift of breeding schemes from conventional field selection to genomic-assisted breeding (Gu et al., 2022). *O. sativa subsp. japonica*/*Geng* and *subsp. Indica*/*Xian* are the two major subspecies of cultivated rice (Zhang et al., 2016a; Nie et al., 2017). The *japonica*/*Geng* rice planting area is 9.87 million ha, accounting for approximately 32.9 percent of the total rice planting area in China (Tang and Chen, 2021). Recently, the early-matured *japonica*/*Geng* rice is becoming more and more important, and its growing area was more than 4 million ha in Northeast China (Cui et al., 2022). Two genome draft sequences of the cultivated rice subspecies *japonica*/*Geng* Nipponbare and *Indica*/*Xian* 93-11 were released in 2002 (Goff et al., 2002; Yu et al., 2002). In 2005, the International Rice Genome Sequencing Project (IRGSP) published the first completed version of the Nipponbare sequence (International-Rice-Genome-Sequencing-Project, 2005). Over the last two decades, several pan-genomes including 66 rice genomes (Zhao et al., 2018b), 33 rice genomes (Qin et al., 2021), 111 rice genomes (Zhang et al., 2022a),251 rice genomes (Shang et al., 2022), and 12 *japonica* rice genome (Wang et al., 2023) were built including IR64, R498, Zhenshan 97, Minghui 63, Taichung Native 1, LTH, Kitaake, IR8, N22, Huajingxian74, HR12, Basmati 334, Dom Sufid, Huazhan and Tianfeng at the chromosome level, and Shennong265, DJ123, WR04-6, Suijing18, Koshihikari, Basmati, Kongyu-131, and Guangluai-4 at scaffold level have been assembled and released with unprecedented speed (Mahesh et al., 2016; Zhang et al., 2016b; Du et al., 2017; Li et al., 2017; Nie et al., 2017; Stein et al., 2018; Zhao et al., 2018b; Jain et al., 2019; Choi et al., 2020; Tanaka et al., 2020; Panibe et al., 2021; Li et al., 2021a; Yang et al., 2022; Zhang et al., 2022b). These assembled genome sequences will be helpful in pinpointing new causal variants that underlie complex agronomic traits and identifying many of the genome-specific loci that were absent from the Nipponbare reference genome. However, most of these varieties are *Indica*/*Xian* rice or landrace. Nevertheless, the genome of *japonica*/*Geng* differs significantly from that of *indica*/*Xian* (Nie et al., 2017). Since the release of the finished version genome of Nipponbare, only seven genomes at the scaffold level of early-matured *japonica*/*Geng* varieties in northern region of China including Shennong265, Liaogeng5, Yanfeng47, Suijing18, Longgeng31, Daohuaxiang2 (Wuyoudao4), and Kongyu-131 were released (Nie et al., 2017; Li et al., 2018; Zhao et al., 2018b; Wang et al., 2023). Only Daohuaxiang2 and Suijing18 were belong to early-mature aromatic type. The public availability of *japonica*/*Geng* genomes at the chromosome level, especially for the early-mature aromatic type, remains largely blank (Nie et al., 2017). Moreover, a few genomes are not enough to represent the whole genomic content of the *japonica* rice. The novo assembled genomes of an early-mature aromatic variety would be advantageous for functional genomics and genome research. For example, if there is structural variation in the particular variety and the reference genome in the candidate region, the guiding role of the reference genome would be limited. So, there is still a need for *de novo* genome assembly for various purposes especially in early-mature aromatic rice breeding research.

Fragrant and long grain are key grain quality traits that directly influence the global market price of rice (Hui et al., 2022). The basmati rice and jasmine rice are the two most popular fragrant *indica* rice in the world. However, consumers from East Asia, including North China, Japan, and Korea tend to prefer *japonica* rice (Lu et al., 2022). So, the aromatic long grain rice from Northeast China, represented by Wuyoudao4 (WYD4), is the most famous rice in the Chinese market. WYD4 had a superior quality, but also had a number of defects, most importantly, poor lodging resistance, lack of cold tolerance and blast resistance, and late maturity (Gao et al., 2012). In 2019, Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences developed Qigeng10 (QG10) to solved these defects of WYD4. The plant area of QG10 was 0.4 million ha in the recent three years. QG10 has been a major variety of early matured aromatic long grain rice in Northeast China. The construction of a high-quality chromosome-level genome of QG10 is very important for improving the efficiency of rice genetic mechanism studies for desirable agronomic traits, such as eating quality, cold tolerance, lodging tolerance and early maturity, as well as accelerating the process of high-quality rice breeding in cold region of northeast China by design (Li et al., 2021a). Here, we produced a high-

precision QG10 chromosomal genome by performing whole-genome sequencing in the Nanopore platform (Lin et al., 2021), followed by the Hi-C-assisted assembly mount technology (Van Berkum et al., 2010). Our results provided several functionally important candidate alleles for the grain length, cold tolerance, early heading, disease resistance, lodging resistance, and nitrate-use of rice breeding in cold region of northeast China.

Materials and methods

Materials

The early-matured aromatic long-grain *japonica* rice variety QG10, which was developed by our own group, was licensed for release in 2019 and is now widely planted in Heilongjiang province in Northeast China. It was a semi-dwarf rice variety with a lot of superior traits such as long panicle, long grain, aromatic, and good quality (Figures 1A-D). It was selected from the cross between two aromatic *japonica*/*Geng* rice Wuyoudao4 (WYD4) and Suigeng4 (SG4) (Figure 1E). The seedlings of QG10 were grown on the agricultural farm of the Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences. Field management practices were performed according to the most commonly followed agricultural practices of local farmers. The leaves, stems, roots, and panicles at heading stages from plants grown in the experimental station were collected in liquid nitrogen for isolating RNA. The young leaves of a single young plant were used to isolate genomic DNA.

Oxford Nanopare sequencing and genome assembly

The high molecular weight genomic DNA of QG10 was extracted from the 15-day-old leaf tissues following a modified CTAB method. Whole genome sequencing was done following the standard instructions of the Ligation Sequencing Kit (Nanopore, Oxford shire, UK). The quantified DNA was randomly sheared, and fragments of ~20 kb were enriched and purified. Then, a 20-kb library was constructed and sequenced on the Nanopore PromethION platform according to the manufacturer's protocols (Jiang et al., 2020).

De novo genome assembly of Nanopore sequence was performed as follow: The raw Nanopore reads were error-corrected and assembled using CANU (v1.7.1) (Koren et al., 2017), followed by Smartdenovo (https://github.com/ruanjue/smartdenovo) assembly, followed by three rounds of polishing with Racon (Vaser et al., 2017), followed by three rounds with Pilon v0.3.0 using the Illumina PCR-free paired-end reads (Walker et al., 2014). Genome completeness was also assessed using the algae dataset of BUSCO v2.0 (Simão et al., 2015).

Hi-C library construction and sequencing

A Hi-C fragment library with a 300-700 bp insert size was constructed from the genomic DNA of the QG10. Briefly, adapter sequences of raw reads were trimmed and low-quality PE reads were removed for clean data. The library was sequenced on the Illumina HiSeq4000^{TM} platform by Biomarker Technologies (Beijing, China). The clean Hi-C reads were first truncated at the putative Hi-C junctions and then the resulting trimmed reads were aligned to the assembly results with the software package bwa aligner (Li and Durbin, 2009). Only uniquely alignable pair reads whose mapping quality of more than 20 retained for further analysis. Invalid read pairs, including dangling-end and self-cycle, re-ligation, and dumped products, were filtered by the software package HiC-Pro v2.8.1 (Servant et al., 2015). The 96.25% of unique mapped read pairs were valid interaction pairs and were used for the correction of scaffolds and clustered, ordered, and orientated

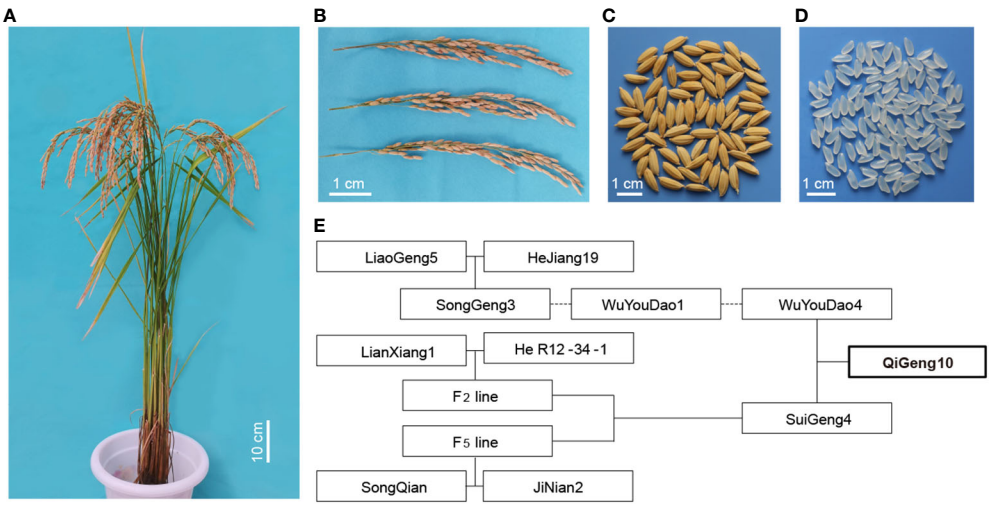


FIGURE 1
The of plant type, panicle phenotype, seeds phenotype, grains phenotype and pedigree of QG10. **(A)** The plant type of QG10; **(B)** The panicle morphology of QG 10; **(C)** The seed morphology of QG10; **(D)** The grain morphology of QG10; **(E)** The pedigree of QG10.

scaffolds onto chromosomes by the software package LACHESIS (Burton et al., 2013).

Genome assembly and Hi-C scaffolding

Before chromosome assembly, we first performed a preassembly for error correction of scaffolds which required the splitting of scaffolds into segments of 50 kb on average. The Hi-C data were mapped to these segments using the software package BWA (version 0.7.10-r789) (Li and Durbin, 2009). The uniquely mapped data were retained to perform assembly by using the software package LACHESIS (Burton et al., 2013). Any two segments which showed an inconsistent connection with information from the raw scaffold were checked manually. These corrected scaffolds were then assembled with the software package LACHESIS. After this step, placement and orientation errors exhibiting obvious discrete chromatin interaction patterns were manually adjusted.

Gene prediction and genome annotation

The RNA of QG10 were isolated from the mixed tissues (leaves, culms, roots, and panicles) following the manufacturer's protocol (Wang et al., 2023). We then performed the sequencing on the Illumina HiSeq 2500 platform according to the manufacturer's instructions. The repetitive sequence of the genome based on the principle of structure prediction and *de-novo* prediction was constructed with the software package LTR_FINDER v1.05 (Xu and Wang, 2007) and the software package RepeatScout v1.0.5 (Price et al., 2005). The PASTE Classifier was used to classify the database (Hoede et al., 2014). Then it was merged with the database of Repbase as the final repeat sequence database (Jurka et al., 2005). And then the software package RepeatMasker v4.0.6 was used to predict the repeat sequence of the QG10 genome based on the constructed repetitive sequence database (Tarailo-Graovac and Chen, 2009). the software packages Genscan (Burge and Karlin, 1997), Augustus v2.4 (Stanke and Waack, 2003), GlimmerHMM v3.0.4 (Majoros et al., 2004), GeneID v1.4 (Alioto et al., 2018), and SNAP (Korf, 2004) were used for *de-novo* prediction. The software package GeMoMa v1.3.1 was used for prediction based on homologous species (Keilwagen et al., 2016; Keilwagen et al., 2018). The software packages Hisat v2.0.4 (Kim et al., 2015) and Stringtie v1.2.3 (Pertea et al., 2015) were used for assembly based on reference transcripts, and the software packages TransDecoder v2.0 and GeneMarkS-T v5.1 (Tang et al., 2015) were used for gene prediction. The software package PASA v2.0.2 was used to predict Unigene sequences without reference based on transcriptome data (Campbell et al., 2006). Finally, the software package EVM v1.1.1 (Haas et al., 2008) was used to integrate the prediction results obtained by the above three methods. The predicted gene sequences were compared with NR, KOG, GO, KEGG, TrEMBL, and other functional databases by the software package BLAST v2.2.31 (-evalue 1e-5) (Altschul et al., 1990) to perform KEGG pathway, KOG function, GO function and other genes functional annotation analysis.

Identification of genomic sequence variation in important genes

The whole-genome assemblies sequences of QG10 were compared with the rice reference genome sequence (Oryza_sativa_MSU7 version) using the software package MUMmer v3 (Kurtz et al., 2004). According to the results from the software package MUMmer, the sequence variations and SVs were further re-called using the software package BLAST. The synteny/inversion comparison were analysis by using GenomeSyn_Win.v1 (Zhou et al., 2022). At the site of each sequence variant, the genotypic information for QG10, Nipponbare, and the elite variety having important genes was called according to the results of the one-to-one alignments. The allelic information of sequence variants was detected based on gff files from the Oryza_sativa_MSU7 version. The software packages ClustalW v1.8.3(Thompson et al., 1994) and BLAST v2.2.31 were used for re-detected the sequence variations and detailed haplotype analyses for the well-characterized genes in rice (Zhao et al., 2018b).

Results

Nanopore sequencing and genome assembly

We sequenced QG10 genomic DNA to generate about 73.52Gb of Nanopore sequencing raw data. After data quality control, the clean data volume was 63.23Gb containing 3,279,893 reads with a total of 166.34-fold sequencing depth. The reads with 10 -20 kb and 20 - 30 kb sequencing length were account for 51.56% (Table 1). The mean reads length of clean sequencing data was 19.28 kb with an N50 length of 25.64 kb. The clean Nanopore sequencing data was re-corrected with the software package Canu (Koren et al., 2017). Then the third-generation sequencing data was re-corrected with the software package Racon (Vaser et al., 2017) for three rounds. Then, the second-generation data were used for three rounds of correction by Pilon (Walker et al., 2014) software, and the stain was removed according to NT alignment. Finally, we obtained a 380.15 Mb genome sequence with the contig N50 was 12.24 Mb. The completeness estimated by Benchmarking Universal Single-Copy Orthologs (BUSCO) was 98.12%.

Genome assembly and Hi-C scaffolding

We conducted the assembly in a stepwise fashion following a previously reported approach (Jiang et al., 2020). The Hi-C sequencing raw data was filtered, and the splice sequences and low-quality reads

TABLE 1 The Reads length distribution of Nanopore sequencing clean data.

Length	ReadsNum	TotalLength	Percent	AveLength
2000-5000	390,870	1,364,569,685	2.15%	3,491.10
5001-10000	506,104	3,711,926,879	5.87%	7,334.31
10001-20000	1,160,503	17,449,235,975	27.59%	15,035.92
20001-30000	620,853	15,157,158,521	23.97%	24,413.44
30001-40000	319,208	10,994,112,170	17.38%	34,441.84
40001-50000	158,764	7,047,880,953	11.14%	44,392.18
50001-60000	73,874	4,015,077,789	6.34%	54,350.35
60001-70000	31,197	2,004,190,092	3.16%	64,243.03
70001-80000	11,756	872,009,657	1.37%	74,175.71
>80000	6,764	616,167,771	0.97%	91,095.17

Length: each length range of reads; ReadsNum: the number of sequences in each length range; TotalLength: indicates the total length of sequences in each length. Percent: indicates the proportion of the number of sequences in each length range to the total number of sequences. AveLength: Average length of sequences in each length range.

were removed to obtain high-quality clean data. The mapped data was obtained by sequence alignment of clean data with the preliminarily assembled genome. Finally, effective Hi-C data were used for further assembly of the draft genome sequence. LACHESIS software was used for clustering, sorting, and orientating the preliminary assembled genome sequence, and finally, the genomic sequence at the chromosome level was obtained. Finally, we assembled the genome (QG10) into 77 contigs with an N50 length of 11.80 Mb in 27 scaffolds with an N50 length of 30.55 Mb. The assembled genome size was 378.31Mb and the 65 contigs constituted approximately 99.59% of the whole genome. Visualization of the Hi-C signals indicated that 12 square matrix areas in the Hi-C heat map displayed significant differences from the background. These scaffolds were anchored into chromosomes 1–12, respectively (Figure 2A; Table 2).

According to the whole genome comparison, the genome of QG10 showed four sequence inversions with a length of about 0.5-2 Mb compared with the Nipponbare genome at the position of about 14-16 Mb on chromosome 4, 30-31 Mb on chromosome 5, 5.5-6 Mb on chromosome 8, and 5.5-6 Mb on chromosome 10 (Figure 2B). We identified 1,080,819 SNPs and 682,392 InDels between QG10 and Nipponbare on 12 chromosomes (Figures 2C, D).

Genome annotation and repeat analysis

We annotated the repeat regions in our QG10 assembly by Repeat Masker and detected 492,503 repetitive regions with 177.52 Mb repeat length that contained 242,211 Class I retrotransposons, 223,439 Class II DNA transposons, 833 Potential Host Gene, and 2,222 simple sequence repeats (Table 3). The repeat regions make up 46.7% of the QG10 assembly genome.

We predicted 39,465 genes by the Ab initio method, 56,999 genes by the Homology-based method, and 24,998 by RNA-seq. Finally, a total of 57,599 genes were integrated with the prediction results obtained by the above three methods by using the software package EVM v1.1.1 (Table 4). A total of 723 tRNA, 306 rRNA, 194 miRNA, and 5,392 pseudogenes were also predicted. 94.11% of the genes could be annotated into NR, GO, KOG, KEGG, and other databases (Table 5).

Sequence variants of the genes controlling grain length

We investigated the sequence variations in 19 cloned genes controlling grain size, including *GW2* (Song et al., 2007), *GS2* (Hu et al., 2015), *qTGW2* (Ruan et al., 2020), *BG1* (Liu et al., 2015), *OsLG3* (Yu et al., 2017), *OsLG3b* (Yu et al., 2018), *GS3* (Fan et al., 2006), *qTGW3* (Hu et al., 2018), *GS5* (Li et al., 2011), *GW5* (Weng et al., 2008), *GS6* (Sun et al., 2013), *GW6* (Shi et al., 2020), *TGW6* (Ishimaru et al., 2013), *GW6a* (Song et al., 2015), *GL6* (Wang et al., 2019), *GLW7* (Si et al., 2016), *GW7* (Wang et al., 2015), *qGW8* (Wang et al., 2012), and *GS9* (Zhao et al., 2018a), to explain the long grain of QG10. A total of five grain size elite alleles ($qTGW2^{Nipponbare}$, $qTGW3^{Nanyangzhan}$, $GW5^{IR24}$, $GW6^{Suyunuo}$, and $qGW8^{Basmati385}$) were identified controlling grain size in QG10 (Figure 3). The *qTGW2* allele in QG10 was identical to Nipponbare having the key variants (G/A) at -1818 bp in the promoter region (Figure 3A). The *qTGW3* allele in QG10 had the key splicing-site mutation as Nanyangzhan, which was a long-grain *indica* rice (Figure 3B). The *GW5* allele in QG10 was found without the critical loss of the 1,212 bp deletion mutation as IR24 a long narrow grain *indica* rice (Figure 3C). The *GW6* allele in QG10 had the key mutation (6 bp) as Suyunuo, which was a wider grain *indica* rice (Figure 3D). The *qGW8* allele in QG10 had the five variants as Basmati385, a long narrow grain *indica* rice (Figure 3E). Tracing the origin of these genes, it was found that *qTGW3* (Nanyangzhan type), *GW5* (IR24 type), *GW6* (Suyunuo type), and *qGW8* (Basmati385 type) belonged to *indica* subspecies, while *qTGW2* (Nipponbare type) were mainly derived from *japonica* subspecies. In addition, these genes controlling grain type are all rare genotypes in *japonica* rice and have important application value in long grain type *japonica* rice breeding.

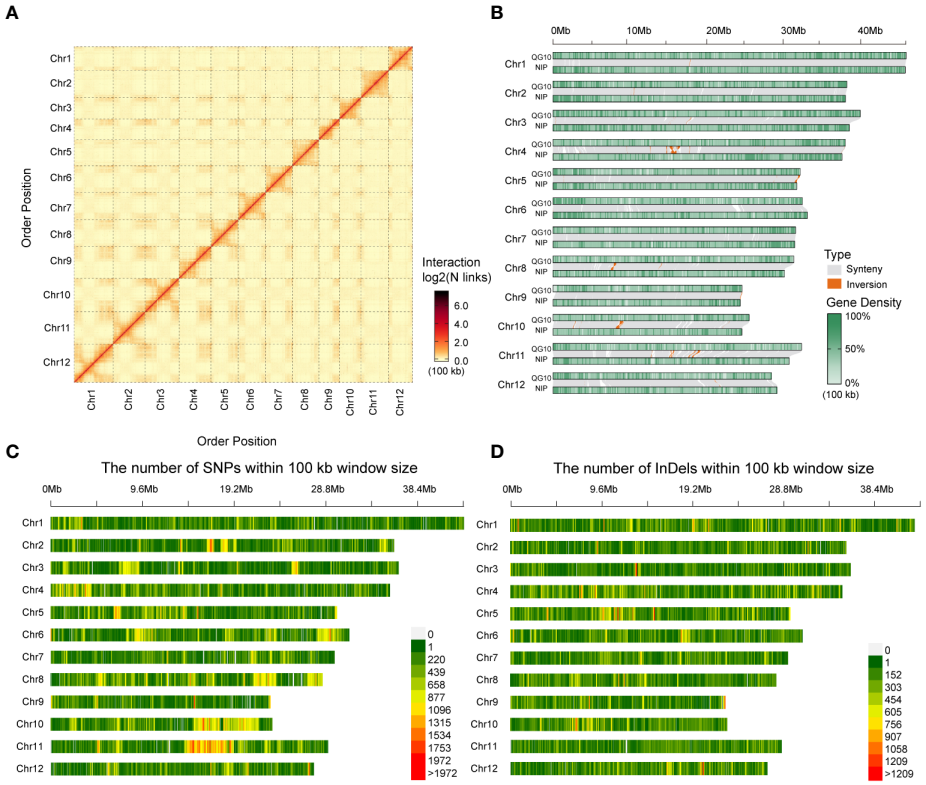


FIGURE 2
Characteristics of Qigeng10 (QG10) genome and synteny examining, SNPs and InDels mining, Synteny/Inversion comparison with Nipponbare. **(A)**, Visualization of the Hi-C signals in the whole genome of QG10. **(B)**, Chromosomal synteny among QG10 and Nipponbare reference genomes of rice. **(C)**, Distribution of SNPs between QG10 and Nipponbare. **(D)**, Distribution of InDels between QG10 and Nipponbare.

TABLE 2 The statistics information of Hi-C assembly data.

Group	Cluster Num	Cluster Len	Order Num	Order Len
Chr1	4	43,382,979	4	43,382,979
Chr2	4	36,173,705	3	36,097,194
Chr3	6	37,757,262	6	37,757,262
Chr4	4	35,918,009	4	35,918,009
Chr5	4	30,375,008	4	30,375,008
Chr6	4	30,649,707	4	30,649,707
Chr7	5	29,809,605	5	29,809,605
Chr8	8	30,022,072	6	29,599,970
Chr9	7	23,219,349	7	23,219,349
Chr10	10	24,170,793	9	24,102,177
Chr11	9	30,667,739	7	30,544,885
Chr12	12	27,708,367	6	26,854,597
Total(Ratio %)	77(96.25)	379854595(99.92)	65(84.42)	378310742(99.59)

TABLE 3 The statistics information of repeat regions in QG10 assembly.

Type	Number	Length	Rate (%)
Class I	242,211	123,543,619	32.50
Class I/DIRS	5,336	6,334,743	1.67
Class I/LARD	87,163	17,713,596	4.66
Class I/LINE	20,531	5,686,323	1.50
Class I/LTR/Copia	25,386	14,961,322	3.94
Class I/LTR/Gypsy	64,824	68,053,651	17.90
Class I/LTR/Unknown	24,079	13,886,683	3.65
Class I/PLE	1,264	547,500	0.14
Class I/SINE	11,311	2,290,288	0.60
Class I/TRIM	1,493	457,017	0.12
Class I/Unknown	824	603,012	0.16
Class II	223,439	57,701,667	15.18
Class II/Crypton	22	1,816	0.00
Class II/Helitron	21,248	6,000,821	1.58
Class II/MITE	18,642	3,713,793	0.98
Class II/Maverick	256	45,395	0.01
Class II/TIR	123,736	34,429,300	9.06
Class II/Unknown	59,535	16,608,349	4.37
Potential Host Gene	833	431,292	0.11
SSR	2,222	754,226	0.20
Unknown	23,798	5,059,225	1.33
Total	492,503	177,517,691	46.70

Type: repeat sequence type, DIRS: Dictyostelium intermediate repeat sequence; LARD: large retrotransposon derivative; LINE: long interspersed nuclear element; LTR: long terminal repeat; PLE: Penelope-like element; SINE: short interspersed nuclear element; TRIM: terminal-repeat retrotransposons in miniature; MITE: miniature inverted-repeat transposable element; TIR: terminal inverted repeat; SSR: simple sequence repeat. Number: the number of repeats obtained; Length: total length of the predicted repeat sequence; Rate (%): the proportion of repeats in the total genome.

Sequence variants of the genes controlling cold tolerance

Cold tolerance is the key agricultural trait controlling rice production and geographic distribution. We investigated the sequence variations in 10 cloned genes controlling cold tolerance, including *bZIP73* (Liu et al., 2018), *COLD1* (Ma et al., 2015), *Ctb1* (Saito et al., 2010), *CTB2* (Li et al., 2021b), *CTB4a* (Zhang et al., 2017), *HAN1* (Mao et al., 2019), *ltt1* (Xu et al., 2020), *OsLTPL159* (Zhao et al., 2020), *qLTG3-1* (Fujino et al., 2008), and *qPSR10* (Xiao et al., 2018), to explain the high cold tolerance of QG10. A total of four elite alleles ($COLD1^{Nipponbare}$, $bZIP73^{Nipponbare}$, $CTB4a^{Kunmingxiaobaigu}$, and $CTB2^{Kunmingxiaobaigu}$) were identified as controlling cold tolerance in QG10 (Figure 4). The *COLD1* allele in QG10 was identical to Nipponbare to have the key SNP in the fourth exon region (Figure 4A). The *bZIP73* allele in QG10 was found having the key SNP mutation (G/A) as Nipponbare (Figure 4B). The *CTB4a* allele in QG10 was found to have the ten mutations as Kunmingxiaobaigu, which was a cold tolerance variety from Yunnan Province (Figure 4C). The *CTB2* allele in QG10 was found also have the ten key SNPs as Kunmingxiaobaigu (Figure 4D). The *COLD1* (Nipponbare type) and *bZIP73* (Nipponbare type) all belong to *japonica* subspecies. The *CTB4a* (Kunmingxiaobaigu type) and *CTB2* (Kunmingxiaobaigu type) were all rare alleles in Northeast *japonica* rice and have important application value in Northeast *japonica* rice breeding.

Sequence variants of the genes controlling early heading

The heading date is one of the most important factors determining rice distribution and the final yield. We investigated the sequence variations in 11 cloned genes related to early heading under long-day conditions, including *DTH7* (Gao et al., 2014), *Ghd7* (Xue et al., 2008), *Ghd8* (Yan et al., 2011), *Ehd1* (Doi et al., 2004), *Ehd3* (Matsubara et al., 2011), *Ehd4* (Gao et al., 2013), *Hd1* (Yano et al., 2000), *Hd3a* (Kojima et al., 2002), *Hd6* (Takahashi

TABLE 4 The statistics information of the predicted genes.

Method	Software	Species	Gene number
Ab initio	Genscan	–	36,995
	Augustus	–	33,891
	GlimmerHMM	–	72,642
	GeneID	–	51,130
	SNAP	–	48,464
Homology-based	GeMoMa	Oryza_sativa_9311	47,215
		Oryza_sativa_MSU7	60,679
		Oryza_sativa_rapdb	38,019
		Oryza_sativa_R498	52,424
		Arabidopsis_thaliana	22,123
RNAseq	TransDecoder	–	57,661
	GeneMarkS-T	–	30,053
	PASA	–	34,939
Integration	EVM	–	57,599

Method: The strategy used in gene prediction; Software: Software used for gene prediction; Species: Species. Gene number: The number of genes predicted.

et al., 2001), *Hd16* (Hori et al., 2013), and *Hd17* (Matsubara et al., 2012). Among these heading date genes, only the *DTH7*, *Ghd7*, and *Hd1* haplotypes were found to have non-functional alleles. The *DTH7* allele in QG10 was found to have the three mutations as Kitaake, which originated at the northern limit of rice cultivation in Hokkaido, Japan (Figure 5A). Kitaake is reported insensitive to day length, short in stature, and completes its life cycle in about 9 weeks (Jain et al., 2019). The *Ghd7* allele in QG10 was found having the critical mutations as Hejiang19, which is an early-maturity rice variety in Heilongjiang Province (Figure 5B). The *Hd1* allele in QG10 was found to have the non-functional allele as Longgeng31, which is the major plant rice variety in Heilongjiang Province (Figure 5C). These results indicated that the heading gene combinations of *Hd1*, *DTH7*, and *Ghd7* determined the early heading in QG10 in the northernmost province of China.

Sequence variants of the genes controlling disease resistance

Blast is one of the most devastating rice diseases in Heilongjiang Province. We investigated the sequence variations in 14 cloned rice blast-resistant genes, including *Pi5* (Lee et al., 2009), *Pi21* (Fukuoka et al., 2009), *Pi36* (Liu et al., 2007), *Pi37* (Lin et al., 2007), *Pi54* (Sharma et al., 2010), *Pi56* (Liu et al., 2013), *Pia* (Okuyama et al., 2011), *Pish* (Takahashi et al., 2010), *Pit* (Hayashi and Yoshida, 2009), *Pita* (Bryan

TABLE 5 The statistics information of the annotated genes.

Annotation database	Annotated number	Percentage (%)
COG_Annotation	12,901	22.40
GO_Annotation	38,790	67.34
KEGG_Annotation	9,038	15.69
KOG_Annotation	20,498	35.59
Pfam_Annotation	32,058	55.66
Swissprot_Annotation	24,932	43.29
TrEMBL_Annotation	54,193	94.09
nr_Annotation	54,007	93.76
All_Annotated	54,206	94.11

Annotation database: the annotation database. Annotated number: the number of genes annotated to the corresponding database; Percentage (%): indicates the percentage of genes annotated to the database.

FIGURE 3
The allelic information of sequence variants in qTGW2 **(A)**, qTGW3 **(B)**, GW5 **(C)**, GW6 **(D)**, and qGW8 **(E)** controlling grain size.

et al., 2000), *Pigm* (Deng et al., 2017), *Pid2* (Chen et al., 2006), *Pid3* (Shang et al., 2009), and *Pid4* (Chen et al., 2018), to explain the blast resistance. Finally, only two blast resistance genes, *Pia* and *Pid4*, were found in QG10. The *Pia* allele in QG10 was found to have the resistant genotype as Akihikari, which encodes a nucleotide-binding site (NBS) and a C-terminal leucine-rich repeat (LRR) domain protein (Figure 6A). The *Pid4* allele in QG10 was found to have the resistant genotype as Digu, which encodes a coiled-coil nucleotide-binding site leucine-rich repeat (CC-NBS-LRR) protein (Figure 6B). The *Pia* (Akihikari type) was the major blast-resistant gene in Northeast China. The *Pid4* (Digu type) was a rare allele in *japonica* rice and has important application value in Northeast *japonica* rice breeding.

Rice stripe virus (RSV), an RNA virus belonging to the genus *Tenuivirus* and transmitted by small brown planthoppers, causes one of the most destructive rice diseases (Wang et al., 2014). RSV has become more and more serious in Heilongjiang province in recent years. But, almost the majority of *japonica* varieties cultivated in Heilongjiang are highly susceptible to RSV (Wang et al., 2014). The *STV11* was the first cloned resistant gene of RSV, which encodes a sulfotransferase (*OsSOT1*) catalyzing the conversion of salicylic acid (SA) into sulphonated SA (SSA) (Wang et al., 2014). The *STV11* allele in QG10 was found to have the resistant genotype as Kasalath, which is a high-resistance *indica* landrace (Figure 6C). The $STV11^{QG10}$ was a useful resistant gene in *japonica* rice breeding.

Sequence variants of fragrance, fertilizer use efficiency and other yield related genes

Fragrant rice is popular among consumers worldwide because its market price is much higher than that of nonfragrant rice. Fragrant was found to be controlled by *BADH2* in rice (Chen et al., 2008). The *BADH2* in QG10 was the typical *badh2-E2* type of a 7-bp deletion as

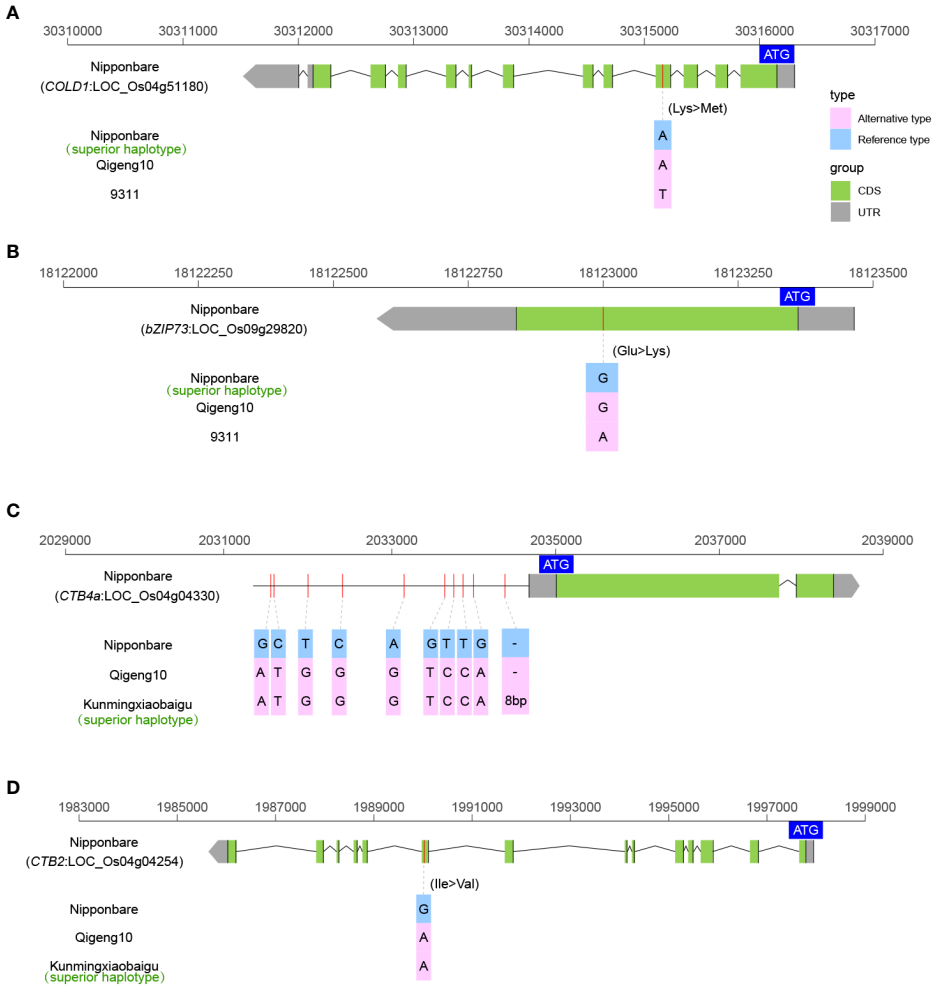


FIGURE 4
The allelic information of sequence variants in COLD1 **(A)**, bZIP73 **(B)**, CTB4a **(C)**, and CTB2 **(D)** controlling cold tolerance in rice.

DHX2 that caused fragrance (Figure 7A). We also investigated the sequence variations in three cloned grain number genes (*Gn1a* (Ashikari et al., 2005), *GNP1* (Wu et al., 2016), and *SPIKE* (Fujita et al., 2013)), and three lodging-resistance genes (*Sd1* (Spielmeyer et al., 2002), *SCM2* (Ookawa et al., 2010), and *SCM3* (Yano et al., 2014)). Only the *SCM3* allele in QG10 was found to have the key allele as Chugoku117, which is a stronger culms rice variety in Japan (Figure 7B). There was no elite grain number gene found in QG10. *NRT1.1B* is a key gene controlling the nitrogen-use efficiency (NUE) in rice (Hu et al., 2015). The *NRT1.1B* allele in QG10 was found to have the *indica* variation, which has higher nitrate absorption activity (Figure 7C).

Discussion

Indica genome introgression and large SVs were found in Qigeng10

Japonica/*Geng* and *Indica*/*Xian* are the two *major* subspecies of Asian cultivated rice (Zhang et al., 2016a). Owing to long-term differentiation and adaptation, both *Indica* and *Japonica* rice contain many favorable genes. Therefore, combining the favorable genes of the two subspecies has great value for creating genotypes with greater yield potential, stronger stress resistance, and better quality (Gu, 2010). Over the past 50 years, the combination of plant ideotypes and favorable vigor through hybridization between *indica* and *japonica* rice has greatly contributed to yield improvements in modern *japonica* rice in Northeast China (Tang and Chen, 2021). In recent years, a series of high-yielding and good-quality *japonica* cultivars have been obtained from hybridization of *Indica*/*Japonica* and the cultivation area of them was more than 4 million ha in Northeast China (Cui et al., 2022). The new fragrant early *japonica* rice cultivar QG10 was derived from a cross between 'Wuyoudao4 and Suigeng4', which were all derived from the hybridization of *Indica*/*Japonica*. In recently, Wang et al. (2023) chose six interrelated modern Chinese temperate *japonica* varieties and six related Japanese *japonica* varieties to investigate genome enhancement in temperate *japonica* varieties during modern breeding. They found many large SVs in Zhonghua11 (ZH11), Liaogeng5 (LG5), and Daohuaxiang2 (DHX2/WYD4). These large-fragment in the same location introgression from *indica* were also found in QG10 on chromosomes 4, 8, and 10 (Figure 2B). Several

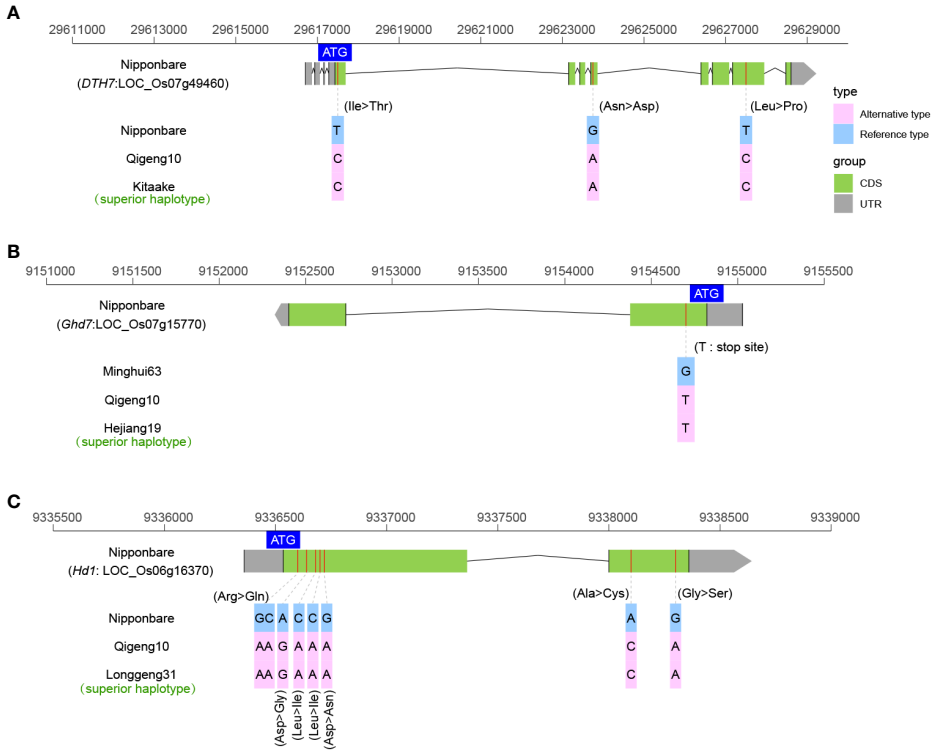


FIGURE 5
The allelic information of sequence variants in DTH7 **(A)**, Ghd7 **(B)**, and Hd1 **(C)** controlling heading date in rice.

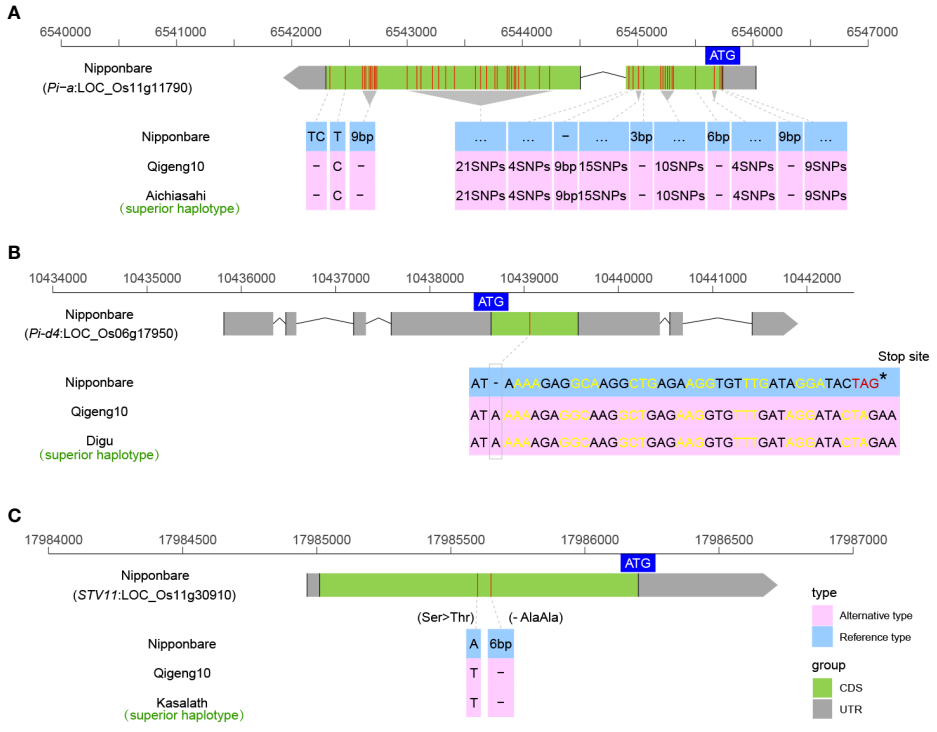


FIGURE 6
The allelic information of sequence variants in Pia **(A)**, Pid4 **(B)**, and STV11 **(C)** controlling disease resistance in rice.

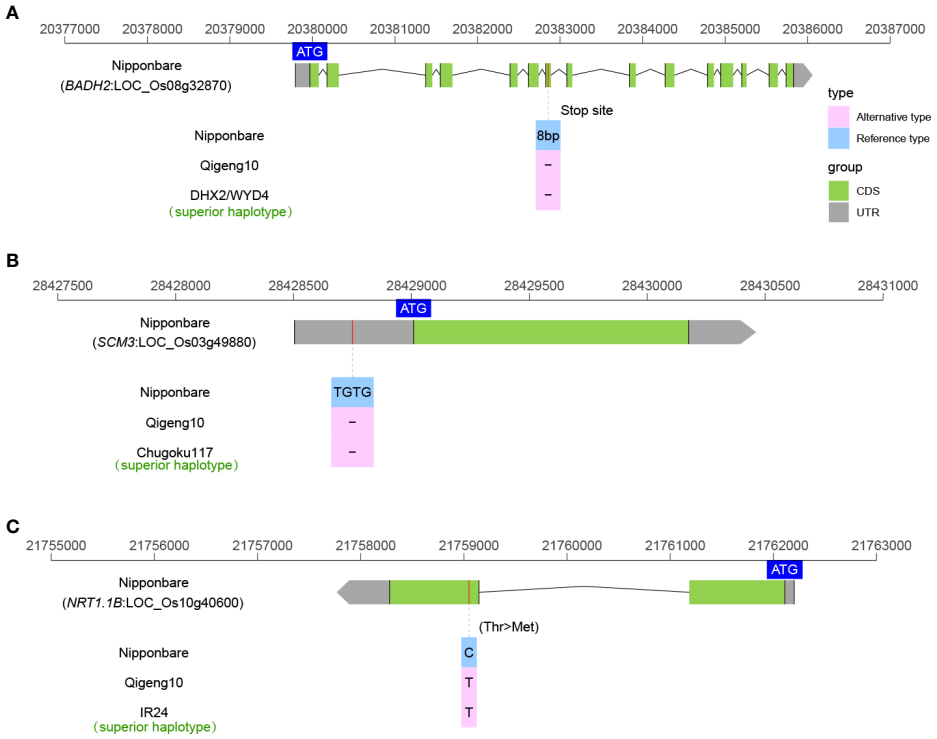


FIGURE 7
The allelic sequence variants of BADH2 **(A)**, SCM3 **(B)**, and NRT1.1B **(C)** in rice.

indica superior alleles including $qTGW3^{Nanyangzhan}$, $GW5^{IR24}$, $GW6^{Suyunuo}$, $qGW8^{Basmati385}$, $Pid4^{Digu}$, $STV11^{Kasalath}$, $NRT1.1B^{IR24}$, and *badh2-E2* were also be found in QG10. This information indicated that the *indica* genome introgression was common in the modern temperate *japonica* rice breeding in Northeast China. Superior alleles were found in Qigeng10 and had important value for breeding in Northeast China

Superior alleles were found in Qigeng10 and had important value for breeding in Northeast China

Greater yield potential, stronger stress resistance, and better quality (longer grain and fragrant) are key agronomy traits that directly influence the market price of rice. Consumers in East Asia, including North China, Japan, and Korea tend to prefer longer fragrant *japonica* rice (Lu et al., 2022). So, the longer fragrant *japonica* rice from Northeast China, represented by Daohuaxiang2 (DHX2/WYD4), is the most famous rice in the Chinese market. QG10 was derived from DHX2 and solved some defects of DHX2 including poor lodging resistance, lack of cold tolerance, weak blast resistance, and late maturity. Therefore, the construction of a high-quality genome of 'QG10' is essential for further improvement of this cultivar or its progenies, as well as accelerating the process of fragrant *japonica* rice breeding, by providing genomic resources that could be directly applied to fragrant *japonica* rice cultivars. In this study, we found five superior alleles ($qTGW2^{Nipponbare}$, $qTGW3^{Nanyangzhan}$, $GW5^{IR24}$, $GW6^{Suyunuo}$, and $qGW8^{Basmati385}$) controlling long grain size in QG10. To compare the phenotype of different gene haplotypes in rice germplasm, we investigated the grain shape traits and days to heading of 3k cultivars in the website (https://www.rmbreeding.cn/) (Wang et al., 2020). The results showed that the functional haplotype of QG10 controlling longer and slider grain (Figures 8A-E) and early heading (Figure 8f-h). Most of them was belong to the rare alleles for controlling longer grain and were less application in rice breeding in Northeast China. The blast resistant alleles ($Pid4^{Digu}$), RSV allele ($STV11^{Kasalath}$), and NUE allele ($NRT1.1B^{IR24}$) were also belong to the rare alleles for rice breeding in Northeast China. In the future, we will develop molecular assisted markers for the improvement of *japonica* rice varieties in Northeast China, and expanded the gene pool of *japonica* rice in Northeast China.

Conclusions

In this study, we present chromosome-level genome assembly of an early-matured aromatic long-grain *japonica* rice variety Qigeng10 by using a combination of Nanopore and Hi-C platforms. The total assembly size is 378.31Mb with an N50 length of 30.55 Mb. A total of 18 superior haplotypes including five long-grain alleles ($qTGW2^{Nipponbare}$, $qTGW3^{Nanyangzhan}$, $GW5^{IR24}$, $GW6^{Suyunuo}$, and $qGW8^{Basmati385}$), four cold tolerant alleles ($COLD1^{Nipponbare}$, $bZIP73^{Nipponbare}$, $CTB4a^{Kunmingxiaobaigu}$, and $CTB2^{Kunmingxiaobaigu}$),

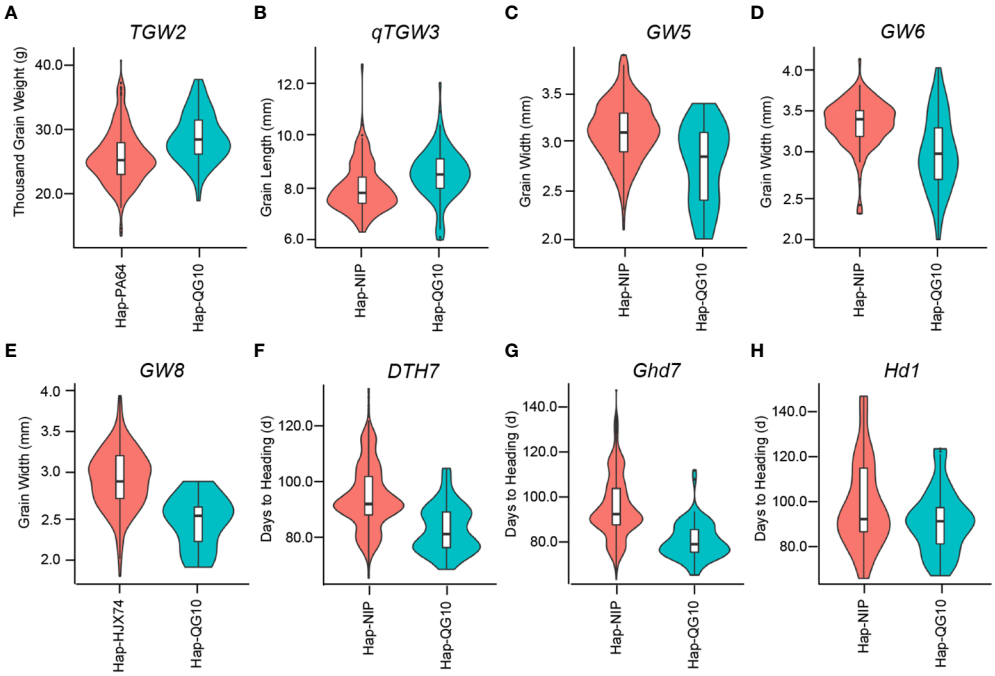


FIGURE 8
Comparison of grain shape traits and days to heading between different gene haplotypes in 3k panel.

three non-functional heading date alleles (*DTH7*^{Kitaake}, *Ghd7*^{Hejiang19}, and *Hd1*^{Longgeng31}), two blast resistant alleles (*Pia* ^{Akihikari} and *Pid4*^{Digu}), a rice stripe virus resistant allele *STV11*^{Kasalath}, a higher nitrate absorption allele *NRT1.1B*^{IR24}, a lodging resistant allele *SCM3*^{Chugoku117}, and the typical aromatic allele *badh2-E2*, were identified in QG10. This information will accelerate the process of fragrant *japonica* rice breeding in Northeast China, by providing genomic resources that could be directly applied to fragrant *japonica* rice cultivars or development of molecular assisted markers for the improvement of *japonica* rice varieties.

Data availability statement

The data presented in the study are deposited in the Data Center of Beijing Institute of Genomics(Big) repository, accession number WGS029943 (PRJCA013131; SAMC988458).

Author contributions

Conceptualization: SJ and JW. Methodology: SJ. Software: XZ, XY, BM. Validation: SJ, HJ and YW. Formal analysis: JH and KT. Investigation: CL and LW. Resources: JW. Data curation: SJ. Writing—original draft preparation: SJ. Writing—review and editing: SJ. Visualization: SJ. Project administration: SJ. Funding acquisition: SJ and XZ. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Fundamental Research Funds for the Research Institutes of Heilongjiang Province, Grant/Award Number: CZKYF2021B009, CZKYF2022-1-B004, National Natural Science Foundation of China, Grant/Award Number: 32071889, Key Research and Development Program of Heilongjiang Province, Grant/Award Number: CZ20210090, and Heilongjiang Province Agricultural Science and Technology Innovation Project, Grant/Award Number: 2021QKPY009, 2021CQJC003.

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

References

Alioto, T., Blanco, E., Parra, G., and Guigó, R. (2018). Using geneid to identify genes. *Curr. Protoc. Bioinf.* 64, e56. doi: 10.1002/cpbi.56

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., et al. (2005). Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745. doi: 10.1126/science.1113373

Bryan, G., Wu, K., Farrall, L., Jia, Y., Hershey, H., Mcadams, S., et al. (2000). A single amino acid difference distinguishes resistant and susceptible alleles of the rice blast resistance gene pi-ta. *Plant Cell* 12, 2033–2046. doi: 10.2307/3871103

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94. doi: 10.1006/jmbi.1997.0951

Burton, J., Adey, A., Patwardhan, R., Qiu, R., Kitzman, J., and Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt.2727

Campbell, M., Haas, B., Hamilton, J., Mount, S., and Buell, C. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with arabidopsis. *BMC Genomics* 7, 327. doi: 10.1186/1471-2164-7-327

Chen, X., Shang, J., Chen, D., Lei, C., Zou, Y., Zhai, W., et al. (2006). A b-lectin receptor kinase gene conferring rice blast resistance. *Plant J.* 46, 794–804. doi: 10.1111/j.1365-313X.2006.02739.x

Chen, S., Yang, Y., Shi, W., Ji, Q., He, F., Zhang, Z., et al. (2008). Badh2, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. *Plant Cell* 20, 1850–1861. doi: 10.1105/tpc.108.058917

Chen, Z., Zhao, W., Zhu, X., Zou, C., Yin, J., Chern, M., et al. (2018). Identification and characterization of rice blast resistance gene Pid4 by a combination of transcriptomic profiling and genome analysis. *J. Genet. Genomics* 45, 663–672. doi: 10.1016/j.jgg.2018.10.007

Choi, J., Lye, Z., Groen, S., Dai, X., Rughani, P., Zaaijer, S., et al. (2020). Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. *Genome Biol.* 21, 21. doi: 10.1186/s13059-020-1938-2

Cui, D., Zhou, H., Ma, X., Lin, Z., Sun, L., Han, B., et al. (2022). Genomic insights on the contribution of introgressions from Xian/Indica to the genetic improvement of Geng/Japonica rice cultivars. *Plant Commun.* 3, 100325. doi: 10.1016/j.xplc.2022.100325

Deng, Y., Zhai, K., Xie, Z., Yang, D., Zhu, X., Liu, J., et al. (2017). Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* 355, 962–965. doi: 10.1126/science.aai8898

Doi, K., Izawa, T., Fuse, T., Yamanouchi, U., Kubo, T., Shimatani, Z., et al. (2004). Ehd1, a b-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. *Genes Dev.* 18, 926–936. doi: 10.1101/gad.1189604

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., et al. (2017). Sequencing and *de novo* assembly of a near complete indica rice genome. *Nat. Commun.* 8, 15324. doi: 10.1038/ncomms15324

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* 112, 1164–1171. doi: 10.1007/s00122-006-0218-1

Fujino, K., Sekiguchi, H., Matsuda, Y., Sugimoto, K., Ono, K., and Yano, M. (2008). Molecular identification of a major quantitative trait locus, qLTG3-1, controlling low-temperature germinability in rice. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12623–12628. doi: 10.1073/pnas.0805303105

Fujita, D., Trijatmiko, K., Tagle, A., Sapasap, M., Koide, Y., Sasaki, K., et al. (2013). NAL1 allele from a rice landrace greatly increases yield in modern indica cultivars. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20431–20436. doi: 10.1073/pnas.1310790110

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., et al. (2009). Loss of function of a proline-containing protein confers durable disease resistance in rice. Aug 21. *Science* 325, 998–1001. doi: 10.1126/science.1175550

Gao, H., Jin, M., Zheng, X., Chen, J., Yuan, D., Xin, Y., et al. (2014). Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16337–16342. doi: 10.1073/pnas.1418204111

Gao, H., Zhao, B., Zhang, J., Song, L., Wu, H., Yu, Y., et al. (2012). High-quality and high-yield cultivation techniques of a new high-quality rice variety wuyoudao no. 4. *Heilongjiang Agric. Sci.* 154-155.

Gao, H., Zheng, X., Fei, G., Chen, J., Jin, M., Ren, Y., et al. (2013). Ehd4 encodes a novel and oryza-genus-specific regulator of photoperiodic flowering in rice. *PloS Genet.* 9, e1003281. doi: 10.1371/journal.pgen.1003281

Goff, S., Ricke, D., Lan, T., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (Oryza sativa l. ssp. japonica). *Science* 296, 92–100. doi: 10.1126/science.1068275

Gross, B., and Zhao, Z. (2014). Archaeological and genetic insights into the origins of domesticated rice. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6190–6197. doi: 10.1073/pnas.1308942110

Gu, H. (2010). Discussion on the aspects of high-yielding breeding in rice. *Acta Agronomica Sin.* 36, 1431–1439.

Gu, H., Liang, S., and Zhao, J. (2022). Novel sequencing and genomic technologies revolutionized rice genomic study and breeding. *Agronomy* 12, 218. doi: 10.3390/agronomy12010218

Haas, B., Salzberg, S., Zhu, W., Pertea, M., Allen, J., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9, R7. doi: 10.1186/gb-2008-9-1-r7

Hayashi, K., and Yoshida, H. (2009). Refunctionalization of the ancient rice blast disease resistance gene pit by the recruitment of a retrotransposon as a promoter. *Plant J.* 57, 413–425. doi: 10.1111/j.1365-313X.2008.03694.x

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., et al. (2014). PASTEC: an automatic transposable element classification tool. *PloS One* 9, e91929. doi: 10.1371/journal.pone.0091929

Hori, K., Ogiso-Tanaka, E., Matsubara, K., Yamanouchi, U., Ebana, K., and Yano, M. (2013). Hd16, a gene for casein kinase I, is involved in the control of rice flowering time by modulating the day-length response. *Plant J.* 76, 36–46. doi: 10.1111/tpj.12268

Hu, Z., Lu, S., Wang, M., He, H., Sun, L., Wang, H., et al. (2018). A novel QTL qTGW3 encodes the GSK3/SHAGGY-like kinase OsGSK5/OsSK41 that interacts with OsARF4 to negatively regulate grain size and weight in rice. *Mol. Plant* 11, 736–749. doi: 10.1016/j.molp.2018.03.005

Hu, J., Wang, Y., Fang, Y., Zeng, L., Xu, J., Yu, H., et al. (2015). A rare allele of GS2 enhances grain size and grain yield in rice. *Mol. Plant* 8, 1455–1465. doi: 10.1016/j.molp.2015.07.002

Hu, B., Wang, W., Ou, S., Tang, J., Li, H., Che, R., et al. (2015). Variation in NRT1.1B contributes to nitrate-use divergence between rice subspecies. *Nat. Genet.* 47, 834–838. doi: 10.1038/ng.3337

Hui, S., Li, H., Mawia, A., Zhou, L., Cai, J., Ahmad, S., et al. (2022). Production of aromatic three-line hybrid rice using novel alleles of BADH2. *Plant Biotechnol. J.* 20, 59–74. doi: 10.1111/pbi.13695

International-Rice-Genome-Sequencing-Project (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800. doi: 10.1038/nature03895

Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., et al. (2013). Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* 45, 707–711. doi: 10.1038/ng.2612

Jain, R., Jenkins, J., Shu, S., Chern, M., Martin, J., Copetti, D., et al. (2019). Genome sequence of the model rice variety KitaakeX. *BMC Genomics* 20, 905. doi: 10.1186/s12864-019-6262-4

Jiang, S., An, H., Xu, F., and Zhang, X. (2020). Chromosome-level genome assembly and annotation of the loquat (Eriobotrya japonica) genome. *Gigascience* 9, giaa015. doi: 10.1093/gigascience/giaa015

Jurka, J., Kapitonov, V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic Genome Res.* 110, 462–467. doi: 10.1159/000084979

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinf.* 19, 189. doi: 10.1186/s12859-018-2203-5

Keilwagen, J., Wenk, M., Erickson, J., Schattat, M., Jan, G., and Frank, H. (2016). Using intron position Conserv. homology-based Gene prediction. *Nucleic Acids Res.* 44, e89. doi: 10.1093/nar/gkw092

Kim, D., Langmead, B., and Salzberg, S. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., et al. (2002). Hd3a, a rice ortholog of the arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol.* 43, 1096–1105. doi: 10.1093/pcp/pcf156

Koren, S., Walenz, B., Berlin, K., Miller, J., Bergman, N., and Phillippy, A. (2017). Canu: scalable and accurate long-read assembly *via* adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinf.* 5. doi: 10.1186/1471-2105-5-59

Kurtz, S., Phillippy, A., Delcher, A., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12. doi: 10.1186/gb-2004-5-2-r12

Lee, S. K., Song, M. Y., Seo, Y. S., Kim, H. K., KO., S., Cao, P. J., et al. (2009). Rice Pi5-mediated resistance to magnaporthe oryzae requires the presence of two coiled-coil-nucleotide-binding-leucine-rich repeat genes. *Genetics* 181, 1627–1638. doi: 10.1534/genetics.108.099226

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43, 1266–1269. doi: 10.1038/ng.977

Li, F., Gao, Y., Wu, B., Cai, Q., Zhan, P., Yang, W., et al. (2021a). High-quality *de novo* genome assembly of huajingxian 74, a receptor parent of single segment substitution lines. *Rice Sci.* 28, 109–113. doi: 10.1016/j.rsci.2020.09.010

Li, G., Jain, R., Chern, M., Pham, N., Martin, J., Wei, T., et al. (2017). The sequences of 1504 mutants in the model rice variety kitaake facilitate rapid functional genomic studies. *Plant Cell Physiol.* 29, 1218–1231. doi: 10.1105/tpc.17.00154

Li, X., Wu, L., Wang, J., Sun, J., Xia, X., Geng, X., et al. (2018). Genome sequencing of rice subspecies and genetic analysis of recombinant lines reveals regional yield- and quality-associated loci. *BMC Biol.* 16, 102. doi: 10.1186/s12915-018-0572-x

Li, J., Zeng, Y., Pan, Y., Zhou, L., Zhang, Z., Guo, H., et al. (2021b). Stepwise selection of natural variations at CTB2 and CTB4a improves cold adaptation during domestication of japonica rice. *New Phytol.* 231, 1056–1072. doi: 10.1111/nph.17407

Lin, F., Chen, S., Que, Z., Wang, L., Liu, X., and Pan, Q. (2007). The blast resistance gene Pi37 encodes a nucleotide binding site leucine-rich repeat protein and is a member of a resistance gene cluster on rice chromosome 1. *Genetics* 177, 1871–1880. doi: 10.1534/genetics.107.080648

Lin, B., Hui, J., and Mao, H. (2021). Nanopore technology and its applications in gene sequencing. *Biosensors* 11, 214. doi: 10.3390/bios11070214

Liu, X., Lin, F., Wang, L., and Pan, Q. (2007). The in silico map-based cloning of Pi36, a rice coiled-coil nucleotide-binding site leucine-rich repeat gene that confers race-specific resistance to the blast fungus. *Genetics* 176, 2541–2549. doi: 10.1534/genetics.107.075465

Liu, Y., Liu, B., Zhu, X., Yang, J., Bordeos, A., Wang, G., et al. (2013). Fine-mapping and molecular marker development for Pi56(t), a NBS-LRR gene conferring broad-spectrum resistance to magnaporthe oryzae in rice. *Theor. Appl. Genet.* 126, 985–998. doi: 10.1007/s00122-012-2031-3

Liu, C., Ou, S., Mao, B., Tang, J., Wang, W., Wang, H., et al. (2018). Early selection of bZIP73 facilitated adaptation of japonica rice to cold climates. *Nat. Commun.* 9, 3302. doi: 10.1038/s41467-018-05753-w

Liu, L., Tong, H., Xiao, Y., Che, R., Xu, F., Hu, B., et al. (2015). Activation of big Grain1 significantly improves grain size by regulating auxin transport in rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11102–11107. doi: 10.1073/pnas.1512748112

Lu R, L. J., Wang, X., Song, Z., Ji, X., Li, N., Ma, G., et al. (2022). Chromosome-level genome assembly of a fragrant japonica rice cultivar 'Changxianggeng 1813' provides insights into genomic variations between fragrant and non-fragrant japonica rice. *Int. J. Mol. Sci.* 23, 9705. doi: 10.3390/ijms23179705

Ma, Y., Dai, X., Xu, Y., Luo, W., Zheng, X., Zeng, D., et al. (2015). COLD1 confers chilling tolerance in rice. *Cell* 160, 1209–1221. doi: 10.1016/j.cell.2015.01.046

Mahesh, H., Shirke, M., Singh, S., Rajamani, A., Hittalmani, S., Wang, G., et al. (2016). Indica rice genome assembly, annotation and mining of blast disease resistance genes. *BMC Genomics* 242. doi: 10.1186/s12864-016-2523-7

Majoros, W., Pertea, M., and Salzberg, S. (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315

Mao, D., Xin, Y., Tan, Y., Hu, X., Bai, J., Liu, Z., et al. (2019). Natural variation in the HAN1 gene confers chilling tolerance in rice and allowed adaptation to a temperate climate. *Proc. Natl. Acad. Sci. U.S.A.* 116, 3494–3501. doi: 10.1073/pnas.1819769116

Matsubara, K., Ogiso-Tanaka, E., Hori, K., Ebana, K., Ando, T., and Yano, M. (2012). Natural variation in Hd17, a homolog of arabidopsis ELF3 that is involved in rice photoperiodic flowering. *Plant Cell Physiol.* 53, 709–716. doi: 10.1093/pcp/pcs028

Matsubara, K., Yamanouchi, U., Nonoue, Y., Sugimoto, K., Wang, Z., Minobe, Y., et al. (2011). Ehd3, encoding a plant homeodomain finger-containing protein, is a critical promoter of rice flowering. *Plant J.* 66, 603–612. doi: 10.1111/j.1365-313X.2011.04517.x

Nie, S., Liu, Y., Wang, C., Gao, S., Xu, T., Liu, Q., et al. (2017). Assembly of an early-matured japonica (Geng) rice genome, Suijing18, ased on PacBio and illumina sequencing. *Sci. Data* 170195. doi: 10.1038/sdata.2017.195

Okuyama, Y., Kanzaki, H., Abe, A., Yoshida, K., Tamiru, M., Saitoh, H., et al. (2011). A multifaceted genomics approach allows the isolation of the rice pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. *Plant J.* 66, 467–479. doi: 10.1111/j.1365-313X.2011.04502.x

Ookawa, T., Hobo, T., Yano, M., Murata, K., Ando, T., Miura, H., et al. (2010). New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. *Nat. Commun.* 1, 132. doi: 10.1038/ncomms1132

Panibe, J., Wang, L., Li, J., Li, M., Lee, Y., Wang, C., et al. (2021). Chromosomal-level genome assembly of the semi-dwarf rice taichung native 1, an initiator of green revolution. *Genomics* 113, 2656–2674. doi: 10.1016/j.ygeno.2021.06.006

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, Sl (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Price, A., Jones, N., and Pa., P. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatics* 21, i351–i358. doi: 10.1093/bioinformatics/bti1018

Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* 184, 3542–3558. doi: 10.1016/j.cell.2021.04.046

Ruan, B., Shang, L., Zhang, B., Hu, J., Wang, Y., Lin, H., et al. (2020). Natural variation in the promoter of TGW2 determines grain width and weight in rice. *New Phytol.* 227, 629–640. doi: 10.1111/nph.16540

Saito, K., Hayano-Saito, Y., Kuroki, M., and Sato, Y. (2010). Map-based cloning of the rice cold tolerance gene Ctb1. *Plant Sci.* 179, 97–102. doi: 10.1016/j.plantsci.2010.04.004

Servant, N., Varoquaux, N., Lajoie, B., Viara, E., Chen, C., Vert, J., et al. (2015). HiC-pro: An optimized and flexible pipeline for Hi-c data processing. *Genome Biol.* 16, 259. doi: 10.1186/s13059-015-0831-x

Shang, L., Li, X., He, H., Yuan, Q., Song, Y., Wei, Z., et al. (2022). A super pan-genomic landscape of rice. *Cell Res.* 32, 878–896. doi: 10.1038/s41422-022-00685-z

Shang, J., Tao, Y., Chen, X., Zou, Y., Lei, C., Wang, J., et al. (2009). Identification of a new rice blast resistance gene, Pid3, by genomewide comparison of paired nucleotide-binding site–leucine-rich repeat genes and their pseudogene alleles between the two sequenced rice genomes. *Genetics* 182, 1303–1311. doi: 10.1534/genetics.109.102871

Sharma, T., Rai, A., Gupta, S., and Singh, N. (2010). Broad-spectrum blast resistance gene pi-kh cloned from rice line tetep designated as Pi54. *J. Plant Biochem. Biotechnol.* 19, 87–89. doi: 10.1007/BF03323441

Shi, C., Dong, N., Guo, T., Ye, W., Shan, J., and Lin, H. (2020). A quantitative trait locus GW6 controls rice grain size and yield through the gibberellin pathway. *Plant J.* 103, 1174–1188. doi: 10.1111/tpj.14793

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., et al. (2016). OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* 48, 447–456. doi: 10.1038/ng.3518

Simão, F., Waterhouse, R., Ioannidis, P., Kriventseva, E., and Zdobnov, E. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Song, X., Huang, W., Shi, M., Zhu, M., and Lin, H. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39, 623–630. doi: 10.1038/ng2014

Song, X., Kuroha, T., Ayano, M., Furuta, T., Nagai, K., Komeda, N., et al. (2015). Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 76–81. doi: 10.1073/pnas.1421127112

Spielmeyer, W., Ellis, M., and Chandler, P. (2002). Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9043–9048. doi: 10.1073/pnas.132266399

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19, ii215–ii225. doi: 10.1093/bioinformatics/btg1080

Stein, J., Yu, Y., Copetti, D., Zwickl, D., Zhang, L., Zhang, C., et al. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus oryza. *Nat. Genet.* 50, 285–296. doi: 10.1038/s41588-018-0040-0

Sun, L., Li, X., Fu, Y., Zhu, Z., Tan, L., Liu, F., et al. (2013). GS6, a member of the GRAS gene family, negatively regulates grain size in rice. *J. Integr. Plant Biol.* 55, 938–949. doi: 10.1111/jipb.12062

Takahashi, A., Hayashi, N., Miyao, A., and Hirochika, H. (2010). Unique features of the rice blast resistance pish locus revealed by large scale retrotransposon-tagging. *BMC Plant Biol.* 10, 175. doi: 10.1186/1471-2229-10-175

Takahashi, Y., Shomura, A., Sasaki, T., and Yano, M. (2001). Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7922–7927. doi: 10.1073/pnas.111136798

Tanaka, T., Nishijima, R., Teramoto, S., Kitomi, Y., Hayashi, T., Uga, Y., et al. (2020). *De novo* genome assembly of the indica rice variety IR64 using linked-read sequencing and nanopore sequencing. *G3 (Bethesda)* 10, 1495–1501. doi: 10.1534/g3.119.400871

Tang, L., and Chen, W. (2021). Development trend and prospect of geng rice in northeast China. *China Rice*, 28(5), 1–4.

Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* 43, e78. doi: 10.1093/nar/gkv227

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* 4 (10), 11–14. doi: 10.1002/0471250953.bi0410s25

Thompson, J., Higgins, D., and Gibson, T. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Van Berkum, N., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., et al. (2010). Hi-C: A method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* 39, 1869. doi: 10.3791/1869

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116

Walker, B., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* 9, e112963. doi: 10.1371/journal.pone.0112963

Wang, A., Hou, Q., Si, L., Huang, X., Luo, J., Lu, D., et al. (2019). The PLATZ transcription factor GL6 affects grain length and number in rice. *Plant Physiol.* 180, 2077–2090. doi: 10.1104/pp.18.01574

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., et al. (2015). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* 47, 949–954. doi: 10.1038/ng.3352

Wang, Y., Li, F., Zhang, F., Wu, L., Xu, N., Sun, Q., et al. (2023). Time-ordering japonica/geng genomes analysis indicates the importance of large structural variants in rice breeding. *Plant Biotechnol. J.* 21, 202–218. doi: 10.1111/pbi.13938

Wang, Q., Liu, Y., He, J., Zheng, X., Hu, J., Liu, Y., et al. (2014). STV11 encodes a sulphotransferase and confers durable resistance to rice stripe virus. *Nat. Commun.* 5. doi: 10.1038/ncomms5768

Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., et al. (2012). Control of grain size, shape and quality by OsSPL16 in rice. *Nat. Genet.* 44, 950–954. doi: 10.1038/ng.2327

Wang, C., Yu, H., Huang, J., Wang, W., Faruquee, M., Zhang, F., et al. (2020). Towards a deeper haplotype mining of complex traits in rice with RFGB v2.0. *Plant Biotechnol. J.* 18, 14–16. doi: 10.1111/pbi.13215

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., et al. (2008). Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* 18, 1199–1209. doi: 10.1038/cr.2008.307

Wu, Y., Wang, Y., Mi, X., Shan, J., Li, X., Xu, J., et al. (2016). The QTL GNP1 encodes GA20ox1, which increases grain number and yield by increasing cytokinin activity in rice panicle meristems. *PloS Genet.* 12, e1006386. doi: 10.1371/journal.pgen.1006386

Xiao, N., Gao, Y., Qian, H., Gao, Q., Wu, Y., Zhang, D., et al. (2018). Identification of genes related to cold tolerance and a functional allele that confers cold tolerance. *Plant Physiol.* 177, 1108–1123. doi: 10.1104/pp.18.00209

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286

Xu, Y., Wang, R., Wang, Y., Zhang, L., and Yao, S. (2020). A point mutation in LTT1 enhances cold tolerance at the booting stage in rice. *Plant Cell Environ.* 43, 992–1007. doi: 10.1111/pce.13717

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., et al. (2008). Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat. Genet.* 40, 761–767. doi: 10.1038/ng.143

Yan, W., Wang, P., Chen, H., Zhou, H., Li, Q., Wang, C., et al. (2011). A major QTL, Ghd8, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Mol. Plant* 4, 319–330. doi: 10.1093/mp/ssq070

Yang, L., Zhao, M., Sha, G., Sun, Q., Gong, Q., Yang, Q., et al. (2022). The genome of the rice variety LTH provides insight into its universal susceptibility mechanism to worldwide rice blast fungal strains. *Comput. Struct. Biotechnol. J.* 20, 1012–1026. doi: 10.1016/j.csbj.2022.01.030

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., et al. (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the arabidopsis flowering time gene CONSTANS. *Plant Cell* 12, 2473–2484. doi: 10.1105/tpc.12.12.2473

Yano, K., Ookawa, T., Aya, K., Ochiai, Y., Hirasawa, T., Ebitani, T., et al. (2014). Isolation of a novel lodging resistance QTL gene involved in strigolactone signaling and its pyramiding with a QTL gene involved in another mechanism. *Mol. Plant* 6, ssu131. doi: 10.1093/mp/ssu131

Yu, J., Hu, S., Wang, J., Wong, G., Li, S., Liu, B., et al. (2002). A draft sequence of the rice genome (Oryza sativa l. ssp. indica). *Science* 296, 79–92.

Yu, J., Miao, J., Zhang, Z., Xiong, H., Zhu, X., Sun, X., et al. (2018). Alternative splicing of OsLG3b controls grain length and yield in japonica rice. *Plant Biotechnol. J.* 16, 1667–1678.

Yu, J., Xiong, H., Zhu, X., Zhang, H., Li, H., Miao, J., et al. (2017). OsLG3 contributing to rice grain length and yield was mined by ho-LAMap. *BMC Biol.* 15, 28. doi: 10.1186/s12915-017-0365-7

Zhang, J., Chen, L., Sun, S., Kudrna, D., Copetti, D., Li, W., et al. (2016a). Building two indica rice reference genomes with PacBio long-read and illumina paired-end sequencing data. *Sci. Data*, 3, 160076.

Zhang, J., Chen, L., Xing, F., Kudrna, D., Yao, W., Copetti, D., et al. (2016b). Extensive sequence divergence between the reference genomes of two elite indica rice varieties zhenshan 97 and minghui 63. *Proc. Natl. Acad. Sci. U.S.A.* 113, E5163–E5171.

Zhang, Z., Li, J., Pan, Y., Li, J., Zhou, L., Shi, H., et al. (2017). Natural variation in CTB4a enhances rice adaptation to cold habitats. *Nat. Commun.* 8, 14788.

Zhang, H., Wang, Y., Deng, C., Zhao, S., Zhang, P., Feng, J., et al. (2022b). High-quality genome assembly of huazhan and tianfeng, the parents of an elite rice hybrid tian-you-hua-zhan. *Sci. China Life Sci.* 65, 398–411.

Zhang, F., Xue, H., Dong, X., Li, M., Zheng, X., Li, Z., et al. (2022a). Long-read sequencing of 111 rice genomes reveals significantly larger pan-genomes. *Genome Res.* 32, 853–863.

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., et al. (2018b). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50, 278–284.

Zhao, D., Li, Q., Zhang, C., Zhang, C., Yang, Q., Pan, L., et al. (2018a). GS9 acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat. Commun.* 9, 1240.

Zhao, J., Wang, S., Qin, J., Sun, C., and Liu, F. (2020). The lipid transfer protein OsLTPL159 is involved in cold tolerance at the early seedling stage in rice. *Plant Biotechnol. J.* 18, 756–769.

Zhou, Z., Yu, Z., Huang, X., Liu, J., Guo, Y., Chen, L., et al. (2022). GenomeSyn: a bioinformatics tool for visualizing genome synteny and structural variations. *J. Genet. Genomics* 49, 1174–1176. doi: 10.1016/j.jgg.2022.03.013

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 March 2023
DOI 10.3389/fpls.2023.1133115

OPEN ACCESS

EDITED BY
Shouvik Das,
Regional Centre for Biotechnology (RCB), India

REVIEWED BY
Jitendra Kumar Thakur,
International Centre for Genetic Engineering and Biotechnology, India
Omar Borsani,
Universidad de la República, Uruguay
Venkata Ravi Prakash Reddy,
Acharya N. G. Ranga Agricultural University, India

*CORRESPONDENCE
Manu Agarwal
agarwalm71@gmail.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 28 December 2022
ACCEPTED 17 February 2023
PUBLISHED 08 March 2023

CITATION
Kumari A, Sharma D, Sharma P, Sahil, Wang C, Verma V, Patil A, Imran M, Singh MP, Kumar K, Paritosh K, Caragea D, Kapoor S, Chandel G, Grover A, Jagadish SVK, Katiyar-Agarwal S and Agarwal M (2023) Meta-QTL and haplo-pheno analysis reveal superior haplotype combinations associated with low grain chalkiness under high temperature in rice. *Front. Plant Sci.* 14:1133115. doi: 10.3389/fpls.2023.1133115

Meta-QTL and haplo-pheno analysis reveal superior haplotype combinations associated with low grain chalkiness under high temperature in rice

Anita Kumari[1], Divya Sharma[1], Priya Sharma[1], Sahil[1], Chaoxin Wang[2], Vibha Verma[3], Arun Patil[4], Md Imran[1], Madan Pal Singh[5], Kuldeep Kumar[6], Kumar Paritosh[7], Doina Caragea[2], Sanjay Kapoor[3], Girish Chandel[4], Anil Grover[3], S. V. Krishna Jagadish[8], Surekha Katiyar-Agarwal[3] and Manu Agarwal[1]*

[1]Department of Botany, University of Delhi, Delhi, India, [2]Department of Computer Science, Kansas State University, Manhattan, KS, United States, [3]Department of Plant Molecular Biology, University of Delhi, New Delhi, India, [4]Department of Plant Molecular Biology and Biotechnology, Indira Gandhi Krishi Vishwavidyalaya, Chattisgarh, India, [5]Division of Plant Physiology, Indian Council of Agricultural Research (ICAR), New Delhi, India, [6]National Institute for Plant Biotechnology, Indian Council of Agricultural Research (ICAR), New Delhi, India, [7]Centre for Genetic Manipulation of Crop Plants, New Delhi, India, [8]Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, United States

Chalk, an undesirable grain quality trait in rice, is primarily formed due to high temperatures during the grain-filling process. Owing to the disordered starch granule structure, air spaces and low amylose content, chalky grains are easily breakable during milling thereby lowering head rice recovery and its market price. Availability of multiple QTLs associated with grain chalkiness and associated attributes, provided us an opportunity to perform a meta-analysis and identify candidate genes and their alleles contributing to enhanced grain quality. From the 403 previously reported QTLs, 64 Meta-QTLs encompassing 5262 non-redundant genes were identified. MQTL analysis reduced the genetic and physical intervals and nearly 73% meta-QTLs were narrower than 5cM and 2Mb, revealing the hotspot genomic regions. By investigating expression profiles of 5262 genes in previously published datasets, 49 candidate genes were shortlisted on the basis of their differential regulation in at least two of the datasets. We identified non-synonymous allelic variations and haplotypes in 39 candidate genes across the 3K rice genome panel. Further, we phenotyped a subset panel of 60 rice accessions by exposing them to high temperature stress under natural field conditions over two Rabi cropping seasons. Haplo-pheno analysis uncovered haplotype combinations of two starch synthesis genes, *GBSSI* and *SSIIa*, significantly contributing towards the formation of grain chalk in rice.

We, therefore, report not only markers and pre-breeding material, but also propose superior haplotype combinations which can be introduced using either marker-assisted breeding or CRISPR-Cas based prime editing to generate elite rice varieties with low grain chalkiness and high HRY traits.

KEYWORDS

starch metabolism, grain chalkiness, meta-QTL analysis, haplotype, haplo-pheno analysis, granule bound starch synthase I, starch synthase IIa

1 Highlights

A combination of natural variations in starch metabolism genes SSIIa and GBSSI is associated with low grain chalkiness and high head yield under high temperature in rice.

2 Introduction

Rice (*Oryza sativa* L.), a primary staple food, is life for about half of the world's population. Besides providing food security, rice farming generates secure livelihood for millions across the world (Le, 2016). Asia is a major producer as well as consumer of rice accounting for over 90% of world's production and consumption (www.fao.org; Hossain et al., 2006; Muthayya et al., 2014). Asia (~4.67 billion) followed by Africa (~1.37 billion) are the two most populous continents and even though the population growth rate in Africa is nearly 3 times (2.45%) more than Asia (0.83%), the absolute population increase in Asia far exceeds that of Africa. Notwithstanding that surplus rice is being produced and consumption has been declining in Asia, it is predicted that the exponential increase in population will soon outstrip the incremental gains in rice production. The predicted gap in demand and supply is likely to be further widened by the climate change-accompanied abiotic factors that adversely affect the yield and quality of rice grains. Predictive modeling by the Intergovernmental Panel on Climate Change [IPCC, Sixth Assessment Report, 2021 (www.ipcc.ch)] estimates that even if all man-made activities are ceased, global temperature will continue to rise by 1.5°C to 2°C in the next two decades, which would be catastrophic to agriculture globally, including rice cultivation. Although rice is primarily cultivated in humid tropical and subtropical climates, minor increase in growth temperatures, especially high night-time temperatures, results in reduced spikelet fertility and sub-optimum grain characteristics. This includes, increased chalky grain ratio which causes significant reduction in grain yield and inferior quality attributes (Lyman et al., 2013; Wu et al., 2016).

Rice grains, in their journey from fields to the consumer, are scrutinized on several interrelated grain quality characteristics such as appearance, cooking, palatability and nutritional value. While the intermediaries, such as millers, are primarily interested in higher recovery of intact grain (also known as head rice yield or HRY), the consumer acceptance depends on the flavor, appearance and cooking quality (He et al., 1999). Grain chalk, an opaque and disordered area in the endosperm due to increased air spaces between the malformed starch granules (Lisle et al., 2000) is a highly unacceptable trait that accounts for many inferior grain characteristics. Based on the position of chalk deposition in rice grain, chalkiness is categorized as white-back, white-base, white-belly, white-core and milky-white (Tashiro and Wardlaw, 1991). The degree of chalkiness influences the recovery of head rice and since rice is preferably consumed as an intact whole kernel head rice fetches 40-50% higher price than the broken grains (Cooper et al., 2008; Chen et al., 2011; Zhao and Fitzgerald, 2013; Sreenivasulu et al., 2015). Consequently, besides boosting rice production, improvement of rice grain milling quality, with respect to high HRY and low grain chalkiness, are the key challenges in rice improvement programs worldwide (Nelson et al., 2011; Zhao and Fitzgerald, 2013).

Rice endosperm starch is composed of two main components: amylose and branched chain amylopectin. While granule-bound starch synthase I (*GBSSI*/Waxy) elongates amylose, soluble starch synthases (SSs), starch branching enzymes (BEs) and starch debranching enzymes (DEBs) are involved in synthesis and branching of amylopectin (Smith et al., 1997). Perturbations in the starch metabolism pathway leads to an aberrant amylose-amylopectin composition and altered starch structure, culminating in formation of chalk (Butardo et al., 2011). Both the expression and activity of starch metabolism genes and proteins, respectively, are modulated by high temperatures for e.g., starch synthesis genes are downregulated, while genes for starch degrading enzymes are upregulated by heat stress (Yamakawa et al., 2007; Liu et al., 2010b; Yamakawa and Hakata, 2010; Lin et al., 2017). Functional analysis studies have identified several genes like pyruvate orthophosphate dikinase (*PPDK*), starch synthase IIIa (*SSIIIa*), UDP-glucose pyrophosphorylase (*UGPase 1*) and cell-wall invertase, associated with chalk formation in rice (Kang et al., 2005; Fujita et al., 2007; Wang et al., 2008; Woo et al., 2008). However, in view of the spatial and temporal complexity of starch synthesis and deposition in developing rice grains, chalkiness appears to be a complex quantitative trait with polygenic inheritance.

10.3389/fpls.2023.1133115

TABLE 1 Summary of the grain chalkiness QTL mapping studies in rice used to perform the meta-analysis.

Mapping Population			Markers Used		QTL identified		Reference
Parents	Type	Size	Type	Number	Trait*	Number	
Sasanishiki x Habataki	ILs	37	SSR	142	PGWC	6	Bian et al., 2013
Pusa1266 X Jaya	RILs	310	SSR	116	PGWC, CS	2	Chandusingh et al., 2013
PYZXXP02430	RIL	192	SNP	2711	PGWC, DC	26	Chen et al., 2016
Chunjiang X Taichung	DH	116	SSR,STS	188, 19	CK, CS CD	10	Dai et al., 2016
XieqingzaoXMilyang	RIL	209	SSR	240	PGWC, DC,ET	40	Mei et al., 2013
KBNT lpa X Zhe733	RIL	187	SSR, SNP	174, 1	Chalk	10	Edwards et al., 2017
GSOR301227XGSOR 301190	RIL	276	SSR	132	Chalk	3	Eizenga et al., 2019
G46BXK1075	RIL	182	SSR	33	PGWC	3	Gao et al., 2016a
9311XPA64s	RIL	104	SNP	4	ACE, DC, PGWC	19	Gao et al., 2016b
AsominoriXIR-24	BC	9711	SSR	10	PGWC	1	Guo et al., 2011
AsominoriXIR-24	RIL, CSSL	7,166	RFLP	375	PGWC	10	Chen et al., 2011
CypressXPanda	RIL	137	AFLP, SSR	53,235	KCAb	3	Kepiro et al., 2008
Ali-Kazemi X Kadous	RIL	157	SSR	300	Chalk	2	Bazrkar-Khatibani et al., 2019
HanaechizanXNiigatawase	F2, F3	180	SSR	407	WBK	3	Kobayashi et al., 2007
HanaechizanXNiigatawase	RIL	178	SSR	175	WBK	4	Kobayashi et al., 2013
V20AXO.glaberrima	BC	308	RFLP, SSR	110, 20	WC, WCA	2	Li et al., 2004
KoshihikariasXGuchao 2	BC	71	SSR	144	PGWC, DPGWC	58	Wang et al., 2016
AsominoriXIR24	CSSL	66	RFLP	116	PGWC, ACE	15	Liu et al., 2010a
KoshihikarixXC602	RIL	261	SSR	236	PGWC	3	Liu et al., 2012
Xiushui79XC Bao	RIL	254	SSR	111	CGR, CD	15	Liu et al., 2013
N22X Nanjing35	F2, BILs	306	SSR	369	PGWC	10	Lu et al., 2013
TeqingXIR BB	BC	190	SSR	11	PGWC, DC	2	Mei et al., 2018
TsukushiromanXChikushi52	RIL	88	SSR, CAPS, dCAPS	70, 4, 2	MW	4	Miyahara et al., 2017
KoshihikariXYamadanishiki	RIL	284	SNP	703	WCE	6	Okada et al., 2017
ZS97XH94	DH	–	SSR	218	WBR, WBA, WCR, CA,CR	70	Peng et al., 2014a
V20BXYVB	BC	100	SNP	694	CR, CA	3	Peng et al., 2016
SamgangXNagdong	DH	120	SSR, STS	56, 116	PWC WCA, WBA, PWB, GCA, PGWC	16	Qin et al., 2009
KoshijiwaseXChiyonishiki	RIL	107	SSR	106	WK	4	Tabata et al., 2007
Zhenshan97XMinghui63	RIL	479	RFLP, SSR	166, 5	Chalk, WB, WC	9	Tan et al., 2000
HabatakiXSasanishiki	CSSLs	39	SSR	23	BW, MW, WBB	6	Terao and Hirose, 2018
TsukushiromanXChikushi52	RIL	88	SSR, SNP	71, 6	WB, BW	9	Wada et al., 2015
AsominoriXIR-24	RIL, CSSL	66	RFLP	116	PGWC, CA, DC	11	Wan et al., 2005
KoshihikariXNona Bokara	CSSLs	154	CAPs, SSR	102	PCRG, DC	2	Hao et al., 2009
O. sativaXO. rufipogon	IL	121	SSR	124	CR	1	Yuan et al., 2010

(Continued)

TABLE 1 Continued

Mapping Population			Markers Used		QTL identified		Reference
Parents	Type	Size	Type	Number	Trait*	Number	
Beilu130XJin23B	RIL	184	SSR	190	WCR, WBR	10	Yun et al., 2016
KoshihikariXKasalath	BC	182	RFLP	162	PGWC	3	Zheng et al., 2012
9311XPA64s	BC	3221	SSR	202	PGWC	1	Zhou et al., 2009
XBXZhonghui	BC4F4	–	SSR	25	PGWC	1	Zhu et al., 2018

*PGWC, Percentage Grain with Chalkiness; DPGWC, Degree of Percentage Grain with Chalkiness; DC, Degree of Endosperm Chalkiness; ET, Endosperm Transparency; CA, Chalkiness Area; WBe, White Belly; WBa, White Back; WBR, White Back Rate; WBA, White Back Area; WC, White core; WCR, White Core Rate; WCA, White Core Area; MW, Milky White; BW, Basal White; WBB, White Back and Basal

Several research groups have mapped quantitative trait loci (QTL) for grain chalkiness and its associated traits (Table 1). Few chalk-associated QTLs have been fine mapped such as Chalk5 that codes for a vacuolar H^+-translocating pyrophosphatase and whose enhanced expression contributes to increased chalkiness (Li Y., et al., 2014). Similarly, non-synonymous variations in OsAGPS1 gene within the qACE9 QTL affected chalk accumulation (Gao et al., 2016a). Takehara et al. (2018) dissected the Apq1 QTL to three genes and based on their temporal expression identified OsSUS3 as the candidate gene. However, the majority of QTLs from independent studies have not yet been fine mapped and have overlapping genomic regions owing to variable experimental designs and environment, genetic background of parents, population size, type and density of molecular markers and the statistical approach utilized for downstream analysis. Traditionally, the fine mapping of QTL to mine the causal gene/(s) involves mapping in the advanced generation (e.g., RILs) with an increased number of markers to obtain a condition where chances of recombination between the markers and traits are subtle. An alternative approach that can help in fine mapping, is meta-analysis of QTLs (MQTL) that are identified from multiple studies. MQTL analysis is fast emerging as an effective approach for reducing the confidence intervals (CI) of overlapping QTLs so that promising markers and genes associated with the concern trait could be rapidly mined.

We performed a Meta-QTL analysis of the QTLs reported for grain chalkiness in rice to discover the meta-regions associated with it. The identified MQTLs encompassed 5262 genes, which were further shortlisted to 49 based on their differential expression patterns in chalk-associated expression datasets. Further, we performed a haplo-pheno analysis for two of the most important starch synthesis genes and identified haplotype combinations that contribute to low grain chalk under high temperature stress. Our study highlights the relationship between high temperatures and rice grain chalk formation, by establishing the (1) relevance of a QTL meta-analysis in the identification of consensus and precise QTLs associated with grain chalkiness, (2) role and importance of differentially regulated genes from the identified meta-QTL genes, (3) identification of common cis-regulating elements in the candidate genes, (4) discovery of haplotypic variations in the candidate genes across the 3K rice genome panel, (5) haplo-pheno analysis of promising candidate genes in a subset of the 3K panel, and (6) association of haplo-pheno data with the starch granules morphology in the rice endosperm. Our work identifies superior allelic combinations which can be readily employed for generation of climate-resilient rice varieties.

3 Material and methods

3.1 Literature survey, QTL data collection, and input file preparation

Survey of literature published till 2021 on QTLs contributing towards rice grain chalkiness was performed resulting in the identification of 403 QTLs encompassing several chalk related traits. A total of 38 independent mapping studies (compiled as Table 1) involving 30 biparental rice populations were included in the MQTL analysis. QTL data constituting name, trait, chromosome, log of odds (LOD) score, phenotypic variance (R^2) value, position and confidence intervals (CI) were compiled in the text format and used as QTL input file. For studies having missing data on LOD score and R^2 we assumed them to be 2.5 and 10%, respectively, as previously suggested by Khahani et al. (2019). A genetic map file containing information about the genetic distances of markers on each linkage group was organized for each study. To incorporate QTLs derived from the studies based on SNP markers, the physical position of the flanking markers was ascertained on the rice genome and closest markers on the reference map were used.

3.2 QTL projection on consensus map and meta-analysis of QTLs

Prior to the meta-analysis, a consensus map was built by integrating the genetic map information of all 38 studies listed in Table 1 with a rice reference map (Temnykh et al., 2001) using BioMercator v4.2.3 (Arcade et al., 2004; Sosnowski et al., 2012). QTLs having variable CI were then projected onto the consensus map and the meta-analysis of only the projected QTLs was performed by merging all traits as rice grain chalk. We applied the two-step algorithm developed by Veyrieras (Veyrieras 2007), wherein first, the best model, having the lowest Akaike Information Criterion (AIC) value, corresponding to the estimated number of MQTLs was computed. In the second step, appropriate parameters

for the meta-analysis were set, including the number of MQTLs to be mapped, following which the MQTLs were generated.

3.3 Identification of candidate genes within the meta-QTL regions

For determining the genes underlying the identified MQTLs for rice grain chalkiness, the physical locations of confidence interval (CI) flanking markers were retrieved from the Gramene Marker Database (http://archive.gramene.org/qtl/) (Tello-Ruiz et al., 2022). In case the physical position of a particular marker was not available, the position of the adjacent genetic marker was utilized for identifying the genomic coordinates. The locus IDs of all genes spanning the marker intervals of respective MQTL regions were batch downloaded from the Rice Annotation Project Database (RAP) (Sakai et al., 2013). The obtained genes were then compared with the rice grain chalkiness expression datasets from five microarray-based studies (Yamakawa et al., 2007; Yamakawa and Hakata, 2010; Liu et al., 2010b; Ishimaru et al., 2019; Bakku et al., 2020) and one RNA-seq study (Lin et al., 2017). Genes common between at least two expression studies and the current analysis were considered as candidate genes regulating rice grain chalk. Gene ontology information of these chalk associated 49 candidate genes was collected from ShinyGO 0.76 database (Ge et al., 2020).

3.4 *In silico* promoter analysis and expression profile

The promoter sequence comprising 1.5 kb upstream from the translational start site (ATG) was downloaded from the Rice Genome Annotation Project Database (http://rice.uga.edu/) for the identified fundamental set of chalk associated candidate genes. All sequences were analyzed for motifs and regulatory elements on the PlantCare database (Lescot et al., 2002). To unravel the expression pattern of the candidate genes in rice endosperms we retrieved their expression data using Genevestigator (Hruz et al., 2008).

3.5 Haplotype analysis

An inbuilt tool of SNP seek database (Mansueto et al., 2016) was used to conduct haplotype analysis of the candidate genes in the 3k rice genome panel using Nipponbare as the reference genome. The number of haplotypes were discovered using the default parameters with calinski criterion (Calinski and Harabasz 1974) for determining the 'k' group. Only non-synonymous SNPs were considered while carrying out the analysis.

3.6 Plant materials and growth conditions

A subset of sixty rice genotypes from the 3K rice genome panel were chosen for grain chalk phenotyping based on the availability of seeds and uniformity in their phenology. Two independent experiments were conducted during Rabi 2020 and 2021 at IGKV, Raipur (21.2514°N, 81.6296°E). Weather data was recorded on a daily basis and plants were exposed to temperatures between 38-41°C during their seed filling stages. In each year four replicates of 15 plants each were grown in two independent fields and seeds were pooled before assessing grain chalk scores. Plants in both years were grown in open fields under fully flooded conditions to study the effect of high temperature stress coinciding with panicle initiation till physiological maturity.

3.7 Rice grain chalk phenotyping

Rice grains harvested at physiological maturity were dehusked using the automatic rice husker TR-250 (Kett, USA) and polished before the chalk imaging analysis. The polished grains were arranged on a transparent sheet kept on the scanner surface in three replicates each containing a non-redundant set of 50 randomly selected seeds. Seeds were scanned at a resolution of 800 dots per inch (dpi) using an Epson Perfection V850 Pro photo scanner. Images were saved in TIFF (.tif) file format for further image analysis. Further image processing and chalk score estimation was performed with Gradient-weighted Class Activation Mapping (Grad-CAM) tool and according to the methodology described by Wang et al. (2022).

3.8 Haplotypic and phenotypic correlation analysis

We phenotyped a subset of sixty rice genotypes from the 3K rice panel and categorized them as high chalk, intermediate chalk and low chalk to find out a correlation between the chalk score phenotype and the allelic variation in genotypes with respect to the identified candidate genes. Significance test of the means of haplotypic groups and Tukey's *post hoc* analyses was computed in R.

3.9 Scanning electron microscopic observation of starch granules

Scanning electron microscopy (SEM) of rice grains was performed to validate the chalk score values as determined by the image analysis. Mature and dry seeds were transversely fractured by putting pressure in the center of the seed by a surgical blade and were coated with 0.1 nm Chromium using Quorum Q150V plus Sputter coater (Peng et al., 2014b). The coated seeds were visualized under 5 keV electron beam energy using TESCAN Clara Scanning Electron Microscope (Tescan Orsay Holding, Czech Republic).

4 Results

4.1 Grain chalk QTLs in rice

Studies on discovery of QTLs associated with grain chalk in rice have used varied attributes for quantifying phenotypic variations. However, the maximum number of studies employed "percentage

of grains with chalkiness" (PGWC) as a measurement of trait. Grain chalk has also been accredited to other quantifiable attributes such as degree of percentage grains with chalkiness (DPGWC), degree of endosperm chalkiness (DC), endosperm transparency (ET), chalkiness area (CA), white belly (WBe), white back (WBa), white back rate (WBR), white back area (WBA), white core (WC), white core rate (WCR), white core area (WCA), milky white (MW), basal white (BW) and white back and basal (WBB) (Figure 1A). We, therefore, collated information on 403 QTLs for all these attributes from 38 different mapping studies (see Table 1 for references) to perform a MQTL analysis for the grain chalkiness trait in rice. These QTLs were discovered from 30 biparental crosses with a population size ranging from 37 to 3221 lines. Diverse molecular markers used in these studies included 4035 Simple Sequence Repeats (SSRs), 3416 Single Nucleotide Polymorphisms (SNPs), 1045 Restriction Fragment Length Polymorphisms (RFLPs), 532 Amplified Fragment Length Polymorphisms (AFLPs), 4 Cleaved Amplified Polymorphisms (CAPs), 2 Derived Cleaved Amplified Polymorphisms (dCAPs) and 135 Sequence Tagged Sites (STSs) (Table 1). The number of QTLs varied between 1 to 70 per population, with chromosome 6 having the highest number of QTLs (52) and chromosome 11 having the lowest number of QTLs (12). Confidence intervals (CI) of QTLs varied from 0.52 to 95.8 cM and the phenotypic variance ranged from 0.09% (Li et al., 2004) to 87.2% (Tan et al., 2000).

4.2 Distribution of grain chalk meta-QTLs on the rice genome

We next generated a consensus map from the 38 QTL maps and the rice reference map (Temnykh et al., 2001). The consensus map consisted of 11,141 markers with a genetic length of 2052.37cM. From the 403 initially retrieved QTLs, 330 (82%) were successfully projected on the consensus maps (Figure 1B). Meta-analysis of the 330 projected QTLs identified 64 MQTLs (Table 2, Figure 2; Supplementary Figure 1) on the basis of Veyrieras algorithm of BioMercator v4.2.3. The best MQTL model for each chromosome was selected on the basis of values of AIC, AIC correction (AICc), AIC 3 candidate (AIC3), Bayesian information criterion (BIC), and Average Weight of Evidence (AWE) for each chromosome. The 64 MQTLs were randomly scattered across all the twelve rice chromosomes and their number varied from 2 on Chromosome 8 to 8 on each Chromosomes 1 and 3 (Figure 2). Going strictly by definition, only MQTLs identified from at least 2 overlapping QTLs clustered on a consensus map are considered for the downstream analysis, while MQTLs derived from a single QTL are omitted (Yang et al., 2019; Kumar and Nadarajah, 2020). Therefore, MQTLs 1.6, 1.8, 4.2, 4.5, 11.1, 11.2, and 11.3 which were derived from a single QTL were not considered for further analysis (Table 2). Clustering of multiple QTLs identified from independent mapping populations that were phenotyped in different environments on the consensus map signifies the importance of the MQTLs identified. We, therefore, used a cut off of 10 QTLs clustering together to predict the most significant MQTLs. MQTL 6.1 derived from 14 overlapping QTLs (mean R^2 value of 23.7%), and MQTL 6.2 derived from 11 QTLs (mean R^2 value of 19.6%) indicated that short arm of Chromosome 6 plays an important role in formation of grain chalk. Similarly, MQTL 5.2 was derived due to clustering of 12 QTLs (mean R^2 value of 18%) suggesting that this genomic region of Chromosome 5 is also important for grain chalkiness in rice. The CI (95% confidence) of the MQTLs ranged from 0.03 to 15.49 cM with a median value of 2.81 cM. Similarly, the physical positions were also reduced and varied from 214 bp to 4.07 Mb with a median value of 0.65 Mb. Notably, the CI of 48 MQTLs (73%) was less than 5 cM, and 52 MQTLs (94%) exhibited a physical distance narrower than 2 Mb. Overall, the meta-analysis resulted in significant reductions in the CIs and physical distance of

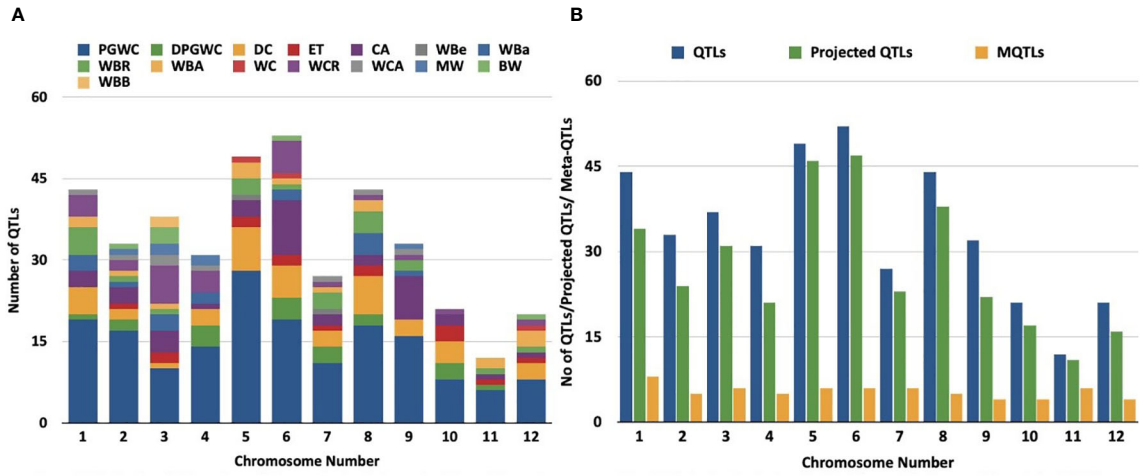


FIGURE 1
Distribution of QTLs and meta-QTLs associated with rice grain chalk on different chromosomes of rice. **(A)** Trait-wise distribution of initial QTLs used for the meta-QTL analysis. (PGWC, Percentage Grain with Chalkiness; DPGWC, Degree of Percentage Grain with Chalkiness; DC, Degree of Endosperm Chalkiness; ET, Endosperm Transparency; CA, Chalkiness Area; WBe, White Belly; WBa, White Back; WBR, ; WBA, White Back Area; WC, White core; WCR, White core rate; WCA, White core area; MW, Milky white; BW, Basal White; WBB, White Back and Basal) **(B)** The distribution of QTLs, projected QTLs and meta-QTLs on twelve rice chromosomes.

10.3389/fpls.2023.1133115

TABLE 2 Summary of the detected meta-QTLs for grain chalkiness in rice.

Sl No.	MQTL	Chr	QTL Cluster	Position (cM)	Left Marker Name	Position (bp)	Right Marker Name	Position (bp)	CI (95%) (cM)	Physical Interval (bp)	No. of genes
1	MQTL 1.1	1	2	13.59	RM10137	2740418	RM10167	3393293	2.62	652875	121
2	MQTL 1.2	1	4	35.85	RM10513	8072500	RM10578	9187636	4.43	1115136	133
3	MQTL 1.3	1	3	58.68	RM6073	13540292	RM10930	15546516	9.54	2006224	174
4	MQTL 1.4	1	2	91.07	RM11249	22796381	RM11323	24153000	5.69	1356619	162
5	MQTL 1.5	1	3	106.49	RM11408	25854706	RM11448	26814864	3.51	960158	115
6	MQTL 1.6	1	0	123.11	RM1152	30092742	RM11655	30957321	3.25	864579	156
7	MQTL 1.7	1	2	154.27	RM11969	38293480	RM12007	38874652	2.25	581172	93
8	MQTL 1.8	1	1	208.67	RG236	–	RZ538	–	3.63	–	–
9	MQTL 2.1	2	3	11.41	RM12456	2606829	RM12491	3091886	1.99	485057	72
10	MQTL 2.2	2	7	43.72	RM12983	10452238	RM13024	11463387	4.1	1011149	81
11	MQTL 2.3	2	3	64.63	RM13148	14821540	RM13231	17237626	9.68	2416086	174
12	MQTL 2.4	2	3	88.09	RM13456	21613848	RM3688	22395018	3.13	781170	82
13	MQTL 2.5	2	3	116.36	RM13844	29026616	RM13854	29103298	0.26	76682	12
14	MQTL3.1	3	5	12.25	RM14367	2271956	RM14433	3450953	4.83	1178997	193
15	MQTL3.2	3	5	20.19	RM14526	5056073	RM14530	5145876	0.2	89803	15
16	MQTL3.3	3	3	41.03	RM14757	9684987	RM14806	10590987	4.18	906000	153
17	MQTL3.4	3	9	62.7	RM15090	15508298	RM15158	16134623	2.8	626325	67
18	MQTL3.5	3	7	71.38	RM15209	17047128	RM15286	18767961	6.94	1720833	142
19	MQTL3.6	3	5	87.89	RM15326	20800545	RM15422	22682051	7.85	1881506	172
20	MQTL3.7	3	3	117.95	RM15826	29274076	RM15874	30052220	3.41	778144	94
21	MQTL3.8	3	2	129.42	RM16017	32519219	RM16024	32630844	0.39	111625	23
22	MQTL 4.1	4	3	9.22	RM16318	1384306	RM2811	2082855	2.82	698549	42
23	MQTL 4.2	4	1	31.91	RM7472	7090660	RM16556	8865320	8.57	1774660	98
24	MQTL 4.3	4	5	89.22	RM17017	21500378	RM17030	21627469	0.55	127091	12
25	MQTL 4.4	4	5	110.28	RM17279	26378417	RM17312	27317909	3.84	939492	127
26	MQTL 4.5	4	1	122.34	RM17401	29617932	RM17417	29860407	1.92	242475	33
27	MQTL 5.1	5	5	15.07	RM17950	3551113	RM17969	4139221	2.37	588108	51

(Continued)

10.3389/fpls.2023.1133115

TABLE 2 Continued

Sl No.	MQTL	Chr	QTL Cluster	Position (cM)	Left Marker Name	Position (bp)	Right Marker Name	Position (bp)	CI (95%) (cM)	Physical Interval (bp)	No. of genes
28	MQTL 5.2	5	12	36.18	RM18185	9063039	RM18189	9163390	0.55	100351	8
29	MQTL 5.3	5	7	71.33	RM6024	17752248	RM7386	18123723	1.55	371475	43
30	MQTL 5.4	5	2	92.78	RM18843	23077962	RM18872	23435326	1.5	357364	45
31	MQTL 5.5	5	2	116.04	RM19132	28241421	RM19224	29722687	7.43	1481266	245
32	MQTL 5.6	5	2	167.19	RG119	–	RG346	–	0.46	–	–
33	MQTL 6.1	6	14	13.31	RM19339	1750988	RM4923	2174616	1.45	423628	73
34	MQTL 6.2	6	11	21.17	RM5754	3762187	RM19477	4117576	1.44	355389	73
35	MQTL 6.3	6	7	32.88	RM19650	6555494	RM5531	7177097	2.46	621603	88
36	MQTL 6.4	6	4	45.06	RM527	9862309	RM527	9862523	0.52	214	1
37	MQTL 6.5	6	5	87.87	RM20215	20867990	RM20235	21162365	1.33	294375	26
38	MQTL 6.6	6	3	100.59	RM20365	–	RM20366	24196636	0.03	–	–
39	MQTL 7.1	7	5	22.52	RM21134	4946994	RM5672	6380197	5.58	1433203	145
40	MQTL 7.2	7	2	70.77	RM445	17462276	RM418	18132520	2.68	670244	48
41	MQTL 7.3	7	3	87.62	RM336	21871205	RM6403	22170728	1.19	299523	37
42	MQTL 7.4	7	4	103.36	RM21963	25338527	RM22046	26605775	4.88	1267248	169
43	MQTL 7.5	7	3	160.61	7042	–	RM1361	–	0.51	–	–
44	MQTL 7.6	7	3	187.12	7042	–	RM1361	–	0.4	–	–
45	MQTL 8.1	8	4	14.86	RM22432	3494017	RM22461	3979625	2.11	485608	58
46	MQTL 8.2	8	4	21.16	RM22543	5271763	RM22546	5345485	0.16	73722	13
47	MQTL 9.1	9	5	13.41	RM23753	3180304	RM2855	3839123	2.64	658819	36
48	MQTL 9.2	9	6	47.75	RM24116	11417353	RM24198	12782400	5.54	1365047	120
49	MQTL 9.3	9	4	78.3	RM24642	19929206	RM3787	20043308	0.35	114102	22
50	MQTL 9.4	9	4	79.66	RM24662	–	RM24663	20318284	0.03	–	–
51	MQTL 10.1	10	4	74.54	RM25607	18067635	RM1108	19161913	4.32	1094278	173

(Continued)

TABLE 2 Continued

Sl No.	MQTL	Chr	QTL Cluster	Position (cM)	Left Marker Name	Position (bp)	Right Marker Name	Position (bp)	CI (95%) (cM)	Physical Interval (bp)	No. of genes
52	MQTL 10.2	10	8	82.45	RM25779	20841416	RM5666	21407562	2.4	566146	81
53	MQTL 10.3	10	2	97.81	RZ421	22493127	RM25934	23075402	5.03	582275	84
54	MQTL 10.4	10	2	130.87	RM18A	–	RG561	–	3.72	–	–
55	MQTL 11.1	11	1	17.74	G1465	–	C6	–	5.76	–	–
56	MQTL 11.2	11	1	31.41	C1003A	–	C1172	–	4.92	–	–
57	MQTL 11.3	11	1	56.38	RM25983	657434	RM26068	2352559	6.82	1695125	256
58	MQTL 11.4	11	2	65.39	RM26127	3086024	RM26194	4372482	5.04	1286458	122
59	MQTL 11.5	11	5	85.78	RM26370	8248744	RM26434	9636542	5.33	1387798	78
60	MQTL 11.6	11	2	132.1	RM26824	18660295	RM27036	22740054	15.49	4079759	383
61	MQTL 12.1	12	3	18.79	RM27580	3327260	RM27657	4413960	4.02	1086700	124
62	MQTL 12.2	12	3	57.05	RM7102	13211325	RM28014	13507963	1.21	296638	12
63	MQTL 12.3	12	4	88.85	RM28331	20694607	RM7018	22163120	7.18	1468513	96
64	MQTL 12.4	12	5	110.31	RM28766	26583689	RM28771	26642770	0.36	59081	6

initial QTLs thereby minimizing the number of candidate genes for further analysis Figure 3A.

4.3 Candidate gene mining within meta-QTL regions

Since the majority of QTL regions from which MQTLs were derived have not yet been fine mapped, we considered all the genes within MQTLs as putative candidate genes. IDs of the genes for each MQTL were batch retrieved from the Rice Annotation Project Database (RAP), and a total of 5262 non-redundant genes were present within the 64 MQTL regions (Supplementary Table 1). Maximum number of genes were in the MQTL 11.6 (383) followed by MQTL 11.3 (256) and MQTL 5.5 (245), while the lowest number of genes were found within MQTLs 6.4 (1) and 12.4 (6) (Table 2). A common approach for further shortlisting the putative candidates is to ascertain the differentially expressed (DEGs) MQTL genes for the studied trait (Kong et al., 2020; Mirdar Mansuri et al., 2020; Raza et al., 2020; Khahani et al., 2021). Therefore, we overlapped all MQTL candidate genes with the 6 previously published expression datasets for grain chalkiness in rice (Yamakawa et al., 2007; Yamakawa and Hakata, 2010; Liu et al., 2010b; Lin et al., 2017; Ishimaru et al., 2019; Bakku et al., 2020). The investigation revealed that 314 MQTL genes (Figure 3B, Supplementary Table 2) were reported to be differentially expressed by Yamakawa and Hakata (2010). Similarly, 60, 56, 50, 48 and 26 genes were categorized as DEGs by Lin et al. (2017); Bakku et al. (2020), Ishimaru et al., (2019), Liu et al. (2010b) and Yamakawa et al. (2007), respectively (Figure 3B, Supplementary Tables 3-7). With an aim to discover a core set of chalk-associated genes we further shortlisted the DEGs on the basis of their presence in at least 2 expression datasets leading to the identification of 49 candidate genes (Supplementary Table 8; Figure 3C). Gene ontology (GO) analysis revealed that out of 49 candidates, 7 genes were involved in the carbohydrate metabolism. Besides these 9 genes belonged to the stress response pathway, 5 genes each belonged to the peptide metabolism pathway and seed storage proteins, 2 genes each were members of alpha amylase inhibitors, serine carboxypeptidases and carboxylic acid metabolic process. One member each was annotated as amino acid transporter and monosaccharide transporter, while 15 genes possessed non-significant annotations (Figure 3D).

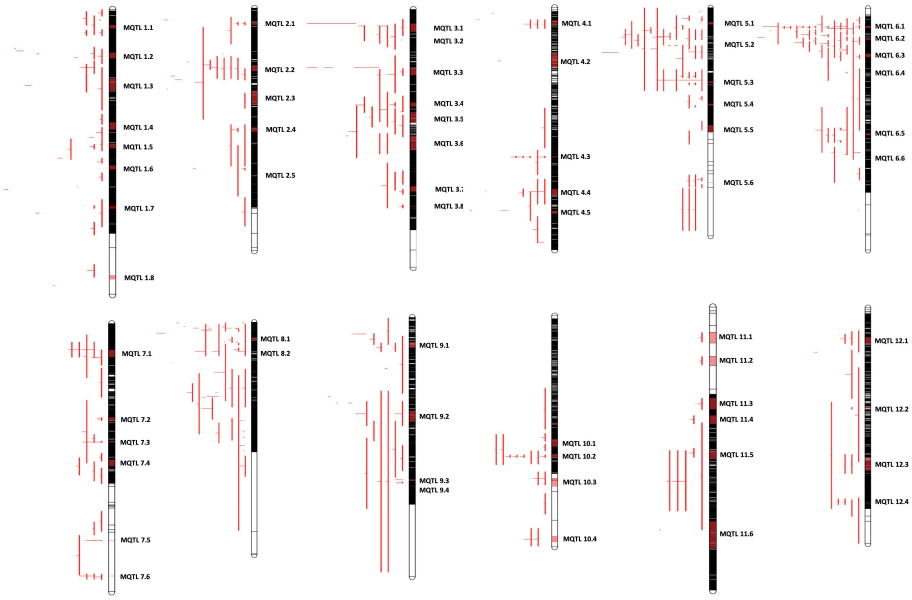


FIGURE 2
Distribution of meta-QTLs on rice chromosomes. Rice chromosomes are represented by the vertical bars. The names of consensus markers along with their position (cM) are present on the right of the chromosomes. QTLs mapped in various studies are highlighted in different colours on the left of the chromosomes. Coloured blocks on the chromosomes represent the identified meta-QTLs. Vertical bars on the left side of the chromosome represent the confidence interval (CI) of QTL and horizontal bars represent the phenotypic variance.

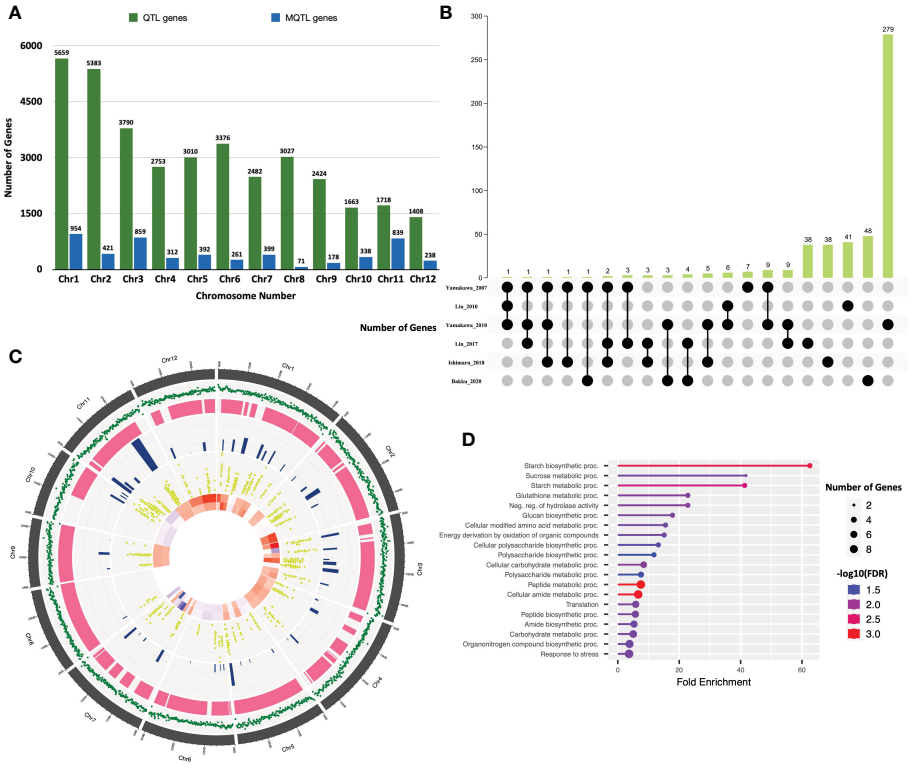


FIGURE 3
Candidate genes underlying the meta-QTL regions. **(A)** Chromosome-wise distribution of genes underlying QTLs and the identified meta-QTLs in rice. **(B)** Upset plot highlighting the number of differentially regulated genes identified in six rice grain chalk-related transcriptomic studies. **(C)** Distribution pattern of the gene density, QTLs, meta-QTLs, meta-QTL spanning genes and candidate genes identified on the rice genome. The outermost circle represents the chromosomes position on the rice genome in Mb. The second circle with green color outlines the gene density on the rice genome. The third and fourth inner circles display the number of initial chalk QTLs and identified meta-QTLs, respectively. The fifth circle represents the MQTL underlying genes., while the innermost circle represents heatmap of the identified candidate genes (red represents upregulation, while blue signifies downregulation). **(D)** Gene Ontology enrichment analysis of 49 candidate genes representing the biological process. Created using ShinyCircos Yu et al., 2018

4.4 Cis-regulatory elements in the promoter of candidate genes

Although synthesis of starch is initiated in leaves where it serves for growth and development, seed endosperm being the sink is at the center of starch metabolism in rice (MacNeill et al., 2017). Consequently, many genes are co-expressed in the endosperm for regulating the optimal quality and quantity of starch granules. We, therefore, investigated whether the 49 candidate genes had common cis-elements in their promoters (Figure 4A) that might regulate their spatial and temporal expression. The 1.5 kb putative promoter region of all these genes were enriched with important cis-regulatory elements (CREs) elements, such as stress-responsive element (STRE; AGGGG), activation sequence-1 like promoter element (AS-1; TGACG), purine (A or G) and pyrimidine (C or T) element (RY; TGAGTCA) and GCN4 (CATGCATG). The AGGGG motif termed as stress-responsive element (STRE), initially identified in yeast, is responsive to several stresses (Martinez-Pastor et al., 1996). The AS-1 cis-element is present in genes regulated by auxin, MeJa and salicylic acid (Miao and Lam, 1995; Abdullah-Zawawi et al., 2021; Kabir et al., 2021). This element has also been reported in the promoters of several Glutathione-S-transferase genes (Hu et al., 2011) and abiotic stress response genes (Garretón et al., 2002). All the 49 investigated genes had the AS-1 and STRE motifs, suggesting that they might play a direct or indirect role in starch metabolism during unfavorable conditions. The GCN4 motif is crucial for seed-specific expression (Wu et al., 1998), while the RY element coordinates with other CREs and also plays a key role in seed-specific gene regulation (Ezcurra et al., 1999). A total of 17 genes either had GCN4 or RY motifs in their promoters indicating that these genes are possibly co-expressed in rice seeds. Further, with the help of Genevestigator expression datasets we observed that 41 out of 49 genes were differentially expressed in the rice endosperms (Figure 4B), indicating their role in the formation of rice grain chalk.

4.5 Haplotype analysis of 49 candidate genes reveals genetic diversity across the 3K rice genome panel

The 3K rice genome (RG) project is a giga dataset of publicly available genome sequences which can be used to mine genotypic variations in a population and link them to the phenotypic variability

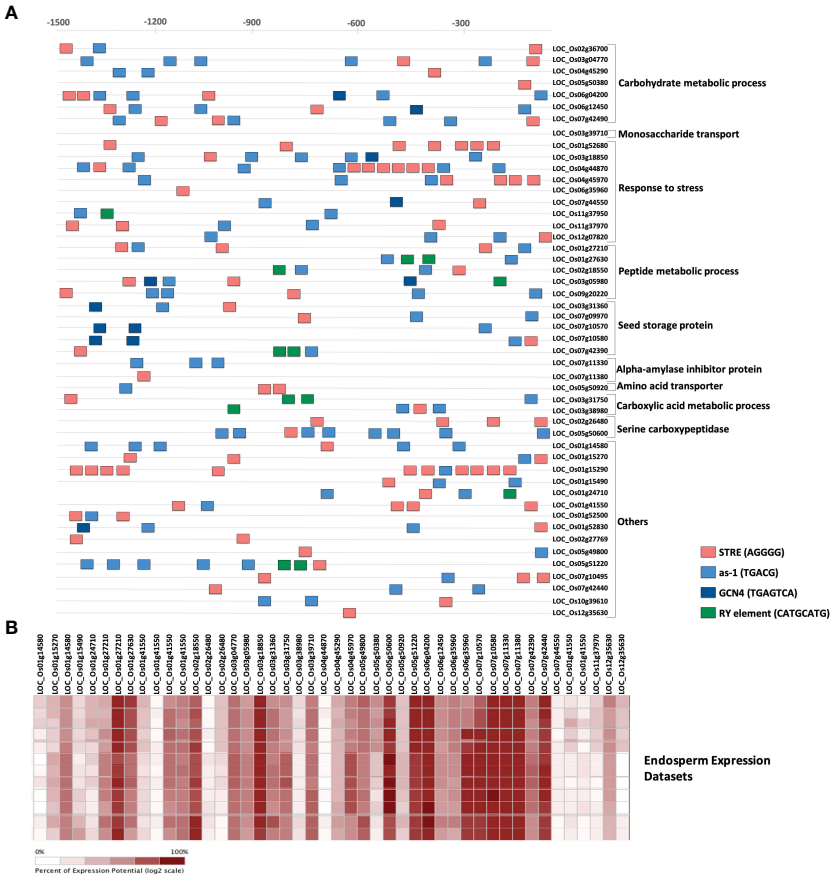


FIGURE 4
Spatio-temporal expression of candidate genes. **(A)** In silico identification of regulatory elements in the putative promoter region of the identified fundamental set and known chalk-responsive genes. Coloured boxes indicate as-1 element (blue), RY element (green), STRE (salmon) and GCN4 motif (dark blue) in 1500 bp upstream promoter regions. Scale is presented above the figure. **(B)** Heatmap highlighting the percent expression pattern of candidate genes in the endosperm, figure created using Genevestigator.

for exploring genetic diversity of different agronomic traits in rice (Li et al., 2014). We, therefore, performed a haplotype analysis of the 49 candidate genes to discover the non-synonymous variations in their coding sequences (CDS) across the 3K RG panel (Supplementary Table 9). Interestingly, 39 of the 49 analyzed candidate genes possessed non-synonymous SNPs ranging from 2 to 8, while the remaining 10 genes did not show any non-synonymous variations. Five of the seven carbohydrate metabolism-related genes contained non-synonymous SNPs in their coding sequences. We discovered 6 haplotypes and 1 allelic variation in sucrose transporter 5 (*SUT5*) and vacuolar invertase (*VIN2*), respectively, which were not reported previously and never before associated with the grain chalkiness and its related attributes. However, haplotypes in *GBSSI* and *SSIIa* have previously been implicated in starch metabolism and therefore we decided to explore the allelic variability in their CDS across the 3K panel and the subpanel of 60 accessions (Figure 5).

Waxy (Wx) gene, which encodes the enzyme *GBSSI*, converts ADP-glucose to amylose in the endosperm of cereals (Sano, 1984; Webb, 1991). Allelic variations of Wx gene in rice; Wx^{lv}, Wx^{a}, Wx^{in}, Wx^{b}, $Wx^{la/mw}$, Wx^{mq}, $Wx^{op/hp}$ and wx (Supplementary Table 10), linked to varied amylose content (AC values), are distinguished by six polymorphic sites namely Int1-1, Ex2-112, Ex4-53, Ex4-77, Ex6-62, and Ex10-115 (Figure 5B) (Cai et al., 1998; Isshiki et al., 1998; Mikami et al., 1999; Sato et al., 2002; Wanchana et al., 2003; Mikami et al., 2008; Mikami et al., 2008; Liu et al., 2009; Yang et al., 2013; Zhang et al., 2019). *GBSSI* has 1 splice site SNP and 5 non-synonymous SNPs that give rise to 8 haplotype combinations. We were able to detect 6 of the 8 haplotype combinations in the entire 3K panel. Two hundred and one accessions (7%) had missing entries in the 3K panel either due to lower sequencing depth or sequencing errors and were therefore excluded from the analysis. Out of the remaining 2826 accessions, haplotype GGAAT (H2) had the highest frequency and was present in 738 accessions, whereas other haplotypes GGAAC (H1), TGAAC (H4), GGACC (H5), GGGAC (H3) and TGACC (H6) were observed in 23.7%, 22.4%, 19.8%, 1.8% and 1% of accessions, respectively (Figure 6B). A 'G' to 'A' change at Ex4-53 (Exon 4 th, 53 rd nucleotide) results in low amylose content and give rise to milky endosperms known as "milky queen" phenotype in seeds (Sato et al., 2002). Interestingly, the entire 3K panel had 'G' at Ex4-53 (Figure 5C) indicating that the "milky queen" phenotype was selected during varietal development. Owing to the presence of G in all the accessions, we were unable to find Wx^{mq} haplotype in the panel. Further, four haplotypes (GGAAC, GGAAT, GGGAC and TGAAC) were represented in the subset panel consisting of 60 accessions.

Soluble starch synthase IIa gene (*SSIIa*) elongates glucan chains from ADP-glucose units to form amylopectin in conjunction with branching and debranching enzymes (Denyer et al., 1995; Denyer et al., 1996; Tetlow and Emes, 2014). We discovered that *SSIIa* had 7 haplotypes arising from a combination of 9 non-synonymous SNPs (H1-CAGCTCGGC, H2-CCGCGGGGC, H3-GCGCGGGGC, H4-CAGCGCAGC, H5-CAGCGCAGT, H6-CAGCGCAAC, H7-CCTAGCGGC), in the entire 3K RG panel (Figure 6A). While 625 accessions had missing data, in the remaining 2402 accessions H4 was most predominant (50% of accessions). Other haplotypes in the descending order of their frequency (Figure 6C) were H7 (0.16%), H1 (0.2%), H3 (1.5%), H6 (2.3%), H2 (4%) and H5 (20.8%). However, we

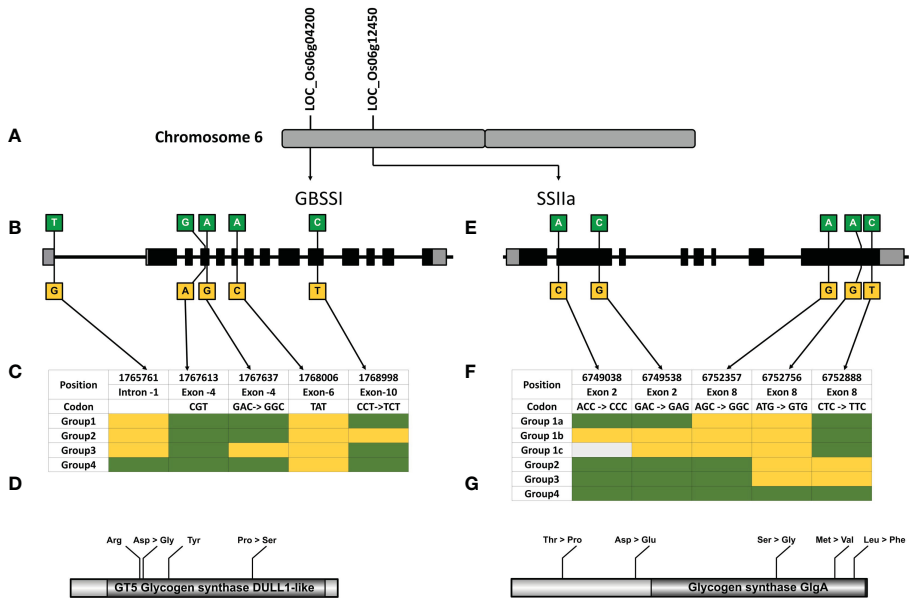


FIGURE 5
Haplotype analysis of Granule Bound Starch Synthase I (GBSSI) and Starch Synthase IIa(SSIIa). **(A)** The genes coding for GBSSI and SSIIa are located on the short arm of chromosome 6. **(B)** Gene diagram showing the non-synonymous variations of GBSSI, comprising 13 exons (grey boxes-UTR, Black boxes- CDS). **(C)** Haplotype table showing the SNP positions within GBSSI, four haplotypes are constructed based on the sequence variations in 25 rice genotypes (green indicates the reference allele Nipponbare, yellow indicates the SNPs and light grey shows missing data). **(D)** Depiction of GBSSI protein and amino acid changes in the GT-5 Glycogen Synthase DULL like domain predicted using NCBI CDD **(E)** Gene diagram showing the non-synonymous variations of SSIIagene comprised of 8 exons. **(F)** Haplotype table showing the SNP positions within SSIIa, four haplotypes, color codes same as **(B)**. **(G)** Depiction of SSIIa protein and amino acid changes in the Glycogen Synthase Glga domain predicted using NCBI CDD.

A

	GBSSI		SSIIa	
Haplotypes	GBSSI in 3k RG panel	GBSSI in subset panel	SSIIa in 3k RG panel	SSIIa in subset panel
H1	GGAAC	GGAAC	CAGCTCGGC	ACGGC
H2	GGAAT	GGAAT	CCGCGGGGC	CGGGC/_GGGC
H3	GGGAC	GGGAC	GCGCGGGGC/G_GCGGGGC	
H4	TGAAC	TGAAC	CAGCGCAGC	ACAGC
H5	GGACC	-	CAGCGCAGT	ACAGT
H6	TGACC	-	CAGCGCAAC	ACAAC
H7	-	-	CCTAGCGGC	-

FIGURE 6
Distribution of haplotypes of GBSSI and SSIIa in rice 3k and its subset panel. **(A)** Table depicting the number of haplotypes in the 3k panel *Vs.* the subset panel for both genes, **(B)** Pie chart showing the % distribution of GBSSI haplotype in the 3k RG panel, **(C)** Pie chart highlighting the % distribution of haplotypes in SSIIa across 3k RG panel.

could only detect 5 SNPs giving rise to 6 haplotypes in the subset panel of 60 accessions (Figurse 6 e, f). The missing 4 SNPs C, G/T, C/A and G/T resulted in 3 combinations "CGCT" (H1), "CGCG" (H2-H6) and "CTAG" (H7). While the combination of missing SNP "CGCT" was present in 21% of accessions, "CTAG" was present in only 2% of accessions. Out of the 6 haplotype combinations in the gene encoding for SSII-a enzyme, 4 haplotype combinations arising from 3 SNPs (H1-GGC, H2-AGC, H3-AGT and H4-AAC), are associated with chain length of amylopectin, activity of starch synthase, gelatinization temperature and alkali spread score (Umemoto et al., 2004; Umemoto and Aoki, 2005; Waters et al., 2006). All the 3 SNPs resulting in 4 haplotypes are present in the catalytic domain of the enzyme (Figure 5G) and therefore are important for enzyme activity, which was more for the haplotypes H1 and H2 and less for H3 and H4 (Umemoto et al., 2004; Umemoto and Aoki, 2005).

4.6 Identification of superior haplotypes for granule bound starch synthase I and starch synthase IIa by haplo-pheno analysis

Since haplotypes in *GBSSI* and *SSIIa* are important for their activity and likely to affect starch quality, we investigated their haplotype behavior with respect to the grain chalkiness in the subset panel of 60 accessions. These accessions were subjected to high temperature stress under natural field conditions during the grain filling stages over two Rabi seasons in Raipur, India. The amount of chalk (expressed as "chalk score") was determined by scanning the polished rice grains and processing the images through a Gradient-weighted Class Activation Mapping (Grad-CAM) tool (Wang et al., 2022). The chalk score ranged from 0.046 to 0.075 and the accessions were distributed in low chalk (CS ≤ 0.05), intermediate chalk (0.05 > CS ≤ 0.065) and high chalk (CS > 0.065) categories (Supplementary Table 11, Figure 7A). Seven accessions were in the low chalk category, thirty-three in the intermediate and twenty were in the high chalk category. Inferential statistics ascertained that the means of all three chalk categories were significantly different (Figure 7B). Analysis of variance (ANOVA) determined that the variability of chalk score with respect to all the 4 haplotypes of SSIIa and GBSSI were significantly different and had p-values of ≤ 0.05 and ≤ 0.1, respectively. However, after Tukey's *post hoc* analyses, differences in the mean CS values for only *GBSSI* haplotypes GGAAC and GGGAC were significant (Figure 7C). Out of the 4 possible haplotypes in the catalytic domain of *SSIIa*, significant differences in the mean CS values were seen between GGC and AGT, GGC and AGC and GGC and AAC (Figure 7D). To find whether and which combination of haplotypes for *GBSSI* and *SSIIa*

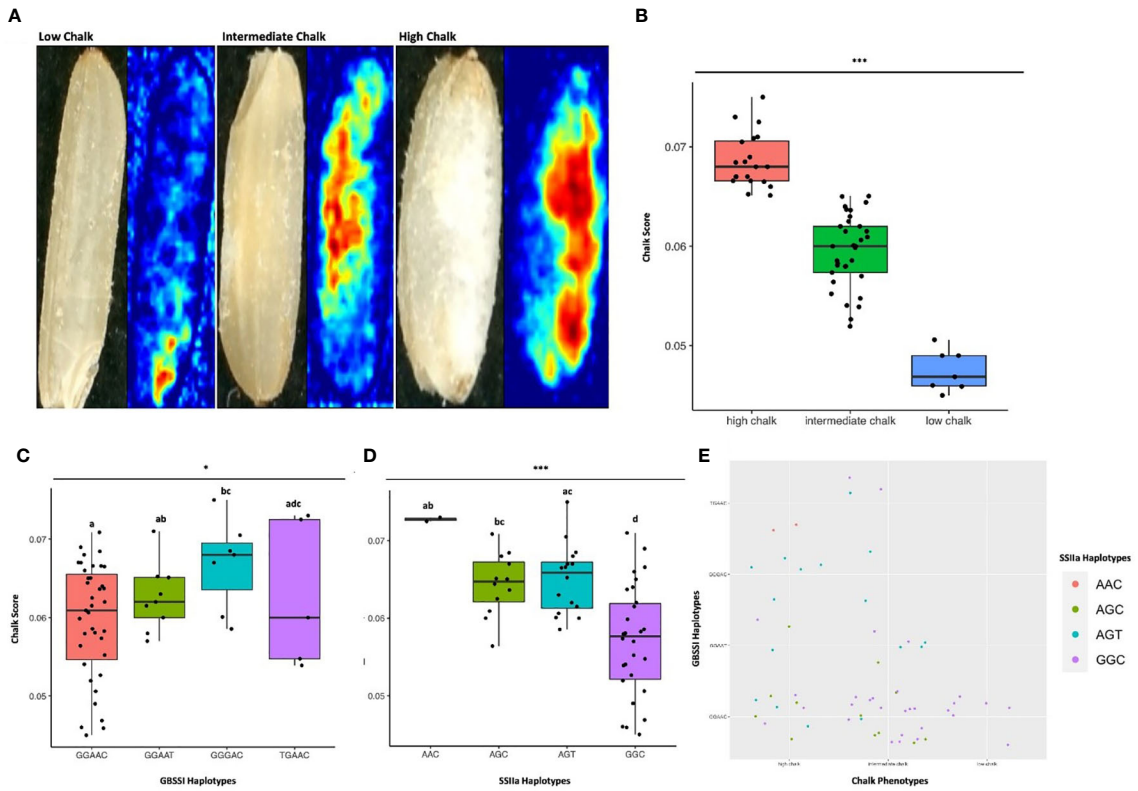


FIGURE 7
Haplo-pheno analysis of Granule Bound Starch Synthase I and Starch Synthase II a. **(A)** Rice seed scans and their corresponding heat maps highlighting the chalky area (red) in less chalk, intermediate chalk and high chalk seed types. **(B)** Box plot depicting the chalky score of 60 rice genotypes (grown in two crop seasons) categorized as high, intermediate and low chalk. **(C)** Box plot of the chalk score data categorized according to the GBSS I haplotypes. **(D)** Box plot of the chalk score data categorized according to the SSIIa haplotypes. **(E)** Comparison of the haplotypic combination of GBSS I and SSIIa. One-way ANOVA was conducted for determining the statistical significance of haplotype means. Chalk score data designated with the same alphabet are not significantly different at $p \leq 0.1$(.) or $p \leq 0.05$ (*) or $p \leq 0.001$ (***) as per Tukey *Post hoc* test for Honest Significance Difference (HSD) analysis.

associate with low grain chalkiness under high temperature stress, we performed a correlation analysis of different haplotype combinations with the chalkiness phenotypes. Interestingly, all the seven accessions with lowest mean CS, had the haplotypic combination of GGAAC (for *GBSSI*) and GGC (for *SSIIa*). Fifteen accessions having an intermediate CS, also had a similar haplotype combination. The other 18 accessions with intermediate CS had the following haplotype combinations for *GBSSI*/*SSIIa*; GGAAC/AGC, GGAAT/GGC, GGAAT/AGC, GGAAT/AGT, GGGAC/AGT, GGAAC/GGC, TGAAC/GGC and TGAAC/AGT. Five high CS accessions had the haplotype combinations (*GBSSI*/*SSIIa*) of GGGAC/AGT, while four had GGAAC/AGC, three each had GGAAC/AGT and GGAAC/GGC, two had TGAAC/AAC and 1 each had GGAAT/AGC, GGAAT/AGT and GGAAT/GGC.

4.7 Scanning electron microscopy validated the correlation between chalk score and haplotypic combinations

Six accessions, two from each category of low, intermediate and high chalk score were selected in order to find correlation between haplotypes and packaging of starch grains in chalky and non-chalky seeds. The seeds of the accessions CX160 and IRIS 313-11462 having low chalk score with GGAAC/GGC haplotypic combination showed dense packaging of starch granules under scanning electron microscope. The amyloplast and the starch grains in them were very tightly packed and no space was visible between them. This probably contributed to the translucent appearance of the grain (Figures 8A1, B1). Whereas, the accessions IRIS 313-10417 and IRIS 313-11245, with GGAAT/AGC, GGAAT/AGT haplotypic combination grouped under the intermediate chalky category, exhibited slightly bulged starch granules with visible spaces between individual amyloplasts (shown with red arrow, Figures 8C1, D1). In contrast, the high chalk genotypes Nipponbare and IRIS 313-11198 with TGAAC/AAC and TGAAC/AAC haplotypic combinations showed visibly round starch granules (indicative of loose packaging) instead of polyhedral and with noticeable air spaces between them. Presence of air spaces and loose packaging scatters light leading to opaque or chalky appearance as shown in Figures 8E1, F1. A magnified view of the starch granules is shown in the lower panel (Figures 8A2-F2). The SEM data validates a clear association between haplotypic combinations and phenotypic data relating to degree of seed chalkiness.

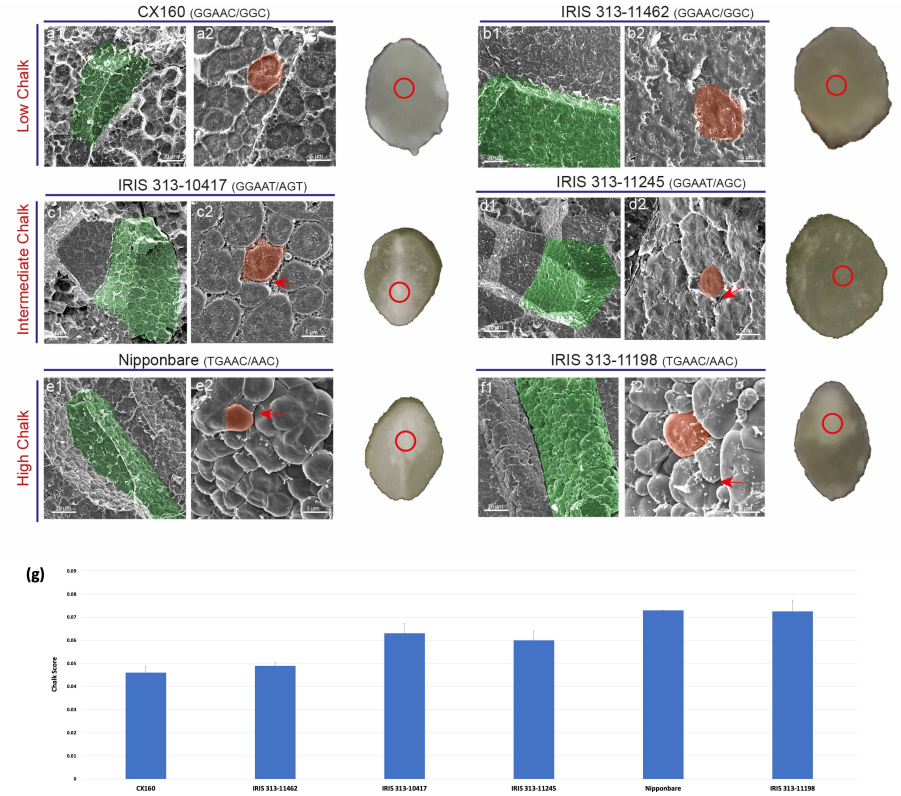


FIGURE 8
Scanning electron micrographs of transversely fractured mature rice seeds. (A1-B2) Endosperm cell (green) of low chalk genotypes is tightly packed with starch grains (orange). (C1-D1) Micrographs of intermediate chalk genotypes with slightly bulged starch granules with visible spaces between individual amyloplasts. (E1-F2) Images of transversely cut chalky endosperm region filled with spherical starch grains instead of polygonal as in a1 and b1. Air spaces (shown by red arrows) can be seen between the grains in e1 and f1. The red circle on a transverse cut seed shows the region selected for analysis.

5 Discussion

Mapping of quantitative trait linked genomic regions in segregating populations from diverse genetic backgrounds has led to the identification of many grain chalk associated QTLs. Majority of these studies employed markers providing low genetic resolution and therefore the identified QTLs span a large CI. Furthermore, many QTLs from independent studies map to similar genomic regions. It is, therefore, important that common and consistent regions of overlapping QTLs are identified for their effective utilization in trait improvement programs. A meta-QTL analysis aptly serves to achieve this goal (Goffinet and Gerber, 2000) and has been successfully applied to several crop plants, including rice. In this study, we performed a meta-analysis of rice grain chalk QTLs to precisely identify the most significant QTL regions and their underlying candidate genes. We were able to restrict the most important genetic information of 403 QTLs within 64 meta-QTLs. To comprehensively identify chalk-associated candidate genes, expression of the genes underlying these meta-QTLs was assessed in the published grain chalk-associated expression datasets, leading to the identification of a core set of 49 genes. We then performed a promoter analysis to identify the common cis-regulating elements, followed by a haplotype analysis of the 49 candidate genes across the 3K RG panel which revealed natural variations in these genes. Owing to the direct connection of genes involved in sugar transport and metabolism, we correlated haplotypes in 6 sugar metabolism genes with the grain chalkiness in a subset of 60 accessions. Our analysis revealed haplotype combinations in *GBSSI* and *SSIIa* that are associated with presence of low chalk.

5.1 Meta-analysis complemented with gene expression underpins key genes involved in grain starch metabolism

Conversion to starch from UDP/ADP-Glucose is a multistep process that begins with the transport of sucrose, its conversion to glucose followed by synthesis of amylose and amylopectin (Figure 9A). Genes and proteins involved in each of these steps are themselves regulated at transcriptional, post-transcriptional and post-translational levels. Majority of the QTL mapping studies for grain chalk have identified genes directly involved in starch synthesis, while other studies have led to the discovery of genes which code for regulators. In the MQTL analysis we identified 7

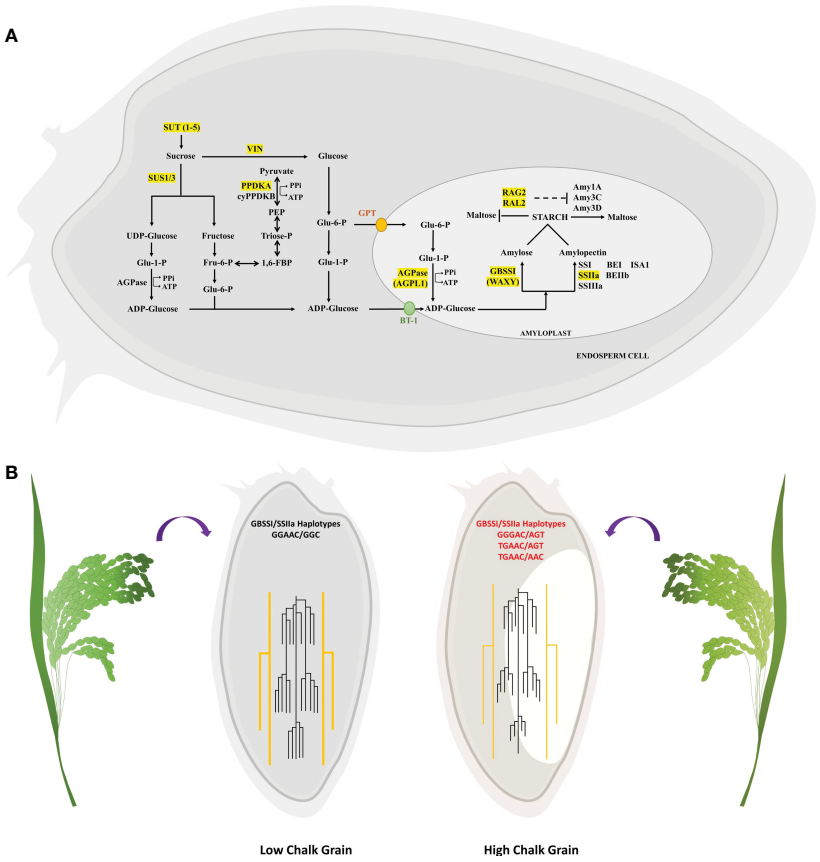


FIGURE 9
Depiction of identified candidate genes in the starch metabolism pathway and their haplotypic combinations affecting grain chalkiness. **(A)** Starch metabolism pathway in seeds begins with the sucrose transporters (SUTs) that translocate sucrose into the seeds, followed by its breakdown into glucose and fructose by the enzymatic action of invertases (VIN) and sucrose synthases (SUS). Finally glucose-1-phosphate (Glu-6-P) is converted to ADP-glucose by ADP-glucose pyrophosphorylase(AGPase), which is converted to starch by the combined action of Granule Bound Starch Synthase I (GBSSI) which forms amylose and Starch Synthase IIa (SSIIa) which forms linear glucan chains leading to formation of amylopectin. During HT amylases (Amy) breakdown starch to malto-oligosaccharides. RAG2/RAL2 from alpha-amylase inhibitor family are thought to suppress the amylase activity. **(B)** Grains with low chalk have high amylose and longer chain length while high chalk grains have low amylose and short chain amylopectin. Low grain chalk seeds have superior haplotypic combination of GBSSI and SSIIa when compared with high chalk seeds.

genes that code for proteins belonging to the starch metabolism pathway, owing to their discovery in overlapping QTLs derived from different mapping studies. This further proves that regulation of sugar metabolism governs the core process of chalk formation.

Sucrose transporters (*SUTs*) mediate the proton symport coupled transport of sucrose across the plasma membrane through the phloem to developing grains (Aoki et al., 2003). The *OsSUT* gene family comprises five members *OsSUT* 1, 2, 3, 4 and 5 that demonstrate overlapping expression patterns during grain filling and therefore might exhibit partial functional redundancy (Aoki et al., 2003). *OsSUT*5, one of the candidate sugar metabolism gene identified in the current study lies in MQTL 2.4, which is contributed by QTLs qPGWC2 (Chen et al., 2011), qAEC2 and qDEC2 (Wan et al., 2005). Structure-based modeling experiments and electrophysiology assays in Xenopus oocytes revealed that five of the six charged amino acids (D177, R188, D331, R335 and E336) are important for sucrose transport by *OsSUT1* (Sun et al., 2012). All the five functionally important amino acids were conserved in *OsSUT5* (unpublished data) across the 3K panel. This signifies the importance of functional redundancy in sucrose transporters for mobilization of sugars to sink to ensure normal grain filling. Further, haplotype analysis revealed 6 non-synonymous haplotypes of *SUT5* gene in the entire 3K panel (Supplementary Table 9). However, none of the haplotypes were in the wobble base of the codons for the five functionally important amino acids. None of the haplotypes grouped with the chalk score data in the 60 accessions which implies that their relevance will need further experimentation. Nonetheless, the formation of grain chalk is regulated by a aleurone-specific TF, *OsNF-YB1*, which binds to the promoters of *OsSUT* genes and transcriptionally activates their expression (Bai et al., 2016). In this respect, it would therefore be important to assess the polymorphism at "CCAAT" boxes in the promoters of *OsSUT* genes in future. Sucrose synthase (SUS) enzyme is encoded by 7 conserved *SUS* genes with compensatory

functions (Hirose et al., 2008; Cho et al., 2011) and catalyzes the conversion of sucrose and uridine diphosphate (UDP) into fructose and UDP-glucose (Huang et al., 2016). *OsSUS3*, a candidate gene in this study present in MQTL 7.4 [contributing QTLs qWBA, qWCA, qWBR and qET(Peng et al., 2014a; Yun et al., 2016; Mei et al., 2013)] was predominantly expressed in the endosperm and the aleurone layers at the onset of starch synthesis (Huang et al., 1996). Our data revealed that the *SUS3* gene had 1 allelic variation in the coding region of the 3K panel (Supplementary Table 9), which is in contrast to previous reports where presence of 2 (6 SNPs) and 11 haplotypes in the coding sequence have been reported (Lestari et al., 2011; Takehara et al., 2018). Additionally, haplotypes were also identified in the non-coding region of the *SUS3* gene which included the promoter (Lestari et al., 2011; Takehara et al., 2018). It was, however, surprising that none of the previously identified haplotypes in the *SUS3* CDS were seen in the 3K panel. Only 3 of the varieties (IR64, Milyang 23 and Lemont) harboring the reported haplotypes were part of the 3K sequenced panel. We believe that these haplotypes were not detected in the variant analysis of the 3K panel because the SNP dataset was a filtered dataset [cutoff values of minor allele frequency (MAF) > 1% and missing < 20%; Mansueto et al., 2016]. Nonetheless the *SUS3* alleles in the promoter and coding region from Habataki, an indica variety were important for low grain chalk under high temperature stress as introgression of Habataki SUS3 alleles in Koshihikari (Apq1-NIL), a japonica variety, led to increased percentage of perfect grains (Takehara et al., 2018).

Sucrose flux in the plant metabolic pathways is regulated by isoforms of invertases (Sturm, 1999) which catalyze the irreversible cleavage of sucrose into glucose and fructose. Rice has eight neutral invertases (*NINs*, cytosolic), nine cell wall invertases (*CWINs*) and two vacuolar invertases (*VINs*). Our study identified *OsINV2* as one of the candidate genes in MQTL 4.4 [contributed by QTLs qDPGWC, qPGWC, qWCR4.1, qWCR4.2, qch (Peng et al., 2014a; Wang et al., 2016; Yun et al., 2016; Edwards et al., 2017)]. *OsVIN2* and its homologue *OsVIN3* regulate the sink size by controlling hexose to sucrose ratio and drive cell expansion by turgor generated due to influx of osmotic solutes (Koch, 2004; Sergeeva et al., 2006; Morey et al., 2019). Consequently, *vin2* and *vin3* mutations lead to decreased seed size and shorter and lighter panicles (Morey et al., 2018; Lee et al., 2019; Morey et al., 2019; Xu et al., 2019; Deng et al., 2020). Other than being present as a candidate gene in the above-mentioned grain chalk-associated QTLs, *OsVIN2* has not been directly implicated in the grain chalkiness. However, the *OsCWIN* member *OsCIN2* mapped to the grain incomplete filling 1 (*OsGIF1*) locus was associated with grain filling and chalkiness in rice (Wang et al., 2008; Wang et al., 2018). Expression of GIF1 is unaffected, whereas both *OsVIN2* and *VIN3* are down regulated by high temperature and therefore it is plausible that the reduced expression of *VINs* contribute to reduced starch synthesis leading to increased grain chalkiness.

Glucose and fructose are also generated from phosphoenolpyruvate (PEP) which is synthesized by the action of pyruvate orthophosphate dikinase (*PPDK*) enzyme. Rice, a C3 plant has 3 *PPDK* genes, *OsPPDKA* encoding a cytosolic isoform (*OsPPDKA*); *OsPPDKB*, expressing a cytosolic *PPDKB* (cy*OsPPDKB*); and the chloroplastic *PPDK* (ch*OsPPDKB*) (Imaizumi et al., 1997; Moons et al., 1998). Our study identified *OsPPDKA* as one of the candidate genes from MQTL 3.5 [contributing QTLs qPGWC3.1, qPGWC3.2, qWB, qWBA, qWBK, qET and qCA (Kobayashi et al., 2007; Mei et al., 2013; Kobayashi et al., 2013; Peng et al., 2014a; Wang et al., 2016)]. It had only 2 non-synonymous haplotypes in the 3K panel, which could not be correlated with the chalk phenotype. Mutation in cy*PPDKB* gene resulted in chalky grains and a floury endosperm (Kang et al., 2005). As compared to *PPDKA*, *PPDKB* has a higher expression during the milky stage in rice kernels (Kang et al., 2005). On the other hand, *PPDKB* is down-regulated, while *PPDKA* is induced in response to heat stress during grain filling in Nipponbare (Yamakawa et al., 2007). We believe that *PPDKB* and *PPDKA* have compensatory roles and their stoichiometric levels determine quantity and quality of starch under high temperature. However, this hypothesis needs further validation by comparing the relative expression of *PPDKA* and *PPDKB* in rice accessions having contrasting grain chalk phenotypes.

One of the key regulatory steps in starch biosynthesis is conversion of glucose-1-phosphate to ADP-glucose catalyzed by the heterotetrameric enzyme complex (Okita, 1992; Tetlow et al., 2004; Lee et al., 2007). The heterotetramer has 2 small (*OsAGPS1* and *AGPS2a/b*) and 2 large subunits (*AGPL1-4*). *AGPL1*, was identified as a candidate gene in MQTL 5.5 [contributed by QTLs qPGWC5.1 and qPGWC5.2 (Chen et al., 2016; Wang et al., 2016)]. Interestingly, we did not find any non-synonymous variations in *AGPL1* throughout the 3K rice genome panel. This is in agreement with Lu and Park (2012) who also could not find any variation in the *AGPL1* gene in 104 rice accessions. ADP-glucose pyrophosphorylase genes, *AGPL2* and *AGPS2b*, exhibit polymorphisms that do not alter their protein (Kharabian-Masouleh et al., 2011; Wei et al., 2017) and conservation of protein sequence is likely a general feature of all the ADP-glucose pyrophosphorylase enzyme subunits. *AGPL1* is localized in amyloplasts and is expressed in the early stages of grain filling, whereas *AGPL2* is localized to cytosol and is expressed in the middle to late stages of grain filling (Lee et al., 2007) and it is likely that their spatio-temporal expression determines the quantity and quality of starch.

Plants exposed to high temperatures require energy to maintain cellular homoeostasis (Sadok and Jagadish, 2020) for which it breaks down reserve starch by increasing the expression and activity of α-amylases in the ripening grains. The timing of HT stress is crucial because if it happens during the grain filling stages, increased expression and activity of α-amylases affects the formation of starch granules and at the same time increases the hydrolysis of the already deposited starch, ultimately producing chalky grains (Zakaria et al., 2002; Iwasawa et al., 2009). RNAi-mediated suppression of α-amylases reduces grain chalk formation

under high temperature, conversely their overexpression enhances chalk even under ambient temperatures (Asatsuma et al., 2006; Hakata et al., 2012; Nakata et al., 2017). *RAG2*, a candidate gene from MQTL7.1 [contributing QTLs qDPGWC7.1, qDPGWC7.2, qPGWC7.1, qPGWC7.2, qPGWC7.3 (Wang et al., 2016)] is a 16-kDa alpha-amylase/trypsin inhibitor in rice. It is expressed in the developing seeds and is involved in both the regulation of grain yield and quality (Zhou et al., 2017). Rice grains of *RAG*2-RNAi lines display high chalk with loosely packed starch as compared to the wild type (Zhou et al., 2017). *RAL2*, also a candidate gene from MQTL 7.1, displays high homology with *RAG2* and may have a similar function. We hypothesize that both *RAG2* and *RAL2* proteins might inhibit α-amylase activity under high temperature but further experimental validation is needed. Our study revealed that *RAG2* and *RAL2* genes have 4 and 2 haplotypes respectively, in the 3K RG panel. We found 2 haplotypes each for *RAG2* and *RAL2* in the subset panel of 60 accessions, which did not correlate with the degree of chalkiness.

We also discovered β-amylase, another starch hydrolyzing enzyme of amylase family, as a candidate gene from MQTL 3.1 [contributing QTLs qWCR, qBW, qWB, qPGWC3.1 and qPGWC 3.2 (Chen et al., 2011; Wada et al., 2015; Wang et al., 2016; Yun et al., 2016)]. Haplotype analysis in the 3K RG panel revealed that beta-amylase is conserved and had no haplotypic variations. The expression of beta-amylase increases under high temperature (Yamakawa et al., 2007) and its activity in chalky grains is higher than in non-chalky grains (Nakamura et al., 2021). An endosperm-specific β-amylase gene from barley, *ESDBamy* is expressed in early endosperm and a shrunken endosperm mutant shows reduced starch deposition due to high levels of its expression (Jung et al., 2001). Absence of non-synonymous haplotypes indicate that differential steady state mRNA levels of β-amylase gene might be a contributing factor for variable amounts of starch in different accessions by inequitably affecting the inhibition of starch deposition. Both *RAG2* and *RAL2* are induced by high temperature and it is plausible that α- and β-amylases are targets of *RAG2* and *RAL2*, respectively. Further experiments involving expression analysis of these genes in multiple accessions with varying degrees of chalk, protein-protein interactions and enzymatic assays will be required to validate these possibilities and possibly provide additional strategies to regulate formation of grain chalk in rice.

Another major group of candidate genes were either seed storage proteins or those involved in nitrogen metabolism. We identified 5 genes each for seed storage proteins and nitrogen metabolism pathway in our meta-QTL analysis. Since both starch and proteins in grains are synthesized using sugar as the carbon source (Bao et al., 2015; Xin et al., 2019), an optimal coordination between the C and N metabolism during grain filling is crucial, but less worked upon. Reports suggest that the nitrogen metabolism pathway and synthesis of storage proteins in particular may lead to grain chalk formation in rice (Han et al., 2012; Xi et al., 2021). A significant negative correlation between prolamin content and chalkiness has also been discovered (Wang et al., 2021). It would be interesting to systematically investigate these promising candidate genes to define their roles in grain chalkiness.

5.2 Natural variations of GBSSI and SSIIa influence the quality of rice grains

Amylose and amylopectin synthesis in the rice endosperm are regulated GBSSI and SSIIa enzymes, respectively, genes for which are located on chromosome 6 (Wang et al., 1990, Figure 5A). *GBSSI*, a candidate gene identified in the current study from MQTL 6.1, was found in 14 mapped QTLs [qWCR, qCR, qWBA, qPCG1, qPCG2, qDC1, qDC2, qDC3, qACE, qPGWC1, qPGWC2, qCA, qWC and qWB (Peng et al., 2014a; Mei et al., 2013; Kobayashi et al., 2007; Zheng et al., 2012; Lu et al., 2013; Gao et al., 2016a)]. Similarly, MQTL 6.3 from which *SSIIa* gene was identified as a candidate, was derived from 7 overlapping QTLs [qDPGWC, qWCR, qCA, qET, qCA, qch, qWBR (Wang et al., 2016; Mei et al., 2013; Tan et al., 2000; Peng et al., 2014a; Yun et al., 2016)]. Such a high number of QTLs from many different studies signifies that these genomic regions and the underlying candidate genes immensely contribute to the grain chalk trait.

It is known that the diversity of grain amylose content is largely due to allelic variations at the Wx locus (Tian et al., 2009; Biselli et al., 2014; Zhang et al., 2019). To date, at least eight Wx alleles, Wx^{lv}, Wx^{a}, Wx^{in}, Wx^{b}, $Wx^{op/hp}$, Wx^{mq}, Wx^{mp}, and wx, have been shown to be associated with the five possible amylose types observed in rice cultivars (Cai et al., 1998; Sato et al., 2002; Larkin and Park, 2003; Wanchana et al., 2003; Mikami et al., 2008; Liu et al., 2009; Yang et al., 2013; Zhang et al., 2019). In the present study we identified Wx alleles, Wx^{lv}, Wx^{a}, Wx^{b} and $Wx^{op/hp}$ leading to the detection of 4 *GBSSI* haplotypes. Previously Wx^{lv} has been associated with high amylose content (Supplementary Table 10, Zhang et al., 2019), which could result in a low chalk accumulation phenotype. Wx^{lv} corresponded with 4 low, 4 intermediate and 3 high chalk phenotypes in our study, indicating that allelic combinations of other alleles with the Wx^{lv} allele possibly determines the degree of chalkiness. While Wx^{a} allele was primarily associated with 5 intermediate and 1 high chalk type rice, accessions having high chalk almost always had Wx^{b} and $Wx^{op/hp}$ alleles.

Most of the indica rice cultivars possess active-type *SSIIa* (wild-type), whereas typical japonica rice cultivars possess mutant *SSIIa* harboring three amino acid substitutions in the active site, thus exhibiting only 10% of *SSIIa* activity relative to indica rice (Nakamura et al., 2005). Low SSIIa activity in japonica rice leads to a reduction in amylopectin branches with degree of polymerization (DP) of 13-24, and an increase in short amylopectin chains with DP ≤ 12 (Umemoto et al., 1999; Umemoto et al., 2002; Nakamura et al., 2005). The four haplotypic combinations in our study were derived from the three substitutions. The GGC haplotype was associated with low and intermediate chalk accessions, intermediate chalk accessions had AGC, while AGT and AAC were linked with high chalk accessions, indicating that a combination of substitutions in the active site of SSIIa enzyme is a critical determinant of chalk formation.

Since, elevated temperatures affect the accumulation of starch (Umemoto et al., 1995) by disturbing the proportion of amylose to amylopectin (Inouchi et al., 2000), therefore, understanding how various allelic combinations of closely located starch biosynthetic genes, *GBSSI* and *SSIIa*, affect starch properties and grain chalkiness

is of utmost importance. Haplotypes GGGAC, TGAAC in *GBSSI* (Cai et al., 1998; Mikami et al., 1999; Larkin and Park, 2003) and AGT and AAC in *SSIIa* (Umemoto et al., 1999; Umemoto et al., 2002; Nakamura et al., 2005) have low amylose and low amylopectin. Our chalky score and scanning electron microscopy data revealed that the combinations GGGAC/AGT, TGAAC/AGT and TGAAC/AAC in accessions invariably led to high chalk accumulation (Figure 9B). Starch quality in rice grains is a trait that has evolved in response to strong selection of starch synthesis pathway genes during domestication (Whitt et al., 2002). Rice varieties with high amylose levels (20–30%) tend to form discrete, non-cohesive grains when cooked and are found in South and Southeast Asian rices classified as indica and tropical japonica variety groups, whereas varieties with low amylose, a characteristic of temperate japonica variety, are common in northeast Asia and form cohesive cooked grains (Morishima and Sano, 1992; Juliano and Villareal, 1993). Because *GBSSI* and *SSIIa* are closely located on chromosome 6, it is possible that these two genes may have undergone co-selection during domestication. Further, rice germplasm harboring different haplotypic combinations of these starch synthesizing genes have likely been selected on the basis of regional palatability. Nevertheless, the haplotype combinations associated with low grain chalk will be helpful in developing elite varieties with high head rice yield.

6 Conclusions and future directions

Meta-QTL analysis is a powerful tool for detecting the most precise and concurrent QTLs for any agronomic trait in crops. Identification of differentially expressed MQTL genes utilizing trait-specific transcriptome datasets, provides additional means for refining the process of shortlisting the significant candidate genes. Harnessing information of haplotypic diversity from sequenced accessions and associating it with the phenotype can provide information on novel allelic/haplotype combinations contributing to traits of economic importance. Our work combined all these approaches and unveiled haplotype combinations in two main starch synthesis genes, which can be genetically linked with varying degrees of grain chalkiness in rice. The favorable allelic combinations that promote biosynthesis of optimum quality starch and thereby contribute to reduced grain chalk, can be backcrossed with high-yielding elite varieties or CRISPR-Cas based prime editing to generate superior rice plants with low grain chalkiness and high HRY traits.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

MA and AK planned and designed the research. AK, DS, and KK performed the meta-analysis. AK and DS mined the candidate genes and performed the in-silico promoter analysis. AK and S analyzed expression data. AK, DS, CW, DC, AP, GC, and MS did the phenotyping. AK, MI, and PK performed the haplotype analysis. AK and PS performed the haplo-pheno analysis. SK and VV prepared the SEM images. AK, MA, SK-A, AG, and SJ wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Support to the laboratory of MA was provided by National Agricultural Science Fund (NASF) grant (NASF/ABP-7O14/2018-19) and Faculty Research Program (FRP) grant ((IoE/FRP/LS/2020/27) from Institute of Eminence (IoE), Delhi University.

Acknowledgments

SK-A acknowledges the Faculty Research Program (FRP) grant (IoE/2021/12/FRP) from Institute of Eminence (IoE), Delhi University. AK is grateful for the research fellowship from the Council of Scientific and Industrial Research (CSIR), India.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1133115/full#supplementary-material

References

Abdullah-Zawawi, M. R., Ahmad-Nizammuddin, N. F., Govender, N., Harun, S., Mohd-Assaad, N., and Mohamed-Hussein, Z. A. (2021). Comparative genome-wide analysis of WRKY, MADS-box and MYB transcription factor families in arabidopsis and rice. *Sci. Rep.* 11 (1), 1–18. doi: 10.1038/s41598-021-99206-y

Aoki, N., Hirose, T., Scofield, G. N., Whitfeld, P. R., and Furbank, R. T. (2003). The sucrose transporter gene family in rice. *Plant Cell Physiol.* 44 (3), 223–232. doi: 10.1093/pcp/pcg030

Arcade, A., Labourdette, A., Falque, M., Mangin, B., Chardon, F., Charcosset, A., et al. (2004). BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* 20 (14), 2324–2326. doi: 10.1093/bioinformatics/bth230

Asatsuma, S., Sawada, C., Kitajima, A., Asakura, T., and Mitsui, T. (2006). α-amylase affects starch accumulation in the rice grain. *J. Appl. Glycosci.* 53, 187–192. doi: 10.5458/jag.53.187

Bai, A. N., Lu, X. D., Li, D. Q., Liu, J. X., and Liu, C. M. (2016). NF-YB1-regulated expression of sucrose transporters in aleurone facilitates sugar loading to rice endosperm. *Cell Res.* 26 (3), 384–388. doi: 10.1038/cr.2015.116

Bakku, R. K., Rakwal, R., Shibato, J., Cho, K., Kikuchi, S., Yonekura, M., et al. (2020). Transcriptomics of mature rice (Oryza sativa l. koshihikari) seed under hot conditions by DNA microarray analyses. *Atmosphere* 11 (5), 528.

Bao, A., Zhao, Z., Ding, G., Shi, L., Xu, F., and Cai, H. (2015). The stable level of glutamine synthetase 2 plays an important role in rice growth and in carbon-nitrogen metabolic balance. *Int. J. Mol. Sci.* 16 (6), 12713–12736. doi: 10.3390/ijms160612713

Bazrkar-Khatibani, L., Fakheri, B. A., Hosseini-Chaleshtori, M., Mahender, A., Mahdinejad, N., and Ali, J. (2019). Genetic mapping and validation of quantitative trait loci (QTL) for the grain appearance and quality traits in rice (Oryza sativa l.) by using recombinant inbred line (RIL) population. *Int. J. Genomics* 2019.

Bian, J. M., Shi, H., Li, C. J., Zhu, C. L., Yu, Q. Y., Peng, X. S., et al. (2013). QTL mapping and correlation analysis for 1000-grain weight and percentage of grains with chalkiness in rice. *J. Genet.* 92 (2), 281–287. doi: 10.1007/s12041-013-0267-6

Biselli, C., Cavalluzzo, D., Perrini, R., Gianinetti, A., Bagnaresi, P., Urso, S., et al. (2014). Improvement of marker-based predictability of apparent amylose content in japonica rice through GBSSI allele mining. *Rice* 7 (1), 1–18. doi: 10.1186/1939-8433-7-1

Butardo, V. M., Fitzgerald, M. A., Bird, A. R., Gidley, M. J., Flanagan, B. M., Larroque, O., et al. (2011). Impact of down-regulation of starch branching enzyme IIb in rice by artificial microRNA-and hairpin RNA-mediated RNA silencing. *J. Exp. Bot.* 62 (14), 4927–4941. doi: 10.1093/jxb/err188

Cai, X. L., Wang, Z. Y., Xing, Y. Y., Zhang, J. L., and Hong, M. M. (1998). Aberrant splicing of intron 1 leads to the heterogeneous 5′ UTR and decreased expression of waxy gene in rice cultivars of intermediate amylose content. *Plant J.* 14 (4), 459–465. doi: 10.1046/j.1365-313X.1998.00126.x

Caliński, T., and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 3(1), 1–27.

Chandusingh, P. R., Singh, N. K., Prabhu, K. V., Vinod, K. K., and Singh, A. K. (2013). Molecular mapping of quantitative trait loci for grain chalkiness in rice (Oryza sativa l.). *Indian J. Genet. Plant Breed.* 73 (3), 244–251. doi: 10.5958/j.0975-6906.73.3.036

Chen, L., Gao, W., Chen, S., Wang, L., Zou, J., Liu, Y., et al. (2016). High-resolution QTL mapping for grain appearance traits and co-localization of chalkiness-associated differentially expressed candidate genes in rice. *Rice* 9 (1), 1–17. doi: 10.1186/s12284-016-0121-6

Chen, H. M., Zhao, Z. G., Jiang, L., Wan, X. Y., Liu, L. L., Wu, X. J., et al. (2011). Molecular genetic analysis on percentage of grains with chalkiness in rice (Oryza sativa l.). *Afr. J. Biotechnol.* 10 (36), 6891–6903.

Cho, J. I., Kim, H. B., Kim, C. Y., Hahn, T. R., and Jeon, J. S. (2011). Identification and characterization of the duplicate rice sucrose synthase genes OsSUS5 and OsSUS7 which are associated with the plasma membrane. *Molecules Cells* 31 (6), 553–561. doi: 10.1007/s10059-011-1038-y

Cooper, N. T. W., Siebenmorgen, T. J., and Counce, P. A. (2008). Effects of nighttime temperature during kernel development on rice physicochemical properties. *Cereal Chem.* 85 (3), 276–282. doi: 10.1094/CCHEM-85-3-0276

Dai, L., Wang, L., Leng, Y., Yang, Y., Huang, L., Chen, L., et al. (2016). Quantitative trait loci mapping for appearance quality in short-grain rice. *Crop Sci.* 56 (4), 1484–1492. doi: 10.2135/cropsci2015.07.0404

Deng, X., Han, X., Yu, S., Liu, Z., Guo, D., He, Y., et al. (2020). OsINV3 and its homolog, OsINV2, control grain size in rice. *Int. J. Mol. Sci.* 21 (6), 2199. doi: 10.3390/ijms21062199

Denyer, K., Clarke, B., Hylton, C., Tatge, H., and Smith, A. M. (1996). The elongation of amylose and amylopectin chains in isolated starch granules. *Plant J.* 10 (6), 1135–1143. doi: 10.1046/j.1365-313X.1996.10061135.x

Denyer, K., Hylton, C. M., Jenner, C. F., and Smith, A. M. (1995). Identification of multiple isoforms of soluble and granule-bound starch synthase in developing wheat endosperm. *Planta* 196 (2), 256–265. doi: 10.1007/BF00201382

Edwards, J. D., Jackson, A. K., and McClung, A. M. (2017). Genetic architecture of grain chalk in rice and interactions with a low phytic acid locus. *Field Crops Res.* 205, 116–123. doi: 10.1016/j.fcr.2017.01.015

Eizenga, G. C., Chen, M. H., Jia, M. H., Jackson, A. K., and Edwards, J. D. (2019). Registration of the estrela× NSFTV199 rice recombinant inbred line mapping population. *J. Plant Registrations* 13 (3), 469–478. doi: 10.3198/jpr2019.04.0019crmp

Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K., and Rask, L. (1999). Interaction between composite elements in the napA promoter: both the b-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol. Biol.* 40 (4), 699–709. doi: 10.1023/A:1006206124512

FAO (2019) *Food and agricultural organization of the united nations. FAOSTAT*. Available at: http://www.fao.org/faostat/en/#data/QC.

Fujita, N., Yoshida, M., Kondo, T., Saito, K., Utsumi, Y., Tokunaga, T., et al. (2007). Characterization of SSIIIa-deficient mutants of rice: the function of SSIIIa and pleiotropic effects by SSIIIa deficiency in the rice endosperm. *Plant Physiol.* 144 (4), 2009–2023. doi: 10.1104/pp.107.102533

Gao, Y., Liu, C., Li, Y., Zhang, A., Dong, G., Xie, L., et al. (2016b). QTL analysis for chalkiness of rice and fine mapping of a candidate gene for qACE9. *Rice* 9 (1), 1–10. doi: 10.1186/s12284-016-0114-5

Gao, F. Y., Zeng, L. H., Ling, Q. I. U., Lu, X. J., Ren, J. S., Wu, X. T., et al. (2016a). QTL mapping of grain appearance quality traits and grain weight using a recombinant inbred population in rice (Oryza sativa l.). *J. Integr. Agric.* 15 (8), 1693–1702. doi: 10.1016/S2095-3119(15)61259-X

Garretón, V., Carpinelli, J., Jordana, X., and Holuigue, L. (2002). The as-1 promoter element is an oxidative stress-responsive element and salicylic acid activates it *via* oxidative species. *Plant Physiol.* 130 (3), 1516–1526. doi: 10.1104/pp.009886

Ge, S. X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36 (8), 2628–2629. doi: 10.1093/bioinformatics/btz931

Goffinet, B., and Gerber, S. (2000). Quantitative trait loci: a meta-analysis. *Genetics* 155 (1), 463–473. doi: 10.1093/genetics/155.1.463

Guo, T., Liu, X., Wan, X., Weng, J., Liu, S., Liu, X., et al. (2011). Identification of a stable quantitative trait locus for percentage grains with white chalkiness in rice (Oryza sativa). *J. Integr. Plant Biol.* 53 (8), 598–607. doi: 10.1111/j.1744-7909.2011.01041.x

Hakata, M., Kuroda, M., Miyashita, T., Yamaguchi, T., Kojima, M., Sakakibara, H., et al. (2012). Suppression of α-amylase genes improves quality of rice grain ripened under high temperature. *Plant Biotechnol. J.* 10 (9), 1110–1117. doi: 10.1111/j.1467-7652.2012.00741.x

Han, X., Wang, Y., Liu, X., Jiang, L., Ren, Y., Liu, F., et al. (2012). The failure to express a protein disulphide isomerase-like protein results in a floury endosperm and an endoplasmic reticulum stress response in rice. *J. Exp. Bot.* 63 (1), 121–130. doi: 10.1093/jxb/err262

Hao, W., Zhu, M. Z., Gao, J. P., Sun, S. Y., and Lin, H. X. (2009). Identification of quantitative trait loci for rice quality in a population of chromosome segment substitution lines. *J. Integr. Plant Biol.* 51 (5), 500–512. doi: 10.1111/j.1744-7909.2009.00822.x

He, P., Li, S. G., Qian, Q., Ma, Y. Q., Li, J. Z., Wang, W. M., et al. (1999). Genetic analysis of rice grain quality. *Theor. Appl. Genet.* 98 (3), 502–508. doi: 10.1007/s001220051098

Hirose, T., Scofield, G. N., and Terao, T. (2008). An expression analysis profile for the entire sucrose synthase gene family in rice. *Plant Sci.* 174 (5), 534–543. doi: 10.1016/j.plantsci.2008.02.009

Hossain, M., Bose, M. L., and Mustafi, B. A. (2006). Adoption and productivity impact of modern rice varieties in Bangladesh. *developing economies* 44 (2), 149–166. doi: 10.1111/j.1746-1049.2006.00011.x

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., et al. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinf.* 2008. doi: 10.1155/2008/420747

Hu, T., He, S., Yang, G., Zeng, H., Wang, G., Chen, Z., et al. (2011). Isolation and characterization of a rice glutathione s-transferase gene promoter regulated by herbicides and hormones. *Plant Cell Rep.* 30 (4), 539–549. doi: 10.1007/s00299-010-0964-z

Huang, J. W., Chen, J. T., Yu, W. P., Shyur, L. F., Wang, A. Y., Sung, H. Y., et al. (1996). Complete structures of three rice sucrose synthase isogenes and differential regulation of their expressions. *Biosci. biotechnol. Biochem.* 60 (2), 233–239. doi: 10.1271/bbb.60.233

Huang, Y. C., Hsiang, E. C., Yang, C. C., and Wang, A. Y. (2016). New insight into the catalytic properties of rice sucrose synthase. *Plant Mol. Biol.* 90 (1), 127–135. doi: 10.1007/s11103-015-0401-3

Imaizumi, N., Ku, M. S., Ishihara, K., Samejima, M., Kaneko, S., and Matsuoka, M. (1997). Characterization of the gene for pyruvate, orthophosphate dikinase from rice, a C3 plant, and a comparison of structure and expression between C3 and C4 genes for this protein. *Plant Mol. Biol.* 34 (5), 701–716. doi: 10.1023/A:1005884515840

Inouchi, N., Ando, H., Asaoka, M., Okuno, K., and Fuwa, H. (2000). The effect of environmental temperature on distribution of unit chains of rice amylopectin. *Starch-Stärke* 52 (1), 8–12. doi: 10.1002/(SICI)1521-379X(200001)52:1<8::AID-STAR8>3.0.CO;2-Q

Ishimaru, T., Parween, S., Saito, Y., Shigemitsu, T., Yamakawa, H., Nakazono, M., et al. (2019). Laser microdissection-based tissue-specific transcriptome analysis reveals a novel regulatory network of genes involved in heat-induced grain chalk in rice endosperm. *Plant Cell Physiol.* 60 (3), 626–642. doi: 10.1093/pcp/pcy233

Isshiki, M., Morino, K., Nakajima, M., Okagaki, R. J., Wessler, S. R., Izawa, T., et al. (1998). A naturally occurring functional allele of the rice waxy locus has a GT to TT mutation at the 5′ splice site of the first intron. *Plant J.* 15 (1), 133–138. doi: 10.1046/j.1365-313X.1998.00189.x

Iwasawa, N., Umemoto, T., Hiratsuka, M., Nitta, Y., Matsuda, T., and Kondo, M. (2009). Structural characters of milky-white rice grains caused by high temperature and shading during grain-filling. *Jpn J. Crop Sci.* 78, 322–323.

Juliano, B. O., and Villareal, C. P. (1993). Grain quality evaluation of world rices. *Int. Rice Res. Inst*.

Jung, W., Skadsen, R. W., and Peterson, D. M. (2001). Characterization of a novel barley β-amylase gene expressed only during early grain development. *Seed Sci. Res.* 11 (4), 325–333.

Kabir, S. M. T., Hossain, M. S., Bashar, K. K., Honi, U., Ahmed, B., Emdad, E. M., et al. (2021). Genome-wide identification and expression profiling of AP2/ERF superfamily genes under stress conditions in dark jute (Corchorus olitorius l.). *Ind. Crops Products* 166, 113469. doi: 10.1016/j.indcrop.2021.113469

Kang, H. G., Park, S., Matsuoka, M., and An, G. (2005). White-core endosperm floury endosperm-4 in rice is generated by knockout mutations in the C4-type pyruvate orthophosphate dikinase gene (OsPPDKB). *Plant J.* 42 (6), 901–911. doi: 10.1111/j.1365-313X.2005.02423.x

Kepiro, J. L., McClung, A. M., Chen, M. H., Yeater, K. M., and Fjellstrom, R. G. (2008). Mapping QTLs for milling yield and grain characteristics in a tropical japonica long grain cross. *J. Cereal Sci.* 48 (2), 477–485. doi: 10.1016/j.jcs.2007.12.001

Khahani, B., Tavakol, E., and Shariati J, V. (2019). Genome-wide meta-analysis on yield and yield-related QTLs in barley (Hordeum vulgare l.). *Mol. Breed.* 39 (4), 1–16. doi: 10.1007/s11032-019-0962-y

Khahani, B., Tavakol, E., Shariati, V., and Rossini, L. (2021). Meta-QTL and ortho-MQTL analyses identified genomic regions controlling rice yield, yield-related traits and root architecture under water deficit conditions. *Scientific reports* 11(1), 6942.

Kharabian-Masouleh, A., Waters, D. L., Reinke, R. F., and Henry, R. J. (2011). Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. *Plant Biotechnol. J.* 9 (9), 1074–1085. doi: 10.1111/j.1467-7652.2011.00629.x

Kobayashi, A., Genliang, B., Shenghai, Y., and Tomita, K. (2007). Detection of quantitative trait loci for white-back and basal-white kernels under high temperature stress in japonica rice varieties. *Breed. Sci.* 57 (2), 107–116. doi: 10.1270/jsbbs.57.107

Kobayashi, A., Sonoda, J., Sugimoto, K., Kondo, M., Iwasawa, N., Hayashi, T., et al. (2013). Detection and verification of QTLs associated with heat-induced quality decline of rice (Oryza sativa l.) using recombinant inbred lines and near-isogenic lines. *Breed. Sci.* 63 (3), 339–346.

Koch, K. (2004). Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Curr. Opin. Plant Biol.* 7 (3), 235–246. doi: 10.1016/j.pbi.2004.03.014

Kong, W., Zhang, C., Qiang, Y., Zhong, H., Zhao, G., and Li, Y. (2020). Integrated RNA-seq analysis and meta-QTLs mapping provide insights into cold stress response in rice seedling roots. *Int. J. Mol. Sci.* 21 (13), 4615. doi: 10.3390/ijms21134615

Kumar, I. S., and Nadarajah, K. (2020). A meta-analysis of quantitative trait loci associated with multiple disease resistance in rice (Oryza sativa l.). *Plants* 9 (11), 1491. doi: 10.3390/plants9111491

Larkin, P. D., and Park, W. D. (2003). Association of waxy gene single nucleotide polymorphisms with starch characteristics in rice (Oryza sativa l.). *Mol. Breed.* 12 (4), 335–339. doi: 10.1023/B:MOLB.0000006797.51786.92

Le, T. T. (2016). Effects of climate change on rice yield and rice market in Vietnam. *J. Agric. Appl. Economics* 48 (4), 366–382. doi: 10.1017/aae.2016.21

Lee, S. K., Hwang, S. K., Han, M., Eom, J. S., Kang, H. G., Han, Y., et al. (2007). Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (Oryza sativa l.). *Plant Mol. Biol.* 65 (4), 531–546. doi: 10.1007/s11103-007-9153-z

Lee, D. W., Lee, S. K., Rahman, M. M., Kim, Y. J., Zhang, D., and Jeon, J. S. (2019). The role of rice vacuolar invertase2 in seed size control. *Molecules Cells* 42 (10), 711.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30 (1), 325–327. doi: 10.1093/nar/30.1.325

Lestari, P., Lee, G., Ham, T. H., Woo, M. O., Piao, R., Jiang, W., et al. (2011). Single nucleotide polymorphisms and haplotype diversity in rice sucrose synthase 3. *J. Heredity* 102 (6), 735–746. doi: 10.1093/jhered/esr094

Li, Y., Fan, C., Xing, Y., Yun, P., Luo, L., Yan, B., et al. (2014). Chalk5 encodes a vacuolar h+-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* 46 (4), 398–404. doi: 10.1038/ng.2923

Li, J. Y., Wang, J., and Zeigler, R. S. (2014). The 3,000 rice genomes project: new opportunities and challenges for future rice research. *Gigascience* 3 (1), 2047–217X. doi: 10.1186/2047-217X-3-8

Li, J., Xiao, J., Grandillo, S., Jiang, L., Wan, Y., Deng, Q., et al. (2004). QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (O. sativa l.) and African (O. glaberrima s.) rice. *Genome* 47 (4), 697–704.

Lin, Z., Wang, Z., Zhang, X., Liu, Z., Li, G., Wang, S., et al. (2017). Complementary proteome and transcriptome profiling in developing grains of a notched-belly rice mutant reveals key pathways involved in chalkiness formation. *Plant Cell Physiol.* 58 (3), 560–573. doi: 10.1093/pcp/pcx001

Lisle, A. J., Martin, M., and Fitzgerald, M. A. (2000). Chalky and translucent rice grains differ in starch composition and structure and cooking properties. *Cereal Chem.* 77 (5), 627–632. doi: 10.1094/CCHEM.2000.77.5.627

Liu, X., Guo, T., Wan, X., Wang, H., Zhu, M., Li, A., et al. (2010b). Transcriptome analysis of grain-filling caryopses reveals involvement of multiple regulatory pathways in chalky grain formation in rice. *BMC Genomics* 11 (1), 1–15. doi: 10.1186/1471-2164-11-730

Liu, L., Ma, X., Liu, S., Zhu, C., Jiang, L., Wang, Y., et al. (2009). Identification and characterization of a novel waxy allele from a yunnan rice landrace. *Plant Mol. Biol.* 71 (6), 609–626. doi: 10.1007/s11103-009-9544-4

Liu, X., Wang, Y., and Wang, S. W. (2012). QTL analysis of percentage of grains with chalkiness in japonica rice (Oryza sativa). *Genet. Mol. Res.* 11 (1), 717–724. doi: 10.4238/2012.March.22.1

Liu, F., Xu, W., Song, Q., Tan, L., Liu, J., Zhu, Z., et al. (2013). Microarray-assisted fine-mapping of quantitative trait loci for cold tolerance in rice. *Mol. Plant* 6 (3), 757–767. doi: 10.1093/mp/sss161

Liu, G., Zhu, H., Liu, S., Zeng, R., Zhang, Z., Li, W., et al. (2010a). Unconditional and conditional QTL mapping for the developmental behavior of tiller number in rice (Oryza sativa l.). *Genetica* 138 (8), 885–893. doi: 10.1007/s10709-010-9471-y

Lu, B., Yang, C., Xie, K., Zhang, L., Wu, T., Li, L., et al. (2013). Quantitative trait loci for grain-quality traits across a rice F2 population and backcross inbred lines. *Euphytica* 192 (1), 25–35. doi: 10.1007/s10681-012-0822-x

Lu, F. H., and Park, Y. J. (2012). Sequence variations in OsAGPase significantly associated with amylose content and viscosity properties in rice (Oryza sativa L.). *Genetics Research* 94(4), 179–189.

Lyman, N. B., Jagadish, K. S., Nalley, L. L., Dixon, B. L., and Siebenmorgen, T. (2013). Neglecting rice milling yield and quality underestimates economic losses from high-temperature stress. *PloS One* 8 (8), e72157. doi: 10.1371/journal.pone.0072157

MacNeill, G. J., Mehrpouyan, S., Minow, M. A., Patterson, J. A., Tetlow, I. J., Emes, M. J., et al. (2017). Starch as a source, starch as a sink: the bifunctional role of starch in carbon allocation. *J. Exp. Bot.* 68 (16), 4433–4453. doi: 10.1093/jxb/erx291

Mansueto, L., Fuentes, R. R., Chebotarov, D., Borja, F. N., Detras, J., Abriol-Santos, J. M., et al. (2016). SNP-seek II: A resource for allele mining and analysis of big genomic data in oryza sativa. *Curr. Plant Biol.* 7, 16–25. doi: 10.1016/j.cpb.2016.12.003

Martinez-Pastor, M. T., Marchler, G., Schüller, C., Marchler-Bauer, A., Ruis, H., and Estruch, F. (1996). The saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO J.* 15 (9), 2227–2235. doi: 10.1002/j.1460-2075.1996.tb00576.x

Mei, D., Yu, Z., Gong, C., Chen, X., Zhu, Q., and Wang, S. (2018). Validation and fine mapping a multifunction region qPCG10/qDC10 conferring chalkiness and grain shape of rice. *Crop Sci.* 58 (2), 639–649. doi: 10.2135/cropsci2017.07.0405

Mei, D. Y., Zhu, Y. J., Yu, Y. H., Fan, Y. Y., Huang, D. R., and Zhuang, J. Y. (2013). Quantitative trait loci for grain chalkiness and endosperm transparency detected in three recombinant inbred line populations of indica rice. *J. Integr. Agric.* 12 (1), 1–11. doi: 10.1016/S2095-3119(13)60199-9

Miao, Z. H., and Lam, E. (1995). Targeted disruption of the TGA3 locus in arabidopsis thaliana. *Plant J.* 7 (2), 359–365. doi: 10.1046/j.1365-313X.1995.7020359.x

Mikami, I., Aikawa, M., Hirano, H. Y., and Sano, Y. (1999). Altered tissue-specific expression at the wx gene of the opaque mutants in rice. *Euphytica* 105 (2), 91–97. doi: 10.1023/A:1003457209225

Mikami, I., Uwatoko, N., Ikeda, Y., Yamaguchi, J., Hirano, H. Y., Suzuki, Y., et al. (2008). Allelic diversification at the wx locus in landraces of Asian rice. *Theor. Appl. Genet.* 116 (7), 979–989. doi: 10.1007/s00122-008-0729-z

Mirdar Mansuri, R., Shobbar, Z. S., Babaeian Jelodar, N., Ghaffari, M., Mohammadi, S. M., and Daryani, P. (2020). Salt tolerance involved candidate genes in rice: an integrative meta-analysis approach. *BMC Plant Biol.* 20 (1), 1–14.

Miyahara, K., Wada, T., Sonoda, J. Y., Tsukaguchi, T., Miyazaki, M., Tsubone, M., et al. (2017). Detection and validation of QTLs for milky-white grains caused by high temperature during the ripening period in japonica rice. *Breed. Sci.* 67(4), 333–339. doi: 10.1270/jsbbs.16203

Moons, A., Valcke, R., and Van, M. (1998). Low-oxygen stress and water deficit induce cytosolic pyruvate orthophosphate dikinase (PPDK) expression in roots of rice, a C3 plant. *Plant J.* 15 (1), 89–98. doi: 10.1046/j.1365-313X.1998.00185.x

Morey, S. R., Hirose, T., Hashida, Y., Miyao, A., Hirochika, H., Ohsugi, R., et al. (2018). Genetic evidence for the role of a rice vacuolar invertase as a molecular sink strength determinant. *Rice* 11 (1), 1–13. doi: 10.1186/s12284-018-0201-x

Morey, S. R., Hirose, T., Hashida, Y., Miyao, A., Hirochika, H., Ohsugi, R., et al. (2019). Characterisation of a rice vacuolar invertase isoform, OsINV2, for growth and yield-related traits. *Funct. Plant Biol.* 46 (8), 777–785. doi: 10.1071/FP18291

Morishima, H. I. R. O. K. O., and Sano, Y. O. S. H. I. O. (1992). Evolutionary studies in cultivated rice. *Oxf. Surv. Evol. Biol.* 8, 135.

Muthayya, S., Sugimoto, J. D., Montgomery, S., and Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. *Ann. New york Acad. Sci.* 1324 (1), 7–14. doi: 10.1111/nyas.12540

Nakamura, Y., Francisco, P. B., Hosaka, Y., Sato, A., Sawada, T., Kubo, A., et al. (2005). Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between japonica and indica rice varieties. *Plant Mol. Biol.* 58 (2), 213–227. doi: 10.1007/s11103-005-6507-2

Nakamura, S., Satoh, A., Aizawa, M., and Ohtsubo, K. I. (2021). Characteristics of physicochemical properties of chalky grains of japonica rice generated by high temperature during ripening. *Foods* 11 (1), 97. doi: 10.3390/foods11010097

Nakata, M., Fukamatsu, Y., Miyashita, T., Hakata, M., Kimura, R., Nakata, Y., et al. (2017). High temperature-induced expression of rice α-amylases in developing endosperm produces chalky grains. *Front. Plant Sci.* 8, 2089. doi: 10.3389/fpls.2017.02089

Nelson, J. C., McClung, A. M., Fjellstrom, R. G., Moldenhauer, K. A. K., Boza, E., Jodari, F., et al. (2011). Mapping QTL main and interaction influences on milling quality in elite US rice germplasm. *Theor. Appl. Genet.* 122 (2), 291–309. doi: 10.1007/s00122-010-1445-z

Okada, S., Suehiro, M., Ebana, K., Hori, K., Onogi, A., Iwata, H., et al. (2017). Genetic dissection of grain traits in yamadanishiki, an excellent sake-brewing rice cultivar. *Theor. Appl. Genet.* 130 (12), 2567–2585. doi: 10.1007/s00122-017-2977-2

Okita, T. W. (1992). Is there an alternative pathway for starch synthesis? *Plant Physiol.* 100 (2), 560–564.

Peng, Y., Hu, Y., Mao, B., Xiang, H., Shao, Y., Pan, Y., et al. (2016). Genetic analysis for rice grain quality traits in the YVB stable variant line using RAD-seq. *Mol. Genet. Genomics* 291 (1), 297–307. doi: 10.1007/s00438-015-1104-9

Peng, B., Wang, L., Fan, C., Jiang, G., Luo, L., Li, Y., et al. (2014a). Comparative mapping of chalkiness components in rice using five populations across two environments. *BMC Genet.* 15 (1), 1–14. doi: 10.1186/1471-2156-15-49

Peng, C., Wang, Y., Liu, F., Ren, Y., Zhou, K., Lv, J., et al. (2014b). FLOURY ENDOSPERM 6 encodes a CBM 48 domain-containing protein involved in compound granule formation and starch synthesis in rice endosperm. *Plant J.* 77 (6), 917–930. doi: 10.1111/tpj.12444

Qin, Y., Kim, S. M., and Sohn, J. K. (2009). Genetic analysis and QTL mapping for grain chalkiness characteristics of brown rice (Oryza sativa l.). *Genes Genomics* 31 (2), 155–164. doi: 10.1007/BF03191148

Raza, Q., Riaz, A., Bashir, K., and Sabar, M. (2020). Reproductive tissues-specific meta-QTLs and candidate genes for development of heat-tolerant rice cultivars. *Plant Mol. Biol.* 104 (1), 97–112. doi: 10.1007/s11103-020-01027-6

Sadok, W., and Jagadish, S. K. (2020). The hidden costs of nighttime warming on yields. *Trends Plant Sci.* 25 (7), 644–651. doi: 10.1016/j.tplants.2020.02.003

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013). Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54 (2), e6–e6. doi: 10.1093/pcp/pcs183

Sano, Y. (1984). Differential regulation of waxy gene expression in rice endosperm. *Theor. Appl. Genet.* 68 (5), 467–473. doi: 10.1007/BF00254822

Sato, H., Suzuki, Y., Sakai, M., and Imbe, T. (2002). Molecular characterization of wx-mq, a novel mutant gene for low-amylose content in endosperm of rice (Oryza sativa l.). *Breed. Sci.* 52 (2), 131–135. doi: 10.1270/jsbbs.52.131

Sergeeva, L. I., Keurentjes, J. J., Bentsink, L., Vonk, J., van der Plas, L. H., Koornneef, M., et al. (2006). Vacuolar invertase regulates elongation of arabidopsis thaliana roots as revealed by QTL and mutant analysis. *Proc. Natl. Acad. Sci.* 103 (8), 2994–2999. doi: 10.1073/pnas.0511015103

Smith, A. M., Denyer, K., and Martin, C. (1997). The synthesis of the starch granule. *Annu. Rev. Plant Biol.* 48 (1), 67–87. doi: 10.1146/annurev.arplant.48.1.67

Sosnowski, O., Charcosset, A., and Joets, J. (2012). BioMercator V3: An upgrade of genetic map compilation and quantitative trait loci meta analysis algorithms. *Bioinformatics* 28 (15), 2082–2083. doi: 10.1093/bioinformatics/bts313

Sreenivasulu, N., Butardo, V. M.Jr., Misra, G., Cuevas, R. P., Anacleto, R., and Kavi Kishor, P. B. (2015). Designing climate-resilient rice with ideal grain quality suited for high-temperature stress. *J. Exp. Bot.* 66 (7), 1737–1748. doi: 10.1093/jxb/eru544

Sturm, A. (1999). Invertases. primary structures, functions, and roles in plant development and sucrose partitioning. *Plant Physiol.* 121 (1), 1–8.

Sun, Y., Lin, Z., Reinders, A., and Ward, J. M. (2012). Functionally important amino acids in rice sucrose transporter OsSUT1. *Biochemistry* 51 (15), 3284–3291. doi: 10.1021/bi201934h

Tabata, M., Hirabayashi, H., Takeuchi, Y., Ando, I., Iida, Y., and Ohsawa, R. (2007). Mapping of quantitative trait loci for the occurrence of white-back kernels associated with high temperatures during the ripening period of rice (Oryza sativa l.). *Breed. Sci.* 57 (1), 47–52. doi: 10.1270/jsbbs.57.47

Takehara, K., Murata, K., Yamaguchi, T., Yamaguchi, K., Chaya, G., Kido, S., et al. (2018). Thermo-responsive allele of sucrose synthase 3 (Sus3) provides high-temperature tolerance during the ripening stage in rice (Oryza sativa l.). *Breed. Sci.* 68(3), 336–342. doi: 10.1270/jsbbs.18007

Tan, Y. F., Xing, Y. Z., Li, J. X., Yu, S. B., Xu, C. G., and Zhang, Q. (2000). Genetic bases of appearance quality of rice grains in shanyou 63, an elite rice hybrid. *Theor. Appl. Genet.* 101 (5), 823–829. doi: 10.1007/s001220051549

Tashiro, T. W. I. F., and Wardlaw, I. F. (1991). The effect of high temperature on kernel dimensions and the type and occurrence of kernel damage in rice. *Aust. J. Agric. Res.* 42 (3), 485–496. doi: 10.1071/AR9910485

Tello-Ruiz, M. K., Jaiswal, P., and Ware, D. (2022). Gramene: A resource for comparative analysis of plants genomes and pathways. *Methods Mol. Biol.* 2443, 101–131. doi: 10.1007/978-1-0716-2067-0_5

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (Oryza sativa l.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* 11 (8), 1441–1452.

Terao, T., and Hirose, T. (2018). Temperature-dependent QTLs in indica alleles for improving grain quality in rice: increased prominence of QTLs responsible for reduced chalkiness under high-temperature conditions. *Mol. Breed.* 38 (5), 1–20. doi: 10.1007/s11032-018-0807-0

Tetlow, I. J., and Emes, M. J. (2014). A review of starch-branching enzymes and their role in amylopectin biosynthesis. *IUBMB Life* 66 (8), 546–558. doi: 10.1002/iub.1297

Tetlow, I. J., Morell, M. K., and Emes, M. J. (2004). Recent developments in understanding the regulation of starch metabolism in higher plants. *J. Exp. Bot.* 55 (406), 2131–2145. doi: 10.1093/jxb/erh248

Tian, Z., Qian, Q., Liu, Q., Yan, M., Liu, X., Yan, C., et al. (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc. Natl. Acad. Sci.* 106 (51), 21760–21765. doi: 10.1073/pnas.0912396106

Umemoto, T., and Aoki, N. (2005). Single-nucleotide polymorphisms in rice starch synthase IIa that alter starch gelatinisation and starch association of the enzyme. *Funct. Plant Biol.* 32 (9), 763–768. doi: 10.1071/FP04214

Umemoto, T., Aoki, N., Lin, H., Nakamura, Y., Inouchi, N., Sato, Y., et al. (2004). Natural variation in rice starch synthase IIa affects enzyme and starch properties. *Funct. Plant Biol.* 31 (7), 671–684. doi: 10.1071/FP04009

Umemoto, T., Nakamura, Y., and Ishikura, N. (1995). Activity of starch synthase and the amylose content in rice endosperm. *Phytochemistry* 40 (6), 1613–1616. doi: 10.1016/0031-9422(95)00380-P

Umemoto, T., Terashima, K., Nakamura, Y., and Satoh, H. (1999). Differences in amylopectin structure between two rice varieties in relation to the effects of temperature during grain-filling. *Starch-Stärke* 51 (2-3), 58–62. doi: 10.1002/(SICI)1521-379X(199903)51:2<58::AID-STAR58>3.0.CO;2-J

Umemoto, T., Yano, M., Satoh, H., Shomura, A., and Nakamura, Y. (2002). Mapping of a gene responsible for the difference in amylopectin structure between japonica-type and indica-type rice varieties. *Theor. Appl. Genet.* 104 (1), 1–8. doi: 10.1007/s001220200000

Wada, T., Miyahara, K., Sonoda, J. Y., Tsukaguchi, T., Miyazaki, M., Tsubone, M., et al. (2015). Detection of QTLs for white-back and basal-white grains caused by high temperature during ripening period in japonica rice. *Breed. Sci.* 65 (3), 216–225. doi: 10.1270/jsbbs.65.216

Wan, X. Y., Wan, J. M., Weng, J. F., Jiang, L., Bi, J. C., Wang, C. M., et al. (2005). Stability of QTLs for rice grain dimension and endosperm chalkiness characteristics across eight environments. *Theor. Appl. Genet.* 110 (7), 1334–1346. doi: 10.1007/s00122-005-1976-x

Wanchana, S., Toojinda, T., Tragoonrung, S., and Vanavichit, A. (2003). Duplicated coding sequence in the waxy allele of tropical glutinous rice (Oryza sativa l.). *Plant Sci.* 165 (6), 1193–1199. doi: 10.1016/S0168-9452(03)00326-1

Wang, C., Caragea, D., Kodadinne Narayana, N., Hein, N. T., Bheemanahalli, R., Somayanda, I. M., et al. (2022). Deep learning based high-throughput phenotyping of chalkiness in rice exposed to high night temperature. *Plant Methods* 18 (1), 1–23. doi: 10.1186/s13007-022-00839-5

Wang, L., Chen, L., Wang, P., Wang, Z., Zheng, H., Ma, H., et al. (2016). Detecting the QTL of rice chalkiness traits using advanced backcrossing population. *J. Nanjing Agric. Univ.* 39 (2), 183–190.

Wang, X., Wang, K., Yin, T., Zhao, Y., Liu, W., Shen, Y., et al. (2021). Nitrogen fertilizer regulated grain storage protein synthesis and reduced chalkiness of rice under actual field warming. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.715436

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat. Genet.* 40 (11), 1370–1374. doi: 10.1038/ng.220

Wang, Z. Y., Wu, Z. L., Xing, Y. Y., Zheng, F. G., Guo, X. L., Zhang, W. G., et al. (1990). Nucleotide sequence of rice waxy gene. *Nucleic Acids Res.* 18 (19), 5898. doi: 10.1093/nar/18.19.5898

Wang, L., Yuan, J., Ma, Y., Jiao, W., Ye, W., Yang, D. L., et al. (2018). Rice interploidy crosses disrupt epigenetic regulation, gene expression, and seed development. *Mol. Plant* 11 (2), 300–314. doi: 10.1016/j.molp.2017.12.006

Waters, D. L., Henry, R. J., Reinke, R. F., and Fitzgerald, M. A. (2006). Gelatinization temperature of rice explained by polymorphisms in starch synthase. *Plant Biotechnol. J.* 4 (1), 115–122. doi: 10.1111/j.1467-7652.2005.00162.x

Webb, B. D. (1991). "Rice quality and grades," in *Rice* (Boston, MA: Springer), 508–538.

Wei, X., Jiao, G., Lin, H., Sheng, Z., Shao, G., Xie, L., et al. (2017). GRAIN INCOMPLETE FILLING 2 regulates grain filling and starch synthesis during rice caryopsis development. *J. Integr. Plant Biol.* 59 (2), 134–153. doi: 10.1111/jipb.12510

Whitt, S. R., Wilson, L. M., Tenaillon, M. I., Gaut, B. S., and Buckler, E. S.IV (2002). Genetic diversity and selection in the maize starch pathway. *Proc. Natl. Acad. Sci.* 99 (20), 12959–12962. doi: 10.1073/pnas.202476999

Woo, M. O., Ham, T. H., Ji, H. S., Choi, M. S., Jiang, W., Chu, S. H., et al. (2008). Inactivation of the UGPase1 gene causes genic male sterility and endosperm chalkiness in rice (Oryza sativa l.). *Plant J.* 54 (2), 190–204. doi: 10.1111/j.1365-313X.2008.03405.x

Wu, Y. C., Chang, S. J., and Lur, H. S. (2016). Effects of field high temperature on grain yield and quality of a subtropical type japonica rice–Pon-Lai rice. *Plant Production Sci.* 19 (1), 145–153. doi: 10.1080/1343943X.2015.1128091

Wu, C. Y., Suzuki, A., Washida, H., and Takaiwa, F. (1998). The GCN4 motif in a rice glutelin gene is essential for endosperm-specific gene expression and is activated by opaque-2 in transgenic rice plants. *Plant J.* 14 (6), 673–683. doi: 10.1046/j.1365-313x.1998.00167.x

Xi, M., Wu, W., Xu, Y., Zhou, Y., Chen, G., Ji, Y., et al. (2021). Grain chalkiness traits is affected by the synthesis and dynamic accumulation of the storage protein in rice. *J. Sci. Food Agric.* 101 (14), 6125–6133. doi: 10.1002/jsfa.11269

Xin, W., Zhang, L., Zhang, W., Gao, J., Yi, J., Zhen, X., et al. (2019). An integrated analysis of the rice transcriptome and metabolome reveals differential regulation of carbon and nitrogen metabolism in response to nitrogen availability. *Int. J. Mol. Sci.* 20 (9), 2349. doi: 10.3390/ijms20092349

Xu, X., Ren, Y., Wang, C., Zhang, H., Wang, F., Chen, J., et al. (2019). OsVIN2 encodes a vacuolar acid invertase that affects grain size by altering sugar metabolism in rice. *Plant Cell Rep.* 38 (10), 1273–1290. doi: 10.1007/s00299-019-02443-9

Yamakawa, H., and Hakata, M. (2010). Atlas of rice grain filling-related metabolism under high temperature: joint analysis of metabolome and transcriptome demonstrated inhibition of starch accumulation and induction of amino acid accumulation. *Plant Cell Physiol.* 51 (5), 795–809. doi: 10.1093/pcp/pcq034

Yamakawa, H., Hirose, T., Kuroda, M., and Yamaguchi, T. (2007). Comprehensive expression profiling of rice grain filling-related genes under high temperature using DNA microarray. *Plant Physiol.* 144 (1), 258–277. doi: 10.1104/pp.107.098665

Yang, L., Christensen, D. A., McKinnon, J. J., Beattie, A. D., and Yu, P. (2013). Effect of altered carbohydrate traits in hulless barley (Hordeum vulgare l.) on nutrient profiles and availability and nitrogen to energy synchronization. *J. Cereal Sci.* 58 (1), 182–190.

Yang, L. M., Liu, H. L., Zhao, H. W., Wang, J. G., Sun, J., Zheng, H. L., et al. (2019). Mapping quantitative trait loci and meta-analysis for cold tolerance in rice at booting stage. *Euphytica* 215 (5), 1–15. doi: 10.1007/s10681-019-2410-9

Yu, Y., Ouyang, Y., and Yao, W. (2018). shinyCircos: an R/Shiny application for interactive creation of circos plot. *Bioinformatics* 34 (7), 1229–1231. doi: 10.1093/bioinformatics/btx763

Yuan, P. R., Kim, H. J., Chen, Q. H., Ju, H. G., Ji, S. D., and Ahn, S. N. (2010). Mapping QTLs for grain quality using an introgression line population from a cross between oryza sativa and o. rufipogon. *J. Crop Sci. Biotechnol.* 13 (4), 205–212. doi: 10.1007/s12892-010-0094-8

Yun, Y. T., Chung, C. T., Lee, Y. J., Na, H. J., Lee, J. C., Lee, S. G., et al. (2016). QTL mapping of grain quality traits using introgression lines carrying oryza rufipogon chromosome segments in japonica rice. *Rice* 9 (1), 1–14. doi: 10.1186/s12284-016-0135-0

Zakaria, S., Matsuda, T., Tajima, S., and Nitta, Y. (2002). Effect of high temperature at ripening stage on the reserve accumulation in seed in some rice cultivars. *Plant production Sci.* 5 (2), 160–168. doi: 10.1626/pps.5.160

Zhang,, Changquan,, Zhu, J., Chen, S., Fan, X., Li, Q., et al. (2019). "Wxlv, the ancestral allele of rice waxy gene.". *Mol. Plant* 12, 1157–1166. doi: 10.1016/j.molp.2019.05.011

Zhao, X., and Fitzgerald, M. (2013). Climate change: implications for the yield of edible rice. *PloS One* 8 (6), e66218. doi: 10.1371/journal.pone.0066218

Zheng, L., Zhang, W., Liu, S., Chen, L., Liu, X., Chen, X., et al. (2012). Genetic relationship between grain chalkiness, protein content, and paste viscosity properties in a backcross inbred population of rice. *J. Cereal Sci.* 56 (2), 153–160. doi: 10.1016/j.jcs.2012.05.003

Zhou, L., Chen, L., Jiang, L., Zhang, W., Liu, L., Liu, X., et al. (2009). Fine mapping of the grain chalkiness QTL qPGWC-7 in rice (Oryza sativa l.). *Theor. Appl. Genet.* 118 (3), 581–590. doi: 10.1007/s00122-008-0922-0

Zhou, W., Wang, X., Zhou, D., Ouyang, Y., and Yao, J. (2017). Overexpression of the 16-kD a α-amylase/trypsin inhibitor RAG 2 improves grain yield and quality of rice. *Plant Biotechnol. J.* 15 (5), 568–580. doi: 10.1111/pbi.12654

Zhu, A., Zhang, Y., Zhang, Z., Wang, B., Xue, P., Cao, Y., et al. (2018). Genetic dissection of qPCG1 for a quantitative trait locus for percentage of chalky grain in rice (Oryza sativa l.). *Front. Plant Sci.* 9, 1173. doi: 10.3389/fpls.2018.01173

frontiers | Frontiers in Plant Science

TYPE Erratum
PUBLISHED 17 July 2023
DOI 10.3389/fpls.2023.1254738

OPEN ACCESS

APPROVED BY
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE
Frontiers Production Office
production.office@frontiersin.org

RECEIVED 07 July 2023
ACCEPTED 07 July 2023
PUBLISHED 17 July 2023

CITATION
Frontiers Production Office (2023) Erratum: Meta-QTL and haplo-pheno analysis reveal superior haplotype combinations associated with low grain chalkiness under high temperature in rice.
Front. Plant Sci. 14:1254738.
doi: 10.3389/fpls.2023.1254738

Erratum: Meta-QTL and haplo-pheno analysis reveal superior haplotype combinations associated with low grain chalkiness under high temperature in rice

Frontiers Production Office*

Frontiers Media SA, Lausanne, Switzerland

KEYWORDS

starch metabolism, grain chalkiness, meta-QTL analysis, haplotype, haplo-pheno analysis, granule bound starch synthase I, starch synthase IIa

An Erratum on

Meta-QTL and haplopheno analysis reveal superior haplotype combinations associated with low grain chalkiness under high temperature in rice

By Kumari A, Sharma D, Sharma P, Sahil, Wang C, Verma V, Patil A, Imran M, Singh MP, Kumar K, Paritosh K, Caragea D, Kapoor S, Chandel G, Grover A, Jagadish SVK, Katiyar-Agarwal S and Agarwal M (2023). *Front. Plant Sci.* 14:1133115. doi: 10.3389/fpls.2023.1133115

Due to a production error, Figure 1 was a repeat of Figure 7 in the published article. The corrected Figures appear below.

The publisher apologizes for this mistake.

The original version of this article has been updated.

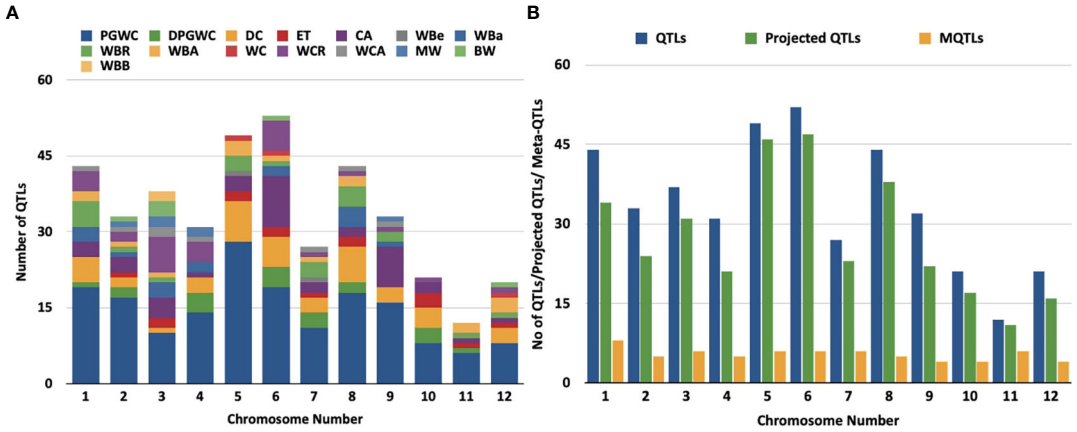


FIGURE 1
Distribution of QTLs and meta-QTLs associated with rice grain chalk on different chmmosomes of rice, **(A)** Trait-wise distribution of initial QTLs used for the meta-QTL analysis, (PGWC, Percentage Grain with Chalkiness; DPGWC, Degree of Percentage Grain with Chalkiness; DC, Degree of Endosperm Chalkiness; ET, Endosperm Transparency; CA, Chalkiness Area; WBe, White Belly; WBa, White Back; WBR,; WBA, White Back Area; WC, White core; WCR, White core rate; WCA, Whlte core area; MW,Mi1ky white; BW, Basal White; WBB, White Back and Basal). **(B)** The distribution of QTLs, projected QTLs and meta-QTLs on twelve rice chromosomes.

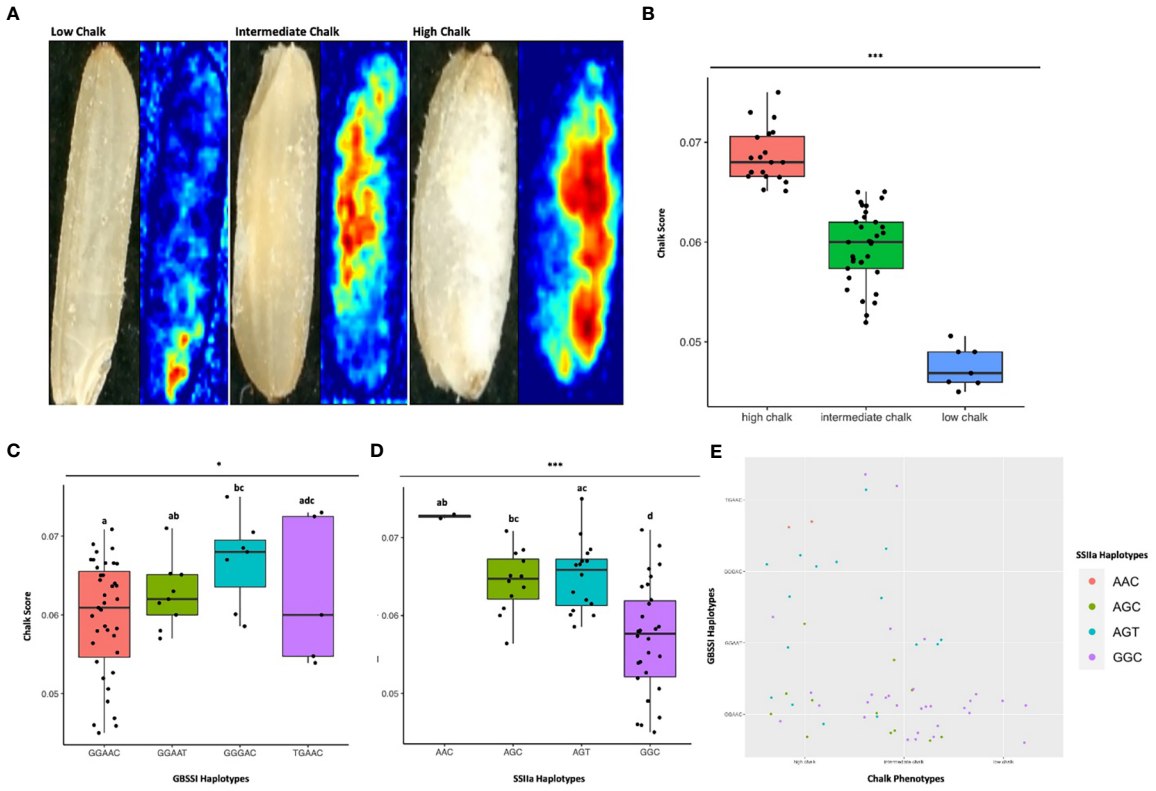


FIGURE 7
Haplo-pheno analysis of Granule Bound Starch Synthase I and Starch Synthase II a, **(A)** Rice seed scans and their corresponding heat maps highlighting the chalky area (red) in less chalk, intermediate chalk and high chalk seed types. **(B)** Box plot depicting the chalky score of 60 rice genotypes (grown in two crop seasons) categorized as high, intermediate and low chalk **(C)** Box plot of the chalk score data categorized according to the GBSS I haplotypes. **(D)** Box plot of the chalk score data categorized according to the SSIIa haplotypes. **(E)** Comparison of the haplotypic combination of GBSS I and SSIIa. One-way ANOVA was conducted for determining the statistical significance of haplotype means. Chalk score data designated with the same alphabet are not significantly different at $p \leq 0.1$(.) or $p \leq 0.05$ (*) or $p \leq 0.001$ (***) as per Turkey *Post hoc* test for Honest Significance Difference (HSD) analysis.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 13 March 2023
DOI 10.3389/fpls.2023.1138523

OPEN ACCESS

EDITED BY
Peng Wang,
Institute of Botany (CAS), China

REVIEWED BY
Ming Zheng,
Oil Crops Research Institute (CAAS), China
Kumar Paritosh,
University of Delhi, India

*CORRESPONDENCE
Jie Yang
yangjie168@aliyun.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 05 January 2023
ACCEPTED 20 February 2023
PUBLISHED 13 March 2023

CITATION
Song X, Chen Z, Du X, Li B, Fei Y, Tao Y, Wang F, Xu Y, Li W, Wang J, Liang G, Zhou Y, Tan X, Li Y and Yang J (2023) Generation of new rice germplasms with low amylose content by CRISPR/CAS9-targeted mutagenesis of the *FLOURY ENDOSPERM 2* gene.
Front. Plant Sci. 14:1138523.
doi: 10.3389/fpls.2023.1138523

Generation of new rice germplasms with low amylose content by CRISPR/CAS9-targeted mutagenesis of the *FLOURY ENDOSPERM 2* gene

Xiaohong Song[1,2†], Zhihui Chen[2,3†], Xi Du[2,3†], Bin Li[1,2], Yunyan Fei[2,3], Yajun Tao[2,3], Fangquan Wang[2,3], Yang Xu[2,3], Wenqi Li[2,3], Jun Wang[2,3], Guohua Liang[2,3], Yong Zhou[2,3], Xiaoli Tan[1], Yulong Li[1] and Jie Yang[1,2,3*]

[1]School of Life Science, Jiangsu University, Zhenjiang, Jiangsu, China, [2]Institute of Food Crops, Jiangsu Academy of Agricultural Sciences/Key Laboratory of Germplasm Innovation in Downstream of Huaihe River Ministry of Agriculture and Rural Affairs, Nanjing, Jiangsu, China, [3]Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou, China

FLOURY ENDOSPERM 2 (*FLO2*), encoding a tetratricopeptide repeat domain (TPR)-containing protein located in the nucleus, is considered to be a regulatory protein that controls the biosynthesis of seed storage substances. The diversity of *flo2* allele is attributable for the variations in grain appearance, amylose content (AC), and physicochemical properties, influencing the eating and cooking quality (ECQ) of rice. In this study, we used CRISPR/Cas9 to introduce loss-of-function mutations into the *FLOURY ENDOSPERM 2* gene in Suken118 (SK118), a widely cultivated elite *japonica* rice variety in Jiangsu, China. Physiochemical analyses of the *flo2* mutants were congruent with previous studies, exhibiting lowered AC and viscosity, risen gel consistency (GC) and gelatinization temperature (GT) values, which were all instrumental to the improvement of ECQ. However, the wrinkled opaque appearance and the decrease in grain width, grain thickness and grain weight imply trade-offs in grain yield. Despite the *ex-ante* estimation for low yielding, the superior ECQ in these novel genotypes generated by using genome editing approach may have the potential for formulating high value specialty food.

KEYWORDS

***FLOURY ENDOSPERM 2*, amylose content (AC), eating and cooking quality (ECQ), japonica rice, CRISPR/Cas9, genome editing**

Abbreviations: FLO2, *FLOURY ENDOSPERM 2*; TPR, tetratricopeptide repeat domain; AC, amylose content; ECQ, eating and cooking quality; SK118 Suken118; GC, gel consistency; GT, gelatinization temperature; Wx, waxy gene; SSII-2, soluble starch synthase ii-2; SBEIIb, starch branching enzyme IIb; BEIIb, branching enzyme IIb; ISA1, isoamylose 1; WT, wild type; Hyg, hygromycin; MNU, nitrogen methyl nitroso groupurea; EMS, methyl sulfonic acid ethyl ester; RVA, rapid visco analyser; PKV, peak paste viscosity; HPV, hot paste viscosity; CPV, cool paste viscosity; GS3, grain size 3; TGW6, thousand-grain weight 6; GS6, grain size 6.

Introduction

Rice (*Oryza sativa* L.), one of the most important crops worldwide, provides staple food for feeding half of the world's population. The steady increase in grain yield, coupled with the rapid improvements in human living conditions, has fostered ever increasing public demands for foods of high quality and high versatility (Concepcion et al., 2015; Huo et al., 2017). Consequently, the rice quality, as measured by a range of standardized parameters, including milling, appearance, nutrition and eating and cooking quality (ECQ), has been commonly used to select for and evaluate the breeding lines prior to commercial release (Zhang et al., 2018; Huang X. et al., 2021; Hu et al., 2021; Zhang H. et al., 2021). Amylose and amylopectin constitute starch that is the dominant component in rice grains and accounts for about 90% of the weight of the rice endosperm (Zhou et al., 2002; Wang et al., 2018). As the most important quality trait, ECQ is affected by multiple factors such as amylose content (AC) and an eclectic list of physicochemical properties of endosperm (Lu et al., 2013; Chun et al., 2015; Van Hung et al., 2016). Despite the recent increase in the awareness of the health benefits of high AC rice, it appears that a seemingly polarized market has emerged, especially for the formulation of specialty food, where the low AC trait is desired as it could well be attributed for the favorable food texture and taste (Sagare et al., 2020; Hu et al., 2021). Further, some rice cultivars, such as Nanjing 46 and Nanjing 9108, with relatively high ECQ and low AC (10-15%) are popular with consumers in China and the broad East Asian region, which are soft yet not sticky after cooking, and are defined as soft japonica rice (Li et al., 2018; Hu et al., 2021; Zhang C. et al., 2021). Therefore, there is impetus to develop rice genotypes with low AC to meet market requirements.

FLOURY ENDOSPERM 2 (*FLO2*), a member of a conserved gene family in plants, encodes a tetratricopeptide repeat domain (TPR)-containing protein located in the nucleus (She et al., 2010; Wu et al., 2015; Kolli et al., 2019; Suzuki et al., 2020). It has been reported that *FLOURY ENDOSPERM 2* could not only modulate the expression of starch synthesis-related genes including *OsAGPL2, OsAGPS2b, OsGBSSI, OsBEI, OsBEIIb, OsISA1* and *OsPUL*, but also directly impact on the amylose biosynthesis by manipulating the splicing efficiency of *Wx* pre-mRNA (Wu et al., 2015; Cai et al., 2022; Feng et al., 2022; Sreenivasulu et al., 2022). Previous studies have shown that *flo2*, the loss-of-function mutant of *FLOURY ENDOSPERM 2*, produced two distinct phenotypes, one with white and floury endosperm, and the other with dull grain (She et al., 2010; Hunt et al., 2013; Boehlein et al., 2018). In the former, starch granules were small, spherical and loosely packed with large air space (Malinova et al., 2017; Jiang et al., 2018; Suzuki et al., 2020). The *FLOURY ENDOSPERM 2* also affected grain size as demonstrated by both overexpression and null mutation approaches (She et al., 2010; Zhang et al., 2013). In addition, *flo2* mutant was also endowed with alterations in grain physicochemical properties including grain breakdown, setback and consistency (Wu et al., 2015; Zhang et al., 2017). Therefore, *FLOURY ENDOSPERM 2* plays an important role in the determination of grain quality by regulating the accumulation of storage substances in the endosperm. Previous methods to generate *flo2* mutants by chemical mutagens such as EMS have the disadvantages of safety risk, low efficiency and uncontrollable direction (Wu et al., 2015; Kihira et al., 2017; Wu et al., 2019).

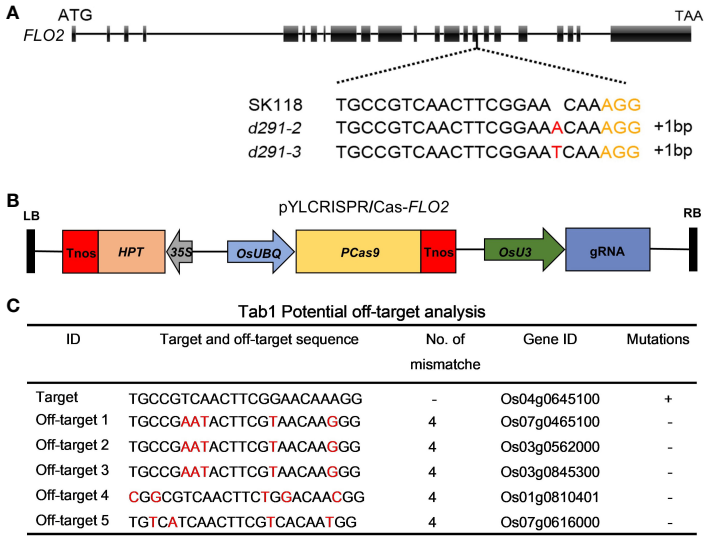
In the last few years, as one of the most advanced systems for genome editing, CRISPR/Cas has been commonly used to make precise and predictable genome modifications in crop plants to obtain desired traits (Zhu et al., 2020; Gao, 2021). This is especially true for rice, where impressive progress has been made and a plethora of traits have been modified by applying the CRISPR/Cas9 editing system, by virtue of rice's economic significance and the ease of transgenic regeneration (Zhou et al., 2014; Ma et al., 2021). For grain quality improvement, *Wx* is a key gene that encodes granule binding starch synthase I (GBSSI) involved in amylose biosynthesis. It has been genetically edited to manipulate AC and/or improve ECQ in a number of recent studies in rice (V.M. Butardo et al., 2017; Bello et al., 2019; Fei et al., 2019; Huang et al., 2020; Zeng et al., 2020; Huang L. et al., 2021; Xu et al., 2021). Likewise, a number of other key genes involved in amylopectin fine structure, such as *SSII-2, SBEIIb, BEIIb* and *ISA1*, have been edited for ECQ improvements (Crofts et al., 2015; Sun et al., 2017; Chao et al., 2019; Tappiban et al., 2022).

In this study, to induce AC and improve ECQ of a widely cultivated elite japonica variety, Suken118 (SK118), two loss-of-function mutants of *FLOURY ENDOSPERM 2* gene were generated by CRISPR/Cas9 technology. Our results show that knockout of *FLOURY ENDOSPERM 2* gene can significantly reduce the content of straight chain powder. The physicochemical properties of *flo2* mutants showed that lower AC and viscosity, higher gelation degree and gelatinization temperature were helpful to improve the edible and cooking quality of rice. Therefore, *FLOURY ENDOSPERM 2* has potential application value in green and healthy rice breeding.

Results

Generation and identification of *flo2* mutants in rice

To generate *flo2* mutants with the expectation to produce null mutations, sgRNA was designed in the coding region of *FLOURY ENDOSPERM 2*. The constructed expression vector *flo2* gRNA was expressed and driven by the *OsU3* promoter, and the Cas9 cassette was driven by the ubiquitin promoter (Figure 1A). We took advantage of these CRISPR/Cas expression vectors to transform rice variety Sk118 by Agrobacterium tumefaciens infecting rice calli (Figure 1B). Positive T0 transgenic plants were identified by PCR amplification of a fragment of the Hyg gene that was used as the selection marker. The target genomic region of *FLOURY ENDOSPERM 2* was amplified by a pair of primers (*FLO2*-TF/TR; Table S1) flanking the target site and sequenced. In addition, we failed to find any mutations in any of the potential off-target sites (Figure 1C). The sequencing results of T0 plants showed that the mutation rate was as high as 74.19%, among which the percentages of mutation by inserting the base "T" and "A" were the highest, being 30.43% and 19.57%, respectively. Two homozygous T1



ID	Target and off-target sequence	No. of mismatche	Gene ID	Mutations
Target	TGCCGTCAACTTCGGAACAAAGG	-	Os04g0645100	+
Off-target 1	TGCCGAATACTTCGTAACAAGGG	4	Os07g0465100	-
Off-target 2	TGCCGAATACTTCGTAACAAGGG	4	Os03g0562000	-
Off-target 3	TGCCGAATACTTCGTAACAAGGG	4	Os03g0845300	-
Off-target 4	CGGCGTCAACTTCTGGACAACGG	4	Os01g0810401	-
Off-target 5	TGTCATCAACTTCGTCACAATGG	4	Os07g0616000	-

FIGURE 1
CRISPR/Cas9-induced mutations in the *FLOURY ENDOSPERM 2* genes. **(A)** Schematic of the *FLOURY ENDOSPERM 2* gene structures and target sites. Exons and introns are indicated with black rectangles and black lines, respectively. The spacer and PAM sequences were marked in black and orange. **(B)** The expression vector of pYLCRISPR/Cas9-*Flo2*. The expression of Cas9 is driven by the maize ubiquitin promoter (*OsUBQ*); the expression of the sgRNA scaffold is driven by the rice *OsU3* small nuclear RNA promoter; the expression of hygromycin (*HPT*) is driven by CaMV35S promoter. Tnos, the terminator; LB and RB, left border and right border, respectively. **(C)** Potential miss analysis. Red fonts indicate different bases.

mutant lines free of T-DNAs, named *d29-2* and *d29-3*, were selected for further research, featuring the insertion of a single base of "T" and "A", respectively, resulting in frame shift and premature termination of translation (Figure 2C).

Knockout of *FLOURY ENDOSPERM 2* gene has multiple effects on agronomic traits

All the plants were grown in the experimental field in Jiangsu Academy of Agricultural Sciences, Nanjing, China, during normal rice-growing seasons. The edited plants derived from *d29-2* and *d29-3* did not exhibit discernible variations from wild type (WT) control in grain length, plant height, the panicle length, the number of grains per panicle and tiller numbers per plant (Figure 2A, Figures 3A, B, E–H). In contrast, the grain width and grain thickness of the *flo2* mutants were significantly lesser than WT control, which is congruent with the previous studies on chemically induced *flo2* mutants (She et al., 2010); (Figures 3C, D). Conceivably, the 1000-grain weights of *d29-2* and *d29-3* were reduced by 26.50% and 20.90%, respectively (Figure 3I).

Rice yield and grain appearance quality are complex quantitative characters, which are influenced by genetic background and environmental factors. Although *FLOURY ENDOSPERM 2* gene knockout mutants have multiple negative effects on grain appearance and yield, rice with low amylose content and high cooking quality has a huge market demand. In future studies, *FLOURY ENDOSPERM 2* gene and yield trait genes can be combined, and the mutual balance between yield and trait can be realized by using the interaction effect between different genes.

Improvement of ECQ and taste value of *flo2* transgenic rice

To further assess the effects of *flo2* mutants on rice ECQ, we performed a number of assays on the grain. The gel consistency (GC) in *d29-2* and *d29-3* increased significantly relative to WT by 11.6 mm and 18.0 mm, respectively (Figure 4A). The gelatinization temperature (GT) of the mutant lines was about 22°C higher than WT (Figure 4D). The increases in GC and GT in *flo2* mutant relative to WT are favorable for improving ECQ in rice (Zhang et al., 2020; Sreenivasulu et al., 2022). The RVA pasting profiles of the *flo2* mutants were also distinct from WT, specifically, the peak paste viscosity (PKV), hot paste viscosity (HPV) and cold paste viscosity (CPV) of the mutants were significantly lower than those of WT (Figure 4B). The protein content (PC) was only increased of 1.68% and 1.64% in *d29-2* and *d29-3* mutants, respectively (Figure 4C), which had no significant effect on ECQ (Sreenivasulu et al., 2022). Generally, the increase of gel consistency is conducive to improving the taste quality of rice. Taken together, these results manifest the complex changes in the physicochemical properties of cooked rice grains in the *flo2* mutants, which are generally conducive to ECQ improvements.

Rice starch with a higher amylose content always shows faster retrogradation, leading to higher setback viscosity (SBV) of the

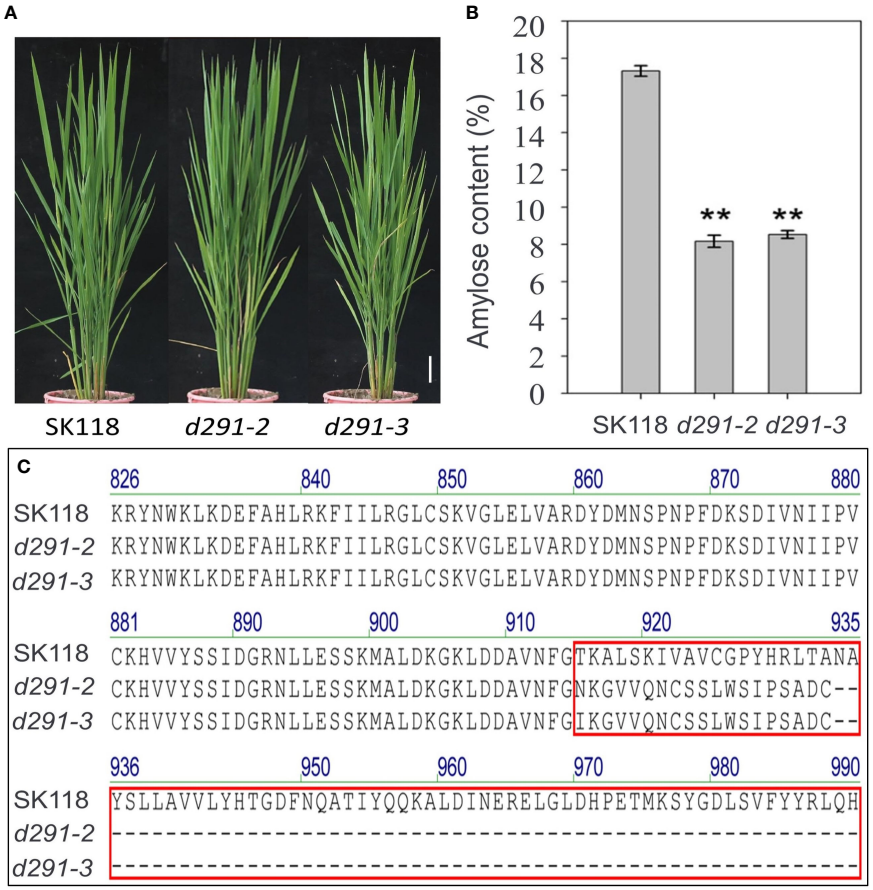


FIGURE 2
Phenotype and amino acid variations assays. **(A)** Phenotypes of *flo2* mutant lines (*d291-2* and *d291-3*) in SK118 background. Bar = 5 cm. **(B)** Amylose content (%). Data are given as the mean ± S.D. ($n \geq 3$). Student's t test: *P < 0.05; **P < 0.01; ns, no significance. **(C)** Amino acid variations of the FLO2 protein in the *flo2* mutant. Wild-type, SK118. Homozygous mutations identified at the target sites of *d291-2* and *d291-3* mutant lines in the T1 generation. Amino acid variations were indicated in red box, with the number representing the order in the proteins. The ellipsis dots represent premature stop codons.

starch, while starch with a lower amylose content has a higher relative crystallinity and short-range molecular order, resulting in higher swelling power, PKV and breakdown viscosity (BDV) (Wang et al., 2017; Cai et al., 2022). Previous studies showed that AC was positively correlated with PKV, HPV, CPV in a significant manner, whereas it was negatively correlated with adhesiveness, cohesiveness, and GT (Kong et al., 2015). Our results in this study on *flo2* lines that were endowed with low AC was concurrent with the reductions in these RVA pasting values, corroborating the previous reports. GC is an important indicator for distinguishing rice grains that do not undergo retrogradation (i.e., rice remains soft after cooking and cooling) (Zhang et al., 2020; Sreenivasulu et al., 2022). GT is also a crucial factor that affects rice ECQ. It is generally recognized that the rice grains with high GT would need more time to cook, and the texture of the cooked rice tends to be less sticky, especially when the cooked rice is cooled (Li et al., 2021; Sreenivasulu et al., 2022). The enhancements in GC and GT values in the *flo2* mutants, as unraveled in this study, are clearly favorable for ECQ amelioration.

Opaque appearance phenotype of *flo2* mutants rice grain

The polished rice grains derived from *d29-2* and *d29-3* were featured with distinctive opaque and shriveled appearance (Figure 5A). There might be some phenotypic variations in grain appearance between the *FLO*-edited lines and the *flo2* mutants that were reported in previous studies, which could well be attributed to the different genetic backgrounds and cultivation practices used between different studies (She et al., 2010; Wu et al., 2015). The images of transverse sections showed that the endosperm of the *flo2* mutants were more opaque and powdery compared to WT at the center of the endosperm, but such differences were diminished away from the center region (Figure 5B).

We speculate that this is mainly due to the fact that there are many gaps between the poorly filled starch bodies in the endosperm. When the light shines on the endosperm in the internal area, the light diffuses between the gaps, causing the endosperm to appear powdery. When the inner amyloplasts are

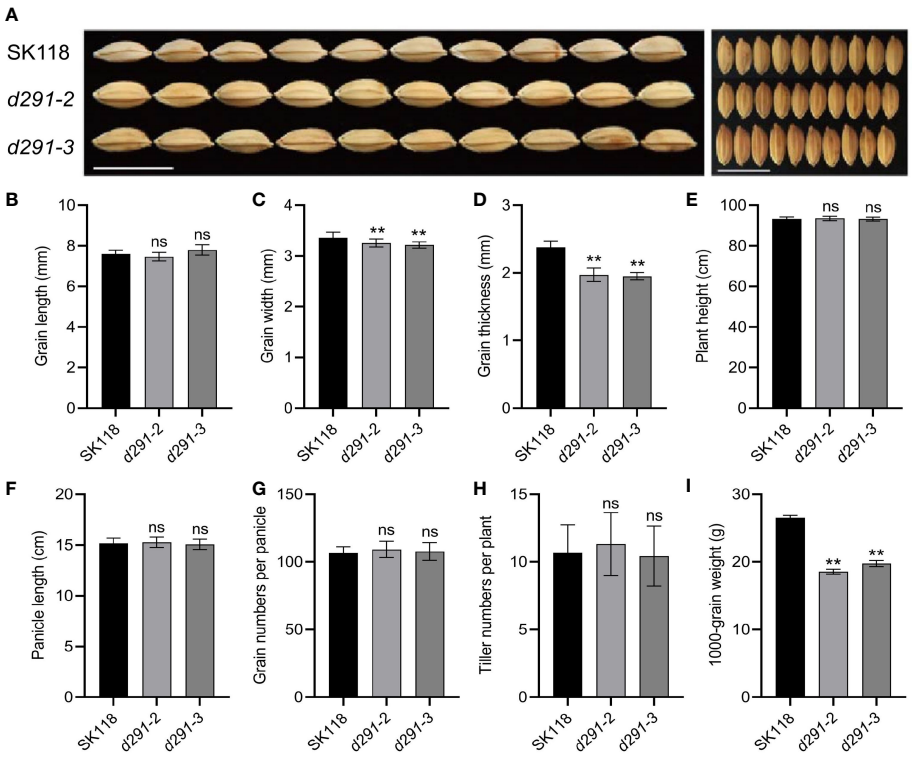


FIGURE 3
Knockout of *FLOURY ENDOSPERM 2* gene has multiple effects on agronomic traits. **(A)** Grain phenotypes of *flo2* mutant lines (*d291-2* and *d291-3*) in SK118 background. Bar = 1 cm. **(B)** grain length (mm). **(C)** grain width (mm). **(D)** grain thickness (mm). **(E)** plant height(cm). **(F)** panicle length (cm). **(G)** grain numbers per panicle. **(H)** tiller numbers per plant. **(I)** 1000-grain weight **(G)**. Data are given as the mean ± S.D. (n ≥ 20). Student's t test: **P < 0.01; ns, not significance.

closely arranged and there is no gap, the light can directly pass through and present transparent endosperm. Such spatial variation in endosperm between *flo2* and WT is intriguing and warrants further investigation.

Genome editing of *FLOURY ENDOSPERM 2* decreases amylose content significantly

Amylose has high binding capacity with iodine. The higher the AC is, the deeper the staining in the starch-iodine reaction becomes (Agasimani et al., 2013). As expected, the iodine staining of the endosperms in the transverse sections of *flo2* seeds were overtly lighter than WT, indicative of AC reduction in the mutants (Figure 5C). The measurement of AC by using AA3 continuous flow analyzer showed that the AC of *d29-2* and *d29-3* were 8.20% and 8.56%, respectively, which are in sharp contrast to that of WT (17.35%) (Figure 2B). Rather, AC in the *flo2* mutants resemble that of a typical soft rice (Wu et al., 2015). Starch is the first kind of storage material in rice. The reduction of 1000 grain weight of *flo2* mutants is mainly caused by the reduction of total starch content in endosperm of seeds. Furthermore, we detected the transcript levels of *FLOURY ENDOSPERM 2* in the WT (SK118), *d291-2* and *d291-3* plants. The transcript levels of *FLOURY ENDOSPERM 2* were decreased in the *d291-2* and *d291-3* plants, consistent with the previous study (Figure S1) (She et al., 2010).

Morphology of starch granules in endosperm cells of *flo2* mutants

To explore the causes of endosperm changes, we observed the endosperm of wild type seeds and two mutant seeds with scanning electron microscopy. Further, scanning electron microscopy images showed that the starch grains of the *flo2* mutants were irregularly spherical and loosely arranged, contrary to the regular spherical shape and uniformity in size, and tight arrangement (Figure 6A). The loose arrangement of *flo2* mutants starch granules is the main reason for its opaque appearance.

We speculated that the starch accumulation in *flo2* mutants was poor, which led to the loose endosperm structure. During the ripening process of rice, the powder region of endosperm is easy to be broken under external pressure such as water loss, forming sub grains. A large number of cavities were left between loose starch grains, which led to diffuse reflection of light to form a silty, opaque endosperm phenotype.

Discussion

In summary, in this study we have generated low AC/high ECQ rice genotypes through targeted mutagenesis of *FLOURY ENDOSPERM 2* by using CRISPR/Cas9 approach and demonstrated that editing of *FLOURY ENDOSPERM 2* could be an

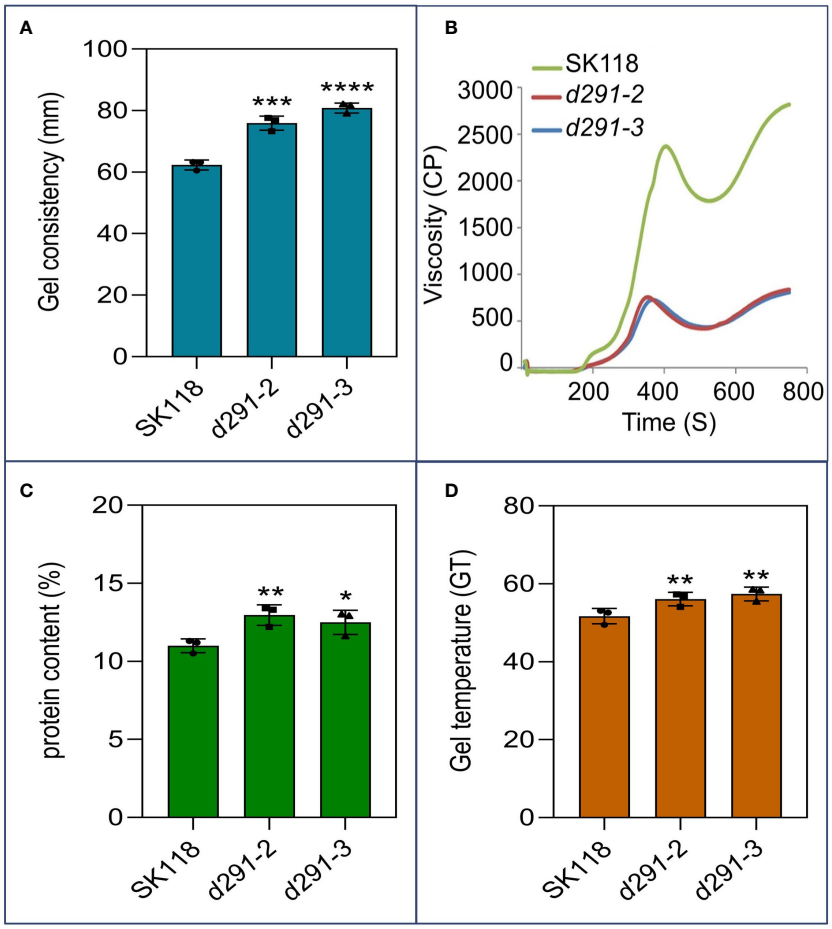


FIGURE 4
Improvement of ECQ and taste value of *flo2* transgenic rice. **(A)** Gel consistency (mm). **(B)** RVA spectra of rice flours. **(C)** Total protein content of the rice flours (%). **(D)** Gelatinization temperature (GT) of rice flours from SK118, *d291-2* and *d291-3*. Data are given as the mean ± S.D. (n ≥ 3). Student's t test: *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

effective route to reduce AC and accelerate rice breeding to achieve desired high ECQ trait. We have also analyzed the physicochemical properties of the *flo2* mutants, which have not only verified the improvements in rice ECQ, but also provided useful data for industrial evaluation of these potentially useful germplasm. Undesired, the mutations of *FLOURY ENDOSPERM 2* have led to a significant reduction in seed weight, which may imply a possible reduction in grain yield. This necessitates further evaluation for their agronomic performance, especially in the field conditions. Nevertheless, there might be a net gain in its output value given that the high value niche market for high-ECQ grains is rapidly expanding. Future studies could also be directed to the incorporation of high yielding traits, for example, by simultaneous targeting a number of negative regulatory genes, such as *GS3* (Zeng et al., 2019), *TGW6* (Hang et al., 2018; Zeng et al., 2019) and *GS6* (Sun et al., 2013), leading to an optimal balance between quality and yield traits.

Materials and methods

Plant materials and growth conditions

Suken118 (SK118) is an elite japonica rice variety with good resistance and quality, which is cultivated widely in the central of Jiangsu Province, China (http://ricedata.cn). The 20bp long target sites were chosen in the *FLOURY ENDOSPERM 2* codon region. The CRISPR/Cas9 vector was constructed as previously described then the cassette was transferred into SK118 callus by Agrobacterium tumefaciens-mediated transformation using the strain EHA105 (Hiei et al., 1997; Ma et al., 2015). Transgenic rice lines were grown in greenhouse and test fields in Jiangsu Academy of Agricultural Sciences in Nanjing, China, during normal rice-growing seasons.

10.3389/fpls.2023.1138523

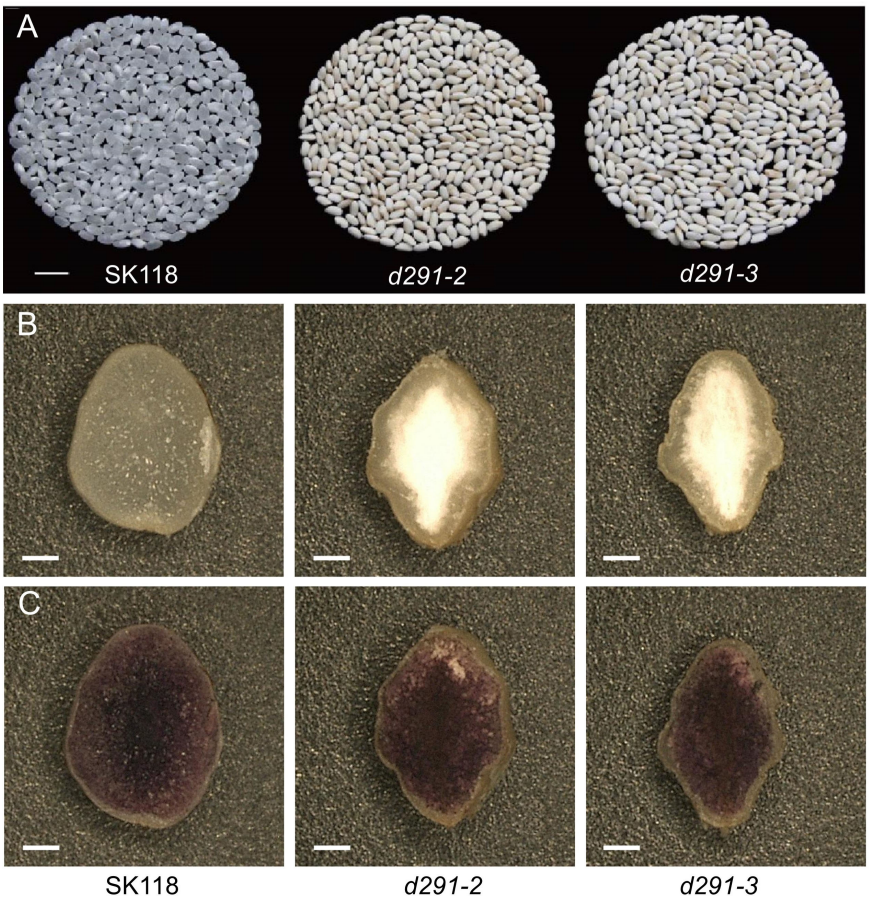


FIGURE 5
Appearance phenotype analysis of *flo2* mutants rice grain. **(A)** Appearance of polished rice from *flo2* mutant lines (*d291-2* and *d291-3*) in SK118 background. Bar = 1 cm. **(B)** Transverse sections of rice grains. Bar = 300 μm. **(C)** Rice grains stained with KI-I2. Bar = 300 μm.

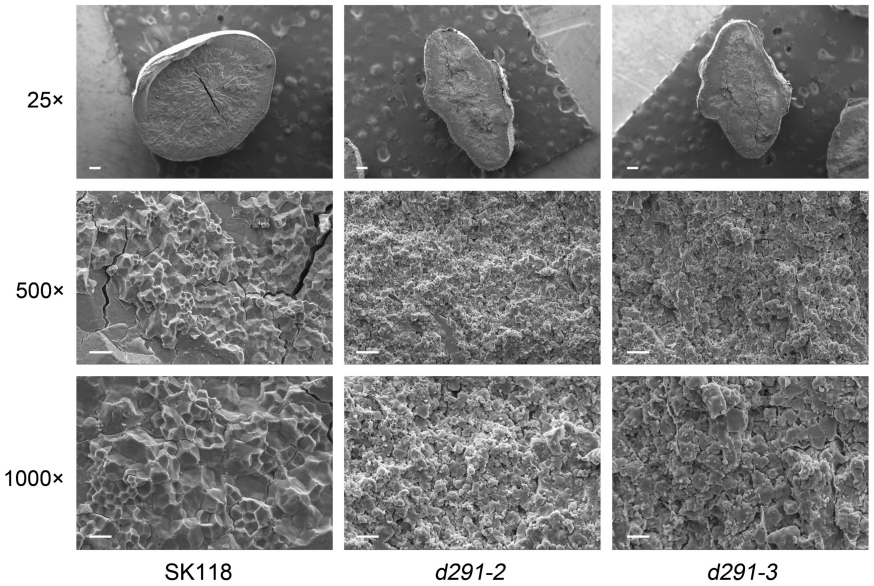


FIGURE 6
Morphology of starch granules in endosperm cells of *flo2* mutants. **(A)** Scanning electron microscopy images of the transverse section of *flo2* mutants and wild-type. 25×, Bars = 200μm. 500×, Bars = 20 μm. 1000×, Bars = 10 μm.

Genotype assays and phenotype

Positive T0 transgenic plants were identified by PCR amplification of a fragment of the Hyg gene that was used as a selection marker. The target genomic region of *FLOURY ENDOSPERM 2* was amplified by a pair of primers (*FLO2*-TF/TR; Table S1) flanking the target site and sequenced. Seeds were collected for each plant from WT and homozygous mutant lines. Grain length, plant height, the panicle length, the number of grains per panicle and tiller numbers per plant were measured in the fields. The 1,000 grains weight was measured after the oven at 45°C until constant weight. The grain length, width, and thickness were measured using a vernier caliper.

Grain phenotype and iodine staining of endosperm

The hulls of rice seeds were removed to observe the external appearance of the grain in both WT and mutant lines. Grains were cut through the center to expose the endosperm; 1μL 0.1% iodine-potassium iodide solution was dropped on the endosperm surface and photographs were taken after 3-5 min by VHX-500FE.

RNA isolation and qPCR

Total RNAs were extracted from the rice endosperm of rice using a TRIzol kit according to the user's manual (Invitrogen) for expression analysis. And then reverse-transcribed into cDNA from 1 μg of total RNA following the manufacturer's instructions (Fastking RT Kit; TIANGEN). qRT-PCR analysis was performed using a Light Cycler 480 system (Roche) and SYBR Green Real-time PCR Master Mix. The relative expression levels of *FLOURY ENDOSPERM 2* were normalized to the rice *UBIQUITIN* (Os03g0234200) gene, which was used as an internal control. All date were measured by three individual replicates. The transcript levels were calculated by the $2^{-\Delta\Delta CT}$ method. The primer sequences are listed in Table S1.

Determination of AC and total protein content

Amylose content (AC) was measured using an AA3 continuous flow analyzer set. Flour from WT, mutant lines and four rice samples with known apparent amylose content was taken 0.05 g to determine. AA3 continuous flow analyzer set was also used to determine total protein content by measuring 0.2 g of flour from WT and mutant lines. The conversion coefficient is 5.95. All samples were repeated three times.

Evaluation of grain GC, GT and viscosity property

Gel consistency (GC) was measured following the procedure described in GB/T 22294-2008/ISO 11747:2012. Gelatinization temperature (GT) was indirectly estimated *via* the alkali digestion test. Rice viscosity properties were determined using a Rapid Visco Analyser (RVA-TecMaster, Newport Scientific, Warriewood, Australia). Rice flour (3g, 12% m.b.) was mixed with 25 g of double-deionized water in the RVA sample can. Thrice measurements were performed for each sample.

Scanning electron microscopy of seed cross-section

After drying, the mature seeds with shelling were frozen in liquid nitrogen for 5S then the cross-section of the samples was manually snapped and sputter-coated with gold palladium on copper studs. Magnifications of about 25×, 500×, and 1000× were used to observe endosperm cross-section and starch granule morphology.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

JY, GL and YF designed and supervised the research. XS, ZC, XD, FW and JW performed most experiments. YX and WL analyzed date. XS, BL, YT and YL wrote the paper. X-LT and YZ provided resources. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by Project of Seeds innovation in Jiangsu (GBGS [2021]001), Natural Science Foundation of Jiangsu Province (BK20190255) and the Exploratory Project of the Jiangsu Academy of Agricultural Sciences (ZX (21)1201).

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/fpls.2023.1138523/full#supplementary-material

References

Agasimani, S., Selvakumar, G., Joel, A. J., and Ganesh Ram, S. (2013). A simple and rapid single kernel screening method to estimate amylose content in rice grains. *Phytochem. Anal.* 24 (6), 569–573. doi: 10.1002/pca.2433

Bello, B. K., Hou, Y., Zhao, J., Jiao, G., Wu, Y., Li, Z., et al. (2019). *NF-YB1-YC12-bHLH144* complex directly activates *Wx* to regulate grain quality in rice (*Oryza sativa l.*). *Plant Biotechnol. J.* 17 (7), 1222–1235. doi: 10.1111/pbi.13048

Boehlein, S. K., Shaw, J. R., Boehlein, T. J., Boehlein, E. C., and Hannah, L. C. (2018). Fundamental differences in starch synthesis in the maize leaf, embryo, ovary and endosperm. *Plant J.* 96 (3), 595–606. doi: 10.1111/tpj.14053

Cai, Y., Zhang, W., Fu, Y., Shan, Z., Xu, J., Wang, P., et al. (2022). Du13 encodes a C2 H2 zinc-finger protein that regulates *Wx*(b) pre-mRNA splicing and microRNA biogenesis in rice endosperm. *Plant Biotechnol. J.* 20 (7), 1387–1401. doi: 10.1111/pbi.13821

Chao, S., Yicong, C., Baobing, F., Guiai, J., Zhonghua, S., Ju, L., et al. (2019). Editing of rice isoamylase gene *ISA1* provides insights into its function in starch formation. *Rice Sci.* 26 (2), 77–87. doi: 10.1016/j.rsci.2018.07.001

Chun, A., Lee, H. J., Hamaker, B. R., and Janaswamy., S. (2015). Effects of ripening temperature on starch structure and gelatinization, pasting, and cooking properties in rice (*Oryza sativa*). *J. Agric. Food Chem.* 63 (12), 3085–3093. doi: 10.1021/jf504870p

Concepcion, J., Ouk, M., Zhao, D., and Fitzgerald, M. A. (2015). The need for new tools and investment to improve the accuracy of selecting for grain quality in rice. *Field Crops Res.* 182, 60–67. doi: 10.1016/j.fcr.2015.05.003

Crofts, N., Abe, N., Oitome, N. F., Matsushima, R., Hayashi, M., Tetlow, I. J., et al. (2015). N. fujita. amylopectin biosynthetic enzymes from developing rice seed form enzymatically active protein complexes. *J. Exp. Bot.* 66 (15), 4469–4482. doi: 10.1093/jxb/erv212

Fei, Y., Jie, Y., Fangquan, W., Fangjun, F., Wenqi, L., Jun, W., et al. (2019). Production of two elite glutinous rice varieties by editing *Wx* gene. *Rice Sci.* 26 (2), 118–124. doi: 10.1016/j.rsci.2018.04.007

Feng, T., Wang, L., Li, L., Liu, Y., Chong, K., Theissen, G., et al. (2022). *OsMADS14* and *NF-YB1* cooperate in the direct activation of *OsAGPL2* and *Waxy* during starch synthesis in rice endosperm. *New Phytol.* 234 (1), 77–92. doi: 10.1111/nph.17990

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell* 184 (6), 1621–1635. doi: 10.1016/j.cell.2021.01.005

Han, Y., Luo, D., Usman, B., Nawaz, G., Zhao, N., Liu, F., et al. (2018). Development of high yielding glutinous cytoplasmic Male sterile rice (*Oryza sativa l.*) lines through CRISPR/Cas9 based mutagenesis of *Wx* and *TGW6* and proteomic analysis of anther. *Agronomy* 8, 290–311. doi: 10.3390/agronomy8120290

Hiei, Y., Komari, T., and Kubo, T. (1997). Transformation of rice mediated by agrobacterium tumefaciens. *Plant Mol. Biol.* 35 (1), 205–218. doi: 10.1023/A:1005847615493

Hu, Y., Cong, S., and Zhang, H. (2021). Comparison of the grain quality and starch physicochemical properties between japonica rice cultivars with different contents of amylose, as affected by nitrogen fertilization. *Agriculture* 11 (7), 616–630. doi: 10.3390/agriculture11070616

Huang, L., Gu, Z., Chen, Z., Yu, J., Chu, R., Tan, H., et al. (2021). Improving rice eating and cooking quality by coordinated expression of the major starch synthesis-related genes, *SSII* and *Wx*, in endosperm. *Plant Mol. Biol.* 106 (4-5), 419–432. doi: 10.1007/s11103-021-01162-8

Huang, L., Li, Q., Zhang, C., Chu, R., Gu, Z., Tan, H., et al. (2020). Creating novel *Wx* alleles with fine-tuned amylose levels and improved grain quality in rice by promoter editing using CRISPR/Cas9 system. *Plant Biotechnol. J.* 18 (11), 2164–2166. doi: 10.1111/pbi.13391

Huang, X., Su, F., Huang, S., Mei, F., Niu, X., Ma, C., et al. (2021). Novel *Wx* alleles generated by base editing for improvement of rice grain quality. *J. Integr. Plant Biol.* 63 (9), 1632–1638. doi: 10.1111/jipb.13098

Hunt, H. V., Moots, H. M., Graybosch, R. A., Jones, H., Parker, M., Romanova, O., et al. (2013). *Waxy* phenotype evolution in the allotetraploid cereal broomcorn millet: Mutations at the GBSSI locus in their functional and phylogenetic context. *Mol. Biol. Evol.* 30 (1), 109–122. doi: 10.1093/molbev/mss209

Huo, X., Wu, S., Zhu, Z., Liu, F., Fu, Y., Cai, H., et al. (2017). *NOG1* increases grain production in rice. *Nat. Commun.* 8 (1), 1497. doi: 10.1038/s41467-017-01501-8

Jiang, J. Z., Kuo, C. H., Chen, B. H., Chen, M. K., Lin, C. S., and Ho, S. L. (2018). Effects of *OsCDPK1* on the structure and physicochemical properties of starch in developing rice seeds. *Int. J. Mol. Sci.* 19 (10), 3247–3262. doi: 10.3390/ijms19103247

Kihira, M., Taniguchi, K., Kaneko, C., Ishii, Y., Aoki, H., Koyanagi, A., et al. (2017). Arabidopsis thaliana *FLO2* is involved in efficiency of photoassimilate translocation, which is associated with leaf growth and aging, yield of seeds and seed quality. *Plant Cell Physiol.* 58 (3), 440–450. doi: 10.1093/pcp/pcw217

Kolli, R., Soll, J., and Carrie, C. (2019). *OXA2b* is crucial for proper membrane insertion of *COX2* during biogenesis of complex IV in plant mitochondria. *Plant Physiol.* 179 (2), 601–615. doi: 10.1104/pp.18.01286

Kong, X., Zhu, P., Sui, Z., and Bao., J. (2015). Physicochemical properties of starches from diverse rice cultivars varying in apparent amylose content and gelatinisation temperature combinations. *Food Chem.* 172, 433–440. doi: 10.1016/j.foodchem.2014.09.085

Li, Q. F., Huang, L. C., Chu, R., Li, J., Jiang, M. Y., Zhang, C. Q., et al. (2018). Down-regulation of *SSSII-2* gene expression results in novel low-amylose rice with soft, transparent grains. *J. Agric. Food Chem.* 66 (37), 9750–9760. doi: 10.1021/acs.jafc.8b02913

Li, H., Yan, S., Yang, L., Xu, M., Ji, J., Mao, H., et al. (2021). Starch gelatinization in the surface layer of rice grains is crucial in reducing the stickiness of parboiled rice. *Food Chem.* 341, 128202–128221. doi: 10.1016/j.foodchem.2020.128202

Lu, S., Cik, T.-T., Lii, C.-y., Lai, P., and Chen., H.-H. (2013). Effect of amylose content on structure, texture and α-amylase reactivity of cooked rice. *LWT - Food Sci. Technol.* 54 (1), 224–228. doi: 10.1016/j.lwt.2013.05.028

Ma, Z., Wei, M., Zhang, Y., Qin, G., Liu, C., Li, Z., et al. (2021). Development of CRISPR_Cas9 genome editing system and its application in rice molecular breeding. *In Vitro Cell. Dev. Biol. - Plant* 57 (4), 700–708. doi: 10.1007/s11627-021-10203-2

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* 8 (8), 1274–1284. doi: 10.1016/j.molp.2015.04.007

Malinova, I., Alseekh, S., Feil, R., Fernie, A. R., Baumann, O., Schottler, M. A., et al. (2017). Starch synthase 4 and plastidal phosphorylase differentially affect starch granule number and morphology. *Plant Physiol.* 174 (1), 73–85. doi: 10.1104/pp.16.01859

Sagare, D. B., Abbai, R., Jain, A., Jayadevappa, P. K., Dixit, S., Singh, A. K., et al. (2020). More and more of less and less: Is genomics-based breeding of dry direct-seeded rice (DDSR) varieties the need of hour? *Plant Biotechnol. J.* 18 (11), 2173–2186. doi: 10.1111/pbi.13454

She, K. C., Kusano, H., Koizumi, K., Yamakawa, H., Hakata, M., Imamura, T., et al. (2010). A novel factor *FLOURY ENDOSPERM2* is involved in regulation of rice grain size and starch quality. *Plant Cell* 22 (10), 3280–3294. doi: 10.1105/tpc.109.070821

Sreenivasulu, N., Zhang, C., Tiozon, R. N.Jr., and Liu, Q. (2022). Post-genomics revolution in the design of premium quality rice in a high-yielding background to meet consumer demands in the 21st century. *Plant Commun.* 3 (3), 100271. doi: 10.1016/j.xplc.2021.100271

Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., et al. (2017). Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* 8, 298. doi: 10.3389/fpls.2017.00298

Sun, L., Li, X., Fu, Y., Zhu, Z., Tan, L., Liu, F., et al. (2013). *GS6*, a member of the GRAS gene family, negatively regulates grain size in rice. *J. Integr. Plant Biol.* 55 (10), 938–949. doi: 10.1111/jipb.12062

Suzuki, R., Imamura, T., Nonaga, Y., Kusano, H., Teramura, H., Sekine, K. T., et al. (2020). H. shimada. A novel *FLOURY ENDOSPERM2* (*FLO2*)-interacting protein, is involved in maintaining fertility and seed quality in rice. *Plant Biotechnol. (Tokyo)* 37 (1), 47–55. doi: 10.5511/plantbiotechnology.19.1212b

Tappiban, P., Hu, Y., Deng, J., Zhao, J., Ying, Y., Zhang, Z., et al. (2022). Relative importance of branching enzyme isoforms in determining starch fine structure and physicochemical properties of indica rice. *Plant Mol. Biol.* 108 (4-5), 399–412. doi: 10.1007/s11103-021-01207-y

Van Hung, P., Chau, H. T., and Phi, N. T. (2016). *In vitro* digestibility and *in vivo* glucose response of native and physically modified rice starches varying amylose contents. *Food Chem.* 191, 74–80. doi: 10.1016/j.foodchem.2015.02.118

V.M. Butardo,, Anacleto, R., Parween, S., Samson, I., de Guzman, K., Alhambra, C. M., et al. (2017). N. sreenivasulu. systems genetics identifies a novel regulatory domain of amylose synthesis. *Plant Physiol.* 173 (1), 887–906. doi: 10.1104/pp.16.01248

Wang, J., Hu, P., Lin, L., Chen, Z., Liu, Q., and Wei, C. (2018). Gradually decreasing starch branching enzyme expression is responsible for the formation of heterogeneous starch granules. *Plant Physiol.* 176 (1), 582–595. doi: 10.1104/pp.17.01013

Wang, S., Li, P., Yu, J., Guo, P., and Wang, S. (2017). Multi-scale structures and functional properties of starches from indica hybrid, japonica and *waxy* rice. *Int. J. Biol. Macromol* 102, 136–143. doi: 10.1016/j.ijbiomac.2017.04.020

Wu, Y. P., Pu, C. H., Lin, H. Y., Huang, H. Y., Huang, Y. C., Hong, C. Y., et al. (2015). Three novel alleles of *FLOURY ENDOSPERM2* (*FLO2*) confer dull grains with low amylose content in rice. *Plant Sci.* 233, 44–52. doi: 10.1016/j.plantsci.2014.12.011

Wu, M., Ren, Y., Cai, M., Wang, Y., Zhu, S., Zhu, J., et al. (2019). Rice *FLOURY ENDOSPERM10* encodes a pentatricopeptide repeat protein that is essential for the trans-splicing of mitochondrial nad1 intron 1 and endosperm development. *New Phytol.* 223 (2), 736–750. doi: 10.1111/nph.15814

Xu, Y., Lin, Q., Li, X., Wang, F., Chen, Z., Wang, J., et al. (2021). Fine-tuning the amylose content of rice by precise base editing of the *Wx* gene. *Plant Biotechnol. J.* 19 (1), 11–13. doi: 10.1111/pbi.13433

Zeng, D., Liu, T., Ma, X., Wang, B., Zheng, Z., Zhang, Y., et al. (2020). Quantitative regulation of *Waxy* expression by CRISPR/Cas9-based promoter and 5'UTR-intron editing improves grain quality in rice. *Plant Biotechnol. J.* 18 (12), 2385–2387. doi: 10.1111/pbi.13427

Zeng, Y., Wen, J., Zhao, W., Wang, Q., and Huang., W. (2019). Rational improvement of rice yield and cold tolerance by editing the three genes *OsPIN5b, GS3*, and *OsMYB30* with the CRISPR-Cas9 system. *Front. Plant Sci.* 10, 1663. doi: 10.3389/fpls.2019.01663

Zhang, C., Chen, S., Ren, X., Lu, Y., Liu, D., Cai, X., et al. (2017). Molecular structure and physicochemical properties of starches from rice with different amylose contents resulting from modification of *OsGBSSI* activity. *J. Agric. Food Chem.* 65 (10), 2222–2232. doi: 10.1021/acs.jafc.6b05448

Zhang, A., Gao, Y., Li, Y., Ruan, B., Yang, S., Liu, C., et al. (2020). Genetic analysis for cooking and eating quality of super rice and fine mapping of a novel locus qGC10 for gel consistency. *Front. Plant Sci* 11, 342. doi: 10.3389/fpls.2020.00342

Zhang, H., Xu, H., Jiang, Y., Zhang, H., Wang, S., Wang, F., et al. (2021). Genetic control and high temperature effects on starch biosynthesis and grain quality in rice. *Front. Plant Sci.* 12, 757997. doi: 10.3389/fpls.2021.757997

Zhang, C., Yang, Y., Chen, Z., Chen, F., Pan, L., Lu, Y., et al. (2020). Characteristics of grain physicochemical properties and the starch structure in rice carrying a mutated *ALK/SSIIa* gene. *J. Agric. Food Chem.* 68 (47), 13950–13959. doi: 10.1021/acs.jafc.0c01471

Zhang, C., Yang, Y., Chen, S., Liu, X., Zhu, J., Zhou, L., et al. (2021). A rare *Waxy* allele coordinately improves rice eating and cooking quality and grain transparency. *J. Integr. Plant Biol.* 63 (5), 889–901. doi: 10.1111/jipb.13010

Zhang, Y. C., Yu, Y., Wang, C. Y., Li, Z. Y., Liu, Q., Xu, J., et al. (2013). Overexpression of microRNA *OsmiR397* improves rice yield by increasing grain size and promoting panicle branching. *Nat. Biotechnol.* 31 (9), 848–852. doi: 10.1038/nbt.2646

Zhang, J., Zhang, H., Botella, J. R., and Zhu, J.-K. (2018). Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the *Waxy* gene in elite rice varieties. *J. Integr. Plant Biol.* 60 (5), 369–375. doi: 10.1111/jipb.12620

Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H., and Yang, B. (2014). Large Chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Res.* 42 (17), 10903–10914. doi: 10.1093/nar/gku806

Zhou, Z., Robards, K., Helliwell, S., and Blanchard, C. (2002). Composition and functional properties of rice. *Int. J. Food Sci. Technol.* 37 (8), 849–868. doi: 10.1046/j.1365-2621.2002.00625.x

Zhu, H., Li, C., and Gao., C. (2020). Applications of CRISPR-cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* 21 (11), 661–677. doi: 10.1038/s41580-020-00288-9

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 April 2023
DOI 10.3389/fpls.2023.1169858

OPEN ACCESS

EDITED BY
Gaoneng Shao,
Zhejiang University, China

REVIEWED BY
Fangjun Li,
China Agricultural University, China
Weihuan Jin,
Henan Agricultural University, China

*CORRESPONDENCE
Lina Jiang
jiangln@htu.edu.cn
Jianhui Ma
cricaas@163.com
Dejun Han
handj@nwsuaf.edu.cn

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 20 February 2023
ACCEPTED 22 March 2023
PUBLISHED 03 April 2023

CITATION
Ma J, Ye M, Liu Q, Yuan M, Zhang D, Li C, Zeng Q, Wu J, Han D and Jiang L (2023) Genome-wide association study for grain zinc concentration in bread wheat (*Triticum aestivum L.*). *Front. Plant Sci.* 14:1169858. doi: 10.3389/fpls.2023.1169858

Genome-wide association study for grain zinc concentration in bread wheat (*Triticum aestivum L.*)

Jianhui Ma[1*], Miaomiao Ye[1], Qianqian Liu[1], Meng Yuan[1,2], Daijing Zhang[1], Chunxi Li[1], Qingdong Zeng[2], Jianhui Wu[2], Dejun Han[2*] and Lina Jiang[1*]

[1]College of Life Science, Henan Normal University, Xinxiang, China, [2]State Key Laboratory of Crop Stress Biology in Arid Areas, Northwest A&F University, Yangling, Shanxi, China

Introduction: Zinc (Zn) deficiency causes serious diseases in people who rely on cereals as their main food source. However, the grain zinc concentration (GZnC) in wheat is low. Biofortification is a sustainable strategy for reducing human Zn deficiency.

Methods: In this study, we constructed a population of 382 wheat accessions and determined their GZnC in three field environments. Phenotype data was used for a genome-wide association study (GWAS) using a 660K single nucleotide polymorphism (SNP) array, and haplotype analysis identified an important candidate gene for GZnC.

Results: We found that GZnC of the wheat accessions showed an increasing trend with their released years, indicating that the dominant allele of GZnC was not lost during the breeding process. Nine stable quantitative trait loci (QTLs) for GZnC were identified on chromosomes 3A, 4A, 5B, 6D, and 7A. And an important candidate gene for GZnC, namely, TraesCS6D01G234600, and GZnC between the haplotypes of this gene showed, significant difference ($P \leq 0.05$) in three environments.

Discussion: A novel QTL was first identified on chromosome 6D, this finding enriches our understanding of the genetic basis of GZnC in wheat. This study provides new insights into valuable markers and candidate genes for wheat biofortification to improve GZnC.

KEYWORDS

***Triticum aestivum* L., grain Zn concentration, genome-wide association study, quantitative trait loci, candidate gene**

1 Introduction

Zinc (Zn) is an essential microelement for human normal metabolism. Zn deficiency causes serious diseases, such as liver cirrhosis, dwarfism, coronary heart disease, visual disorders, reproductive organ development disorders, and even cancer (Hambidge, 2000). Furthermore, approximately 17% of the world population is affected by Zn deficiency,

especially women and children in developing countries (Shah and Sachdev, 2006; Wessells and Brown, 2012; Black et al., 2013). Bread wheat (*Triticum aestivum* L.) is one of the most important cereal crops in the world; its products provide approximately 20% of the energy and protein in the human diet (Ludwig and Slamet-Loedin, 2019). However, the grain Zn concentration (GZnC) is below the minimum level (38 mg/kg) required to meet human needs and cannot satisfy the demand from the people consuming wheat as their main staple food (Bouis et al., 2011). Therefore, increasing wheat GZnC is an important goal in wheat breeding.

Owing to the important role of wheat in human nutrition, wheat biofortification can be used as a cost-effective strategy to alleviate Zn deficiency, especially in low-income countries, where most people rely on cereal foods for a basic diet (Velu et al., 2014; Andersson et al., 2017; Ma et al., 2022). However, owing to the ambiguity of the genetic architecture and molecular processes regulating Zn homeostasis in wheat, breeding outstanding varieties with high GZnC is a difficult challenge (Gupta et al., 2021). Identification of molecular markers that are closely linked to QTLs controlling complex quantitative traits, such as grain iron, zinc, or protein concentration, are a goal on their own to facilitate the development of biofortified wheat cultivars through marker-assisted breeding, whereby, improving our understanding of the genetic basis of GZnC in wheat demands identifying as many causal loci as possible. Currently, genome-wide association study (GWAS) is the most popular method for analysing the genetic basis of complex traits in wheat (Breseghello and Sorrells, 2006; Bradbury et al., 2007).

Recently, some studies have been conducted to identify QTLs for wheat GZnC by GWAS to further improvement. Specifically, to the best of our knowledge, nine GWAS studies have been conducted to identify QTLs and marker-trait associations (MTAs) for wheat GZnC. Thus, for example, Velu et al. identified 39 significant MTAs for wheat GZnC located on chromosomes 1A, 2A, 2B, 2D, 5A, 6B, 6D, 7B, and 7D, explaining 5% to 10.5% of the phenotypic variation (Velu et al., 2018). In turn, Alomari et al. detected 40 MTAs on chromosomes 2A, 3A, 3B, 4A, 4D, 5A, 5B, 5D, 6D, 7A, 7B, and 7D for wheat GZnC, which explained 2.5% to 5.2% of the phenotypic variation (Alomari et al., 2018). Similarly, Cu et al. identified 72 MTAs for wheat GZnC on chromosomes 1A, 2A, 3A, 4A, 5B, and 7A, explaining 3.7% to 5.2% of the phenotypic variation (Cu et al., 2020). Furthermore, Liu et al. detected 16 QTLs for wheat GZnC on chromosomes 1B, 3A, 3D, 4A, 4B, 5A, 5D, 6B, 6D, and 7D, explaining 2.7% to 6.6% of the phenotypic variation (Liu et al., 2021). In turn, Zhou et al. detected 29 QTLs on chromosomes 1A, 1B, 1D, 2B, 2D, 3A, 3B, 3D, 4A, 4B, 5A, 5B, 6A, 6B, 6D, and 7A, which explained 9.75% to 24.77% of the phenotypic variation (Zhou et al., 2020). Tong et al. identified 25 QTLs on chromosomes 1A, 2A, 3A, 3B, 5A, 5D, 6A, 6B, 6D, 7A, 7B, and 7D, which explained 7.73% to 13.57% of the phenotypic variation (Tong et al., 2022). Rathan et al. identified two MTAs on chromosomes 1A and 7B, explaining 6.35% to 7.60% of the phenotypic variation (Rathan et al., 2022). Juliana et al. identified 67 MTAs on chromosomes 1A, 1B, 2A, 2D, 3B, 4A, 5A, 5B, 5D, 6B, 6D, and 7B, among which the maximum phenotypic variation explained is 7.3% (Juliana et al., 2022). Krishnappa et al. found five MTAs on chromosomes 2B, 5B, 6A, and 7B, contributing to 5.7% to 10.9% of the phenotypic variation (Krishnappa et al., 2022). However, they are insufficient for further map-based cloning.

Using of different germplasm resources and high resolution genotyping techniques may contribute to identify important QTLs for wheat GZnC; therefore enriching the genetic information. In this study, 382 wheat accessions were used to determine wheat GZnC in three environments to analysis the changes of wheat GZnC. Additionally, a GWAS using a 660K single nucleotide polymorphism (SNP) array was conducted to identify significant QTLs and candidate genes for GZnC. Thus, this study aimed to provide useful information for further GZnC improvement for in wheat.

2 Materials and methods

2.1 Plant materials

Based on previous diversity assessments, the experimental material used herein comprised 382 representative wheat accessions, including 68 exotic cultivars, 43 landraces, and 271 domestic cultivars. These accessions were planted in the cropping season from 2017 to 2018 in Nanyang (33.03°N, 112.50°E), Suqian (34.02°N, 118.33°E), and Yangling (34.16°N, 108.40°E), where the soil Zn concentrations were 221.38, 384.86, and 233.25 mg/kg, respectively. The field was managed according to the local standard agronomic practices.

2.2 Determination of GZnC

Wheat grains were harvested manually in the field. Three biological replicates were sampled from each wheat accession in each environment. Wheat grains were washed with distilled water and dried to avoid potential contamination. Dried wheat grains were ground to whole wheat flour using a stainless steel grinder (MM400, Retsch, Haan, Germany), and whole wheat flour was subsequently oven-dried at 80 °C for 12 h. Subsequencely, 0.2 g of whole wheat flour was digested in 8 ml of a mixture of high-purity concentrated nitric acid and hydrogen peroxide (HNO_3/H_2O_2, 75/25, v/v), and then diluted with ultrapure water. After filtration, the zinc concentration in the digestion solution was determined using an inductively coupled plasma system (iCAP 7000, Thermo Scientific) according to the method described by Ma et al. (2022). And a standard curve (0.07, 0.14, 0.21, 0.28, 0.35, 0.42 mg/L) was used for each round of determinations. Finally, the GZnC (mg/kg) was calculated on a dried weight basis.

2.3 Statistical analysis

Analysis of variance (ANOVA) and *t*-test were performed using SPSS 26.0. The following formula was used for estimating generalised heritability (H^2):

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{GE}^2/n + \sigma_e^2/nr}$$

10.3389/fpls.2023.1169858

Where σ_G^2 represents the phenotypic variation due to the genotype, σ_{GE}^2 represents the phenotypic variation due to the environment × genotype interaction, σ_e^2 represents error variance, n represents the number of environments, and r represents the number of replicates. The best linear unbiased predictions (BLUPs) of GZnC for each accession across the three environments were calculated using the mixed linear model in the "lem4" package, version 3.5.3 of R, and were used for further GWAS.

2.4 SNP genotyping and screening

The Affymetrix Wheat 660K SNP array was used to genotype the wheat panel. SNP genotype calls and allele clustering were analyzed using the Affmetrix Genotyping Console software. The parameters of allele frequency (MAF) less than 0.05 and missing data greater than 10% were used to filter the SNP marker to ensure the accuracy of genotyping, together with a Hardy-Weinberg equilibrium value greater than 0.01 (Wu et al., 2021), resulting in a total of 405,606 high-quality SNP markers for GWAS. The physical locations of all SNP markers were determined using the bread wheat reference genome of IWGSC RefSeq v1.0 (I.W.G.S.C, 2018).

2.5 Linkage disequilibrium analysis, population structure and GWAS

Genome-wide LD analysis was performed using PLINK. The squared correlation (r^2) of allele frequencies was plotted in R Studio using genetic distance (Mb). The LD decay pattern was determined as the distance at which the LD value was reduced to half of its maximum value. STRUCTURE v2.3.4 (unlinked markers (r^2 = 0)) was used to calculate population structure (Pritchard et al., 2000). Each K value with a burn-in length of 20,000 was calculated five times, and iterations were set to be 10,000. The delta K (ΔK) method was used to determine the most likely number of subpopulations based on the change rate between the K values (Earl and vonHoldt, 2012). The general linear model in the GEMMA software was used to conduct GWAS, and BLUPs of 382 wheat accessions were used as phenotypic data for GWAS. After calculating the suggestive P-value threshold based on Bonferroni's correction (P = 1/Ne, where Ne represents the effective number of independent SNPs), we considered -log10 (P-value) ≥ 4.5 to be significant. The -log10 (P-value) values for all SNPs were plotted on a Manhattan plot.

2.6 Candidate gene identification

Non-synonymous mutations were obtained using significant SNPs in the selected QTLs, and phenotypic differences between haplotypes in the three environments were analyzed to identify the candidate genes. Root, stem, and leaf tissues at the jointing and heading stages, as well as wheat grains at 5, 10, 20, and 25 days post-anthesis (DPA), were collected from cv. Chinese Spring, which has been used for genome sequencing (I.W.G.S.C, 2018), for quantitative real-time PCR (qRT-PCR). Total RNA of each sample was extracted using TRIzol reagent (TaKaRa, Japan) following the manufacturer instructions, and first-strand cDNA was synthesised uisng the HiScript III First-stand cDNA synthesis kit (Vazyme, Nanjing, China). In turn, qPCR was conducted on an ABI 7500 real-time PCR system (Applied Biosystems, USA) using a SYBR Premix Ex Taq Kit (TaKaRa, Japan). *TaTubulin* was used as an internal control, and the related expression levels were calculated using the $2^{-\triangle\triangle Ct}$ method. Primers for qRT-PCR are listed in Table S1.

3 Results

3.1 Phenotypic analysis of GZnC

The GZnC of 382 wheat accessions in three environments was determined (Table S2), and the corresponding maximum, minimum, median, variance, coefficient of variation, skewness, and kurtosis were calculated (Table 1). The highest GZnC values were recorded in Suqian, the environment with the highest soil Zn concentration, and ranged from 22.31 to 92.38 mg/kg (average 35.73 mg/kg). Those recorded in Nanyang followed, ranging from 13.55 to 60.59 mg/kg (average 22.64 mg/kg). In turn, the wheat accessions in Yangling showed lowest GZnC, ranging from 11.92 to 36.65 mg/kg (average 21.50 mg/kg). BLUPs of GZnC based on three environments were calculated in the range 24.78~39.21 mg/kg. ANOVA revealed that the genotype, environment, and interaction of genotype and environment had significant effects on GZnC (P<0.001). The H^2 of GZnC was determined to be 0.32, confirming that both environment and genotype had significant effects on GZnC (Table S3). In addition, the GZnC showed an approximately normal distribution in the three environments and BLUPs indicated by skewness, kurtosis, and GZnC distribution (Table 1; Figure 1). These findings indicated that the wheat panel was suitable for GWAS.

TABLE 1 Phenotypic data of GZnC in three environments and BLUP.

Environment	Min (mg/kg)	Max (mg/kg)	Mean (mg/kg)	Standard deviation (mg/kg)	Coefficient of variation (%)	Skewness	Kurtosis
Yangling	11.92	36.65	21.50	4.64	0.21	0.62	0.65
Nanyang	13.55	60.59	22.64	4.79	0.21	0.67	1.14
Suqian	22.31	92.38	35.73	9.41	0.26	0.57	0.02
BLUP	24.78	38.21	27.10	1.14	0.04	0.38	-0.20

10.3389/fpls.2023.1169858

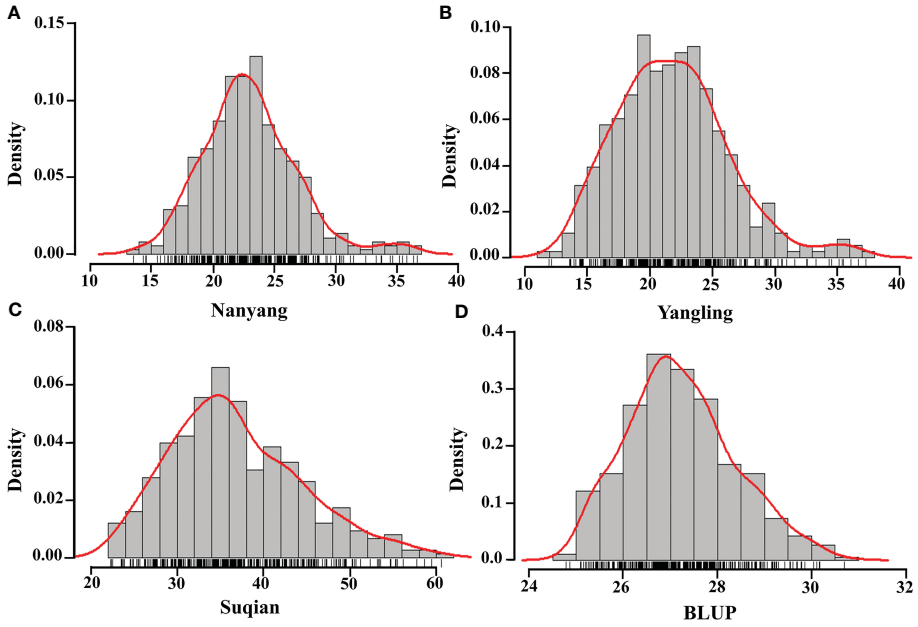


FIGURE 1
Histogram of wheat GZnC density in three environments and BLUP. Nanyang **(A)**, Yangling **(B)**, Suqian **(C)**, BLUP **(D)**, and the red curves and the black short horizontal lines represent the density curve and the rug plot of the distribution of GZnC.

3.2 GZnC variation with source and released years

Based on source, these wheat accessions were divided into three categories: 68 exotic cultivars, 43 landraces, and 271 domestic cultivars. Domestic cultivars showed the highest wheat GZnC, followed by landraces and exotic cultivars (Table 2). These wheat accessions were further divided into five stages according to the released years. In this respect, we found that GZnC increased from the pre-1950s to the post-1990s in Suqian and Yangling (Figures 2A, B), while it showed slightly fluctuation in Nanyang (Figure 2C). These findings indicate that the predominant allele may be selected during wheat breeding.

3.3 Genotyping by SNP array and linkage disequilibrium analysis

The 660K SNP array was used for conducting a GWAS using the phenotypes of 382 wheat accessions. After removing low-quality SNPs (MAF< 0.05 and missing data > 0.1), 412,619 SNPs remained for subsequent analysis. A set of 405,606 SNPs was distributed on 21 chromosomes, of which the B subgenome (185,057) was the most labelled, followed by the A subgenome (162,316), and the D genome (58,233). Chromosomal distribution of SNP markers showed that chromosome 3B contained the largest number of SNP markers (45,867) and the least number was found on chromosome 4D (4,046). Further, marker density distribution was uneven across chromosomes, ranging from 5.67 markers per Mb (4D) to 66.53 markers per Mb (3B). In addition, the polymorphism information content of the three sub-genomes was 0.28, 0.29, and 0.27, respectively (Table 3).

Population structure analysis was performed for these 382 wheat accessions. According to the ΔK method of Bayesian clustering, the slope broke when K = 8. As a result, the 382 wheat accessions were divided into eight subpopulations, SP1-SP8 (Table S2; Figures 3A, B). SP1 contained eight exotic cultivars, two landraces, and 56 domestic cultivars, with an average GZnC of 27.36 mg/kg. SP2 contained four exotic cultivars, 35 landraces, and six domestic cultivars, with an average GZnC of 26.04 mg/kg. SP3 contained 59 domestic cultivars, with an average GZnC of 27.31 mg/kg. SP4 contained one exotic cultivars, two landraces, and 34 domestic cultivars, with an average GZnC of 27.29 mg/kg. SP5 contained two exotic cultivars and 38 domestic cultivars, with an average GZnC of 28.10 mg/kg. SP6 contained one exotic cultivar, one landraces, and 38 domestic cultivars, with an average GZnC of

TABLE 2 GZnC of different source in three environments.

Wheat accessions	Nanyang (mg/kg)	Suqian (mg/kg)	Yangling (mg/kg)
Exotic cultivars	22.93	36.46	19.96
Landraces	23.42	34.56	20.03
Domestic cultivars	22.98	37.30	22.56

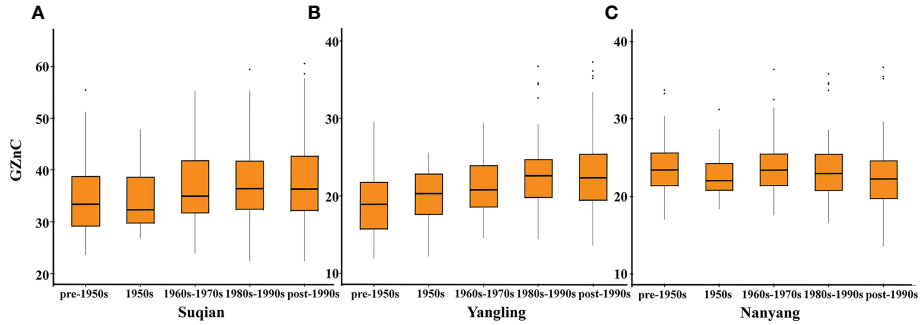


FIGURE 2
The changes of GZnC in Suqian **(A)**, Yangling **(B)**, and Nanyang **(C)** with released years.

27.90 mg/kg. SP7 contained 25 exotic cultivars, with an average GZnC of 26.46 mg/kg. SP8 contained 27 exotic cultivars, three landraces, and 40 domestic cultivars, with an average GZnC of 27.14 mg/kg (Table S4). We found that the GZnC fluctuated slightly among the eight subpopulations. LD was estimated using SNPs and squared allele frequency correlations (r^2) for subgenomes A (162,316), B (185,057), and D (58,233). As expected, LD decayed with increasing physical distance, and it differed among subgenomes. Further, whole-genome-wide LD decayed with genetic distance, and it decayed to half of the genome at a genetic distance of 3.6 Mb. Subgenomes A, B, and D were 3.0 Mb, 5.7 Mb, and 1.6 Mb, respectively (Figure 3C).

3.4 GWAS for GZnC

The BLUPs based on the GZnC across three environments were used for the GWAS using the Lm4 model. SNPs with –log10 (P-value) ≥ 4.5 were deemed significant (Figure 4). A total of nine significant QTLs were selected, locating on chromosomes 3A, 4A, 5B, 6D, and 7A. For convenience, the SNP markers with the highest threshold in each QTL were used to represent the corresponding QTL, and the effects of the nine QTLs ranged from 0.16 to 0.30 (Table 4).

3.5 Identification of candidate genes for wheat GZnC

A total of 494 genes were found in the nien QTLs. The functional annotations of these genes are provided in Table S5. We further analyzed significant SNPs in the nine QTLs that might cause missense mutations. Interestingly, we found that the phenotype between alleles, which was caused by the SNP AX-108884748, showed significant differences in the three environments, whereby it was further analyzed, as it may be closely related to GZnC.

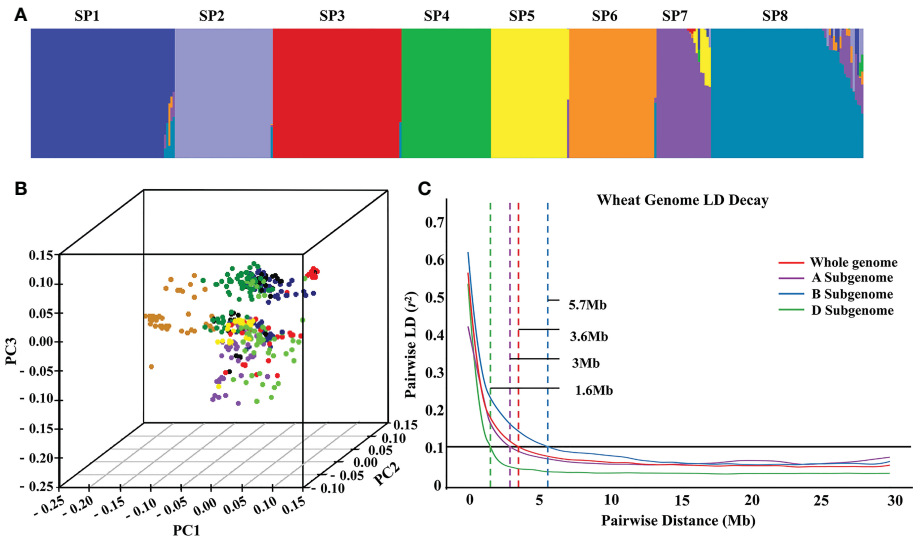


FIGURE 3
The population structure and linkage disequilibrium of 382 wheat accessions. Subpopulations inferred by K-mean structure analysis **(A)**. Principal component analysis of all wheat accessions **(B)**. LD decay **(C)** over different genetic distances (Mb) for A, B and D subgenomes and whole genome in the wheat panel (r^2 = 0.1).

The SNP, AX-108884748, is located at 328,977,424 bp on chromosome 6D (Figure 5A), and may cause an amino acid change from Arg (CC haplotype) to Ser (AA haplotype) at 1047 bp in the GDSL esterase-encoding gene, *TraesCS6D01G234600* (Figure 5B). The 382 wheat accessions were divided into three categories according to genotype: among them, 250 wheat accessions with the CC haplotype, 123 with the AA haplotype, and 9 with heterozygosity or lack of genotype. After analysing the phenotypic data of the two haplotypes, the GZnC of the AA haplotype was found to be significantly higher ($P \leq 0.05$) than that of the CC haplotype across three environments (Figure 5C). Further, qRT-PCR analysis showed that the gene was predominantly expressed in the root, stem, and grain tissues, suggesting a possible role in Zn transfer (Figure 5D). Altogether, these results indicate that *TraesCS6D01G234600* is an important candidate gene for wheat GZnC.

4 Discussion

4.1 Increasing wheat GZnC is needed to alleviate "hidden hunger"

Wheat grain yields have significantly increased over the past 75 years, from the early days of the "Green Revolution" in the 1960s to the recent optimisation of breeding and agronomic crop management (Evenson and Gollin, 2003; Pingali, 2012). However, nutrition-related traits are often neglected in the breeding process, resulting in the low concentration of microelements in wheat grains, giving rise to what has been called "hidden hunger" (Ul-Allah, 2018). Zn is an indispensable microelement for the human body. As an essential microelement, Zn cannot be synthesised by the human body, and plays a very important role in the human nervous, immune, and reproductive systems, as well as in the

TABLE 3 Summary of the SNPs information in three sub-genomes and chromosomes of 382 wheat accessions.

Chromosome	No.of markers	Effective number	Effective Ratio	Suggestive *P*-value	Markers (%)	Length (Mb)	Marker density	He	PIC
1A	28,959	5,687	0.20	1.76E-04	7.14	594.02	48.75	0.66	0.27
1B	20,624	4,796	0.23	2.09E-04	5.08	780.76	26.42	0.76	0.30
1D	10,584	3,090	0.29	3.24E-04	2.61	750.73	14.10	0.67	0.28
2A	28,832	6,170	0.21	1.62E-04	7.11	744.54	38.72	0.70	0.28
2B	28,645	6,708	0.23	1.49E-04	7.06	709.76	40.36	0.73	0.29
2D	10,449	3,744	0.36	2.67E-04	2.58	617.97	16.91	0.69	0.28
3A	19,409	4,611	0.24	2.17E-04	4.79	736.69	26.35	0.69	0.28
3B	45,867	7,828	0.17	1.28E-04	11.31	689.38	66.53	0.77	0.31
3D	7,247	2,678	0.37	3.73E-04	1.79	801.25	9.04	0.67	0.27
4A	17,856	4,258	0.24	2.35E-04	4.40	830.70	21.50	0.67	0.27
4B	13,120	3,021	0.23	3.31E-04	3.23	673.47	19.48	0.71	0.29
4D	4,046	1,666	0.41	6.00E-04	1.00	713.02	5.67	0.66	0.27
5A	22,616	4,947	0.22	2.02E-04	5.58	720.95	31.37	0.78	0.31
5B	33,966	6,539	0.19	1.53E-04	8.37	750.61	45.25	0.80	0.31
5D	8,396	3,267	0.39	3.06E-04	2.07	495.44	16.95	0.65	0.26
6A	16,383	3,834	0.23	2.61E-04	4.04	651.81	25.13	0.73	0.29
6B	25,783	5,565	0.22	1.80E-04	6.36	615.48	41.89	0.69	0.28
6D	7,425	2,924	0.39	3.42E-04	1.83	509.85	14.56	0.65	0.26
7A	28,261	6,280	0.22	1.59E-04	6.97	566.04	49.93	0.69	0.28
7B	17,052	4,499	0.26	2.22E-04	4.20	473.56	36.01	0.72	0.29
7D	10,086	3,682	0.37	2.72E-04	2.49	638.65	15.79	0.64	0.26
A genome	162,316					4934.47	32.89	0.70	0.28
B genome	185,057					5719.38	32.36	0.74	0.29
D genome	58,233					3950.83	14.74	0.66	0.27
Total	405,606					14064.68	28.84	0.70	0.28
Average				2.51E-04					

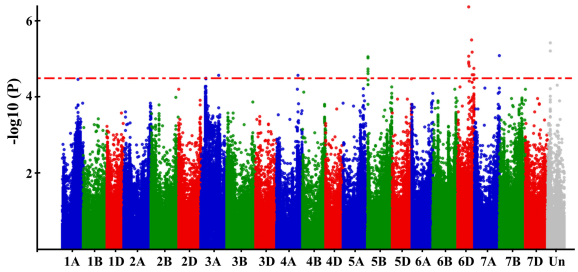


FIGURE 4
Manhattan plots of the GWAS results for GZnC. The resulting BLUPs of GZnC of 382 wheat accessions across three environments was used for GWAS. -log10 (*P*-value) of each SNP was shown in the Manhattan plot.

growth and development of children. Hao et al. analyzed wheat GZnC over the past 80 years in China, and found no decrease with released years (Hao et al., 2021). In turn, this study revealed that wheat GZnC has increased with the released years. Therefore, the dominant allele of wheat GZnC was never lost during selective breeding.

As one of the main food crops in the world, bread wheat is the staple food and major source of microelements for 30% to 40% of the world population (Poudel and Bhatta, 2017). Given the significant impact of Zn on human health (Shewry et al., 2012), it is unfortunate that, as many studies have shown, wheat GZnC is low worldwide. Thus, for example, Rehman et al. analyzed 28 wheat accessions in Pakistan and found that wheat GZnC ranged from 21.20 to 54.40 mg/kg (Rehman et al., 2018). Similarly, Khokhar et al. analyzed 245 wheat accessions from Woestkin and found that GZnC ranged from 24.0 to 49.0 mg/kg (Khokhar et al., 2020). In

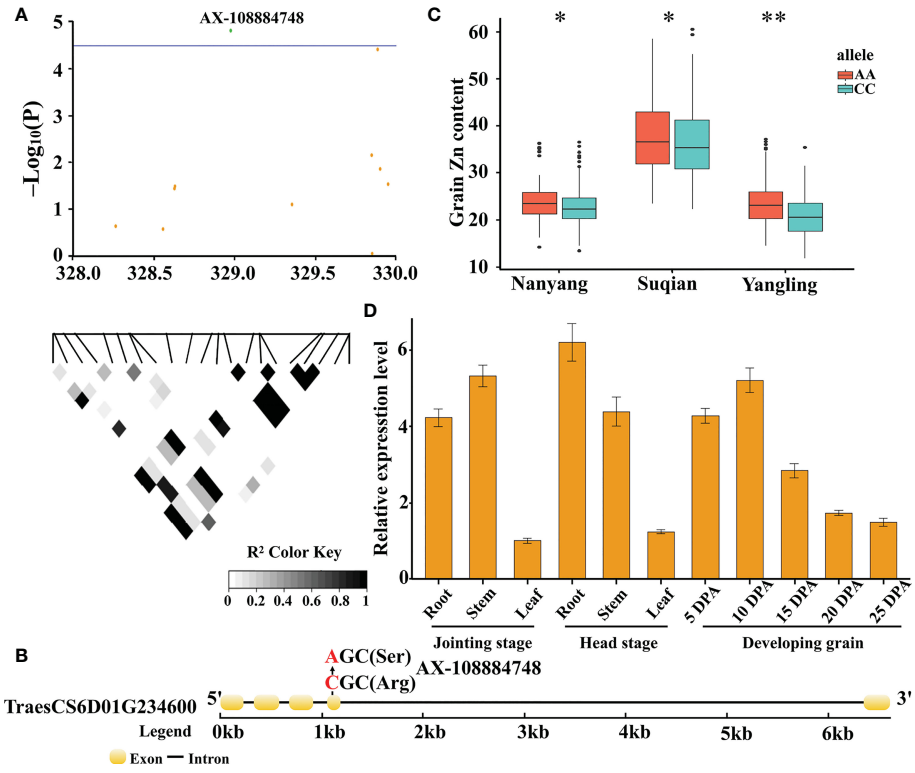


FIGURE 5
Candidate gene analysis based on the SNP AX-108884748 on chromosome 6D. Local Manhattan plot for the candidate region on chromosome 6D. The green dot represents the significant SNP AX-108884748. The corresponding LD block analysis of SNPs in this region is shown below. The degree of linkage is represented by the coefficient of r^2 **(A)**. Gene structure and location of the non-synonymous SNP for *TraesCS6D01G234600*. Yellow rectangles and black lines represent exons and introns, respectively **(B)**. Boxplots for GZnC based on haplotype analysis **(C)**. Upper and lower edges of the box represent the 75th and 25th quantiles, respectively, and the whiskers show the 90th and 10th quantiles; the horizontal solid lines represent the median. Statistical significance was analyzed by *t*-test, * $P \leq 0.05$; ** $P \leq 0.01$. Expression profile of *TraesCS6D01G234600* in different tissues, as determined by qRT-PCR. The data was shown as the means ± Sd of $n = 3$ **(D)**.

turn, Maryami et al. analyzed 158 Iranian wheat accessions and found that GZnC ranged from 27.9 to 65.0 mg/kg (Maryami et al., 2020). In particular, according to a summary and analysis of a large number of studies worldwide by Wang et al., the average global zinc concentration in wheat grains is only 28.48 mg/kg (Wang et al., 2020). Consistently, the results reported herein revealed that the GZnC ranged from 11.92 to 92.83 mg/kg. Although domestic wheat cultivars contained the highest GZnC, they were still far below human requirements, at approximately 40 to 60 mg/kg (Cakmak, 2008). Therefore, there is an urgent need to improve GZnC in wheat.

4.2 Genetic structure analysis was performed by comparing known loci

As a complementary strategy, GWAS is a powerful tool to for detecting QTLs for complex traits (Hamblin et al., 2011). With the development of wheat genome sequencing, different SNP arrays have been developed and gradually become the main tools for wheat GWAS (Rasheed et al., 2017). To reveal the QTLs for wheat GZnC, a 35K SNP array was used to identify two QTLs located on chromosomes 1A and 7B (Rathan et al., 2022). A wheat 660K SNP array was used to map seven QTLs on chromosomes 1B, 3B, 3D, 4A, 5A, 5B, and 7A (Zhou et al., 2020). Subsequently, the 90K and 660K SNP arrays were used to identify 17 QTLs on chromosomes 1A, 2A, 3A, 3B, 5A, 5D, 6A, 6B, 6D, 7A, 7B, and 7D, which were related to wheat GZnC (Tong et al., 2022). However, the molecular mechanism underlying Zn accumulation in wheat grains remains unclear, and no molecular marker conducive to GZnC has been used for wheat breeding. In this study, we identified nine QTLs on chromosomes 3A, 4A, 5B, 6D, and 7A. Specifically, the QTL (AX-108884748) was first identified on chromosome 6D, and is most likely a novel QTL located in the chromosomal region. This finding enriches our understanding of the genetic basis of GZnC in wheat.

4.3 Candidate genes for GZnC in wheat

A few genes have been functionally verified aiming to improve wheat GZnC. For example, the NAC transcription factor, NAM-B1, promotes the transfer of Zn from the leaves to the grains to increase GZnC (Uauy et al., 2006). However, wheat GZnC is a quantitative trait that is contributed by many genes. To date, many studies have been performed using GWAS to screen candidate genes for complex traits in wheat. Tong et al. used three methods: haplotype analysis, gene function comparison, and lineal homologues to screen a total of 28 promising candidate genes that might be involved in zinc/iron absorption, transport, storage, and regulation (Tong et al., 2022). In turn, Zhou et al. used superior allele estimation to identify seven candidate genes for GZnC, encoding NAC transcription factor and TPR-like superfamily proteins (Zhou et al., 2020). Meanwhile, using haplotype analysis, Krishnappa et al. identified two important candidate genes for GZnC, encoding a late embryogenic rich protein, *LEA-18*, and RNA recognition motif domains (Krishnappa et al., 2022). These findings indicate that GWAS is a highly useful tool for identifying candidate genes for complex traits. In addition, these studies showed haplotype analysis is reliable for screening candidate genes for complex traits, which laid a solid theoretical foundation for our subsequent research.

In this study, we identified the candidate gene *TraesCS6D01G234600* on chromosome 6D through haplotype analysis, which was caused by the SNP AX-108884748 (CC/AA). This candidate gene encodes GDSL esterase, which shows hydrolytic enzyme activity for thioesters, aryl esters, phospholipids and amino acids, and others (Akoh et al., 2004). The AA allele had a significant positive effect on GZnC in all three environments ($P \leq 0.05$), indicating the importance of *TraesCS6D01G234600*. In addition, this gene was mainly expressed in the root, stem, and grain tissues, suggesting that it may play an important role in Zn transfer. These results provide important information for further improvement of wheat GZnC, and indicate that the pyramid effect should be considered in the next step of selective breeding.

TABLE 4 Significant QTLs associated with GZnC.

Trait[a]	Chromosome	Marker[b]	Physical position[c] (bp)	Predominant allele and its ratio[d]	-log$_{10}$(*P-value*)	Effect
GZnC	3A	AX-109368604	140,810,282	G/A (0.12)	4.50	0.19
	4A	AX-109872817	621,769,441	G/A (0.38)	4.57	0.20
	5B	AX-109350479	13,177,703	G/T (0.46)	5.06	0.21
	6D	AX-108814900	307,646,180	G/A (0.39)	6.38	0.30
	6D	AX-108884748	328,977,424	A/C (0.33)	4.81	0.26
	6D	AX-89576368	403,654,827	T/C (0.45)	5.51	0.27
	6D	AX-108812106	410,167,022	C/T (0.49)	5.19	0.16
	6D	AX-94571657	469,831,635	C/G (0.11)	4.76	0.29
	7A	AX-109999591	726,436,979	T/C (0.33)	5.08	0.29

[a]GZnC, grain zinc concentration.
[b]The SNP markers with the highest threshold in each QTL were used to represent the corresponding QTL.
[c]Physical positions of single nucleotide polymorphism (SNP) markers were based on IWGSC RefSeq v.1.0.
[d] "–" indicates the predominant allele with its ratio on GZnC.

5 Conclusion

We constructed a diversity panel comprising 382 wheat accessions, and conducted field trials in three environments. The GZnC of 382 wheat accessions was determined, and the data showed that wheat GZnC increased with the release years, indicating that the predominant allele of wheat GZnC was not lost during selected breeding. Phenotypic data were further analyzed using GWAS, and nine QTLs for GZnC were identified, with effects ranging from 0.16 to 0.30. In addition, one candidate gene for GZnC was screened using haplotype analysis. Overall, our study provides novel insights that increase our understanding of the genetic information of GZnC and will facilitate the improvement of GZnC in wheat breeding programs.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

JM, DH, and LJ conceived and designed the research. JM, MYe, QL, MYuan, DZ, CL, QZ, and JW performed the experiments. JM, MYe, and MYuan prepared the figures and provided the materials. JM wrote the manuscript. All authors have read and approved the final manuscript.

Funding

This study was supported by the Open Project Program (CSBAA2020001) of State Key Laboratory of Crop Stress Biology for Arid Areas, NWAFU, Yangling, Shaanxi, 712100, China; the National Key Research and Development Program of China (2018YFD0300703-2 and 2017YFD0301101); Technological development for wheat production (H2022028).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1169858/full#supplementary-material

References

Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H., and Shaw, J. F. (2004). GDSL family of serine esterases/lipases. *Prog. Lipid Res.* 43, 534–552. doi: 10.1016/j.plipres.2004.09.002

Alomari, D. Z., Eggert, K., von Wirén, N., Alqudah, A. M., Polley, A., Plieske, J., et al. (2018). Identifying candidate genes for enhancing grain zn concentration in wheat. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01313

Andersson, M. S., Saltzman, A., Virk, P. S., and Pfeiffer, W. (2017). Progress update: Crop development of biofortified staple food crops under HarvestPlus. *Afr. J. Food Agric. Nutr. Dev.* 17, 11905–11935. doi: 10.18697/ajfand.78.HarvestPlus05

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., et al. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet* 382 (9890), 427–451. doi: 10.1016/S0140-6736(13)60937-X

Bouis, H. E., Hotz, C., McClafferty, B., Meenakshi, J. V., and Pfeiffer, W. H. (2011). Biofortification: A new tool to reduce micronutrient malnutrition. *Food Nutr. Bull.* 32 (1 Suppl), S31–S40. doi: 10.1177/15648265110321S105

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

Breseghello, F., and Sorrells, M. E. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.* 46, 1323–1330. doi: 10.2135/cropsci2005.09-0305

Cakmak, I. (2008). Enrichment of cereal grains with zinc: Agronomic or genetic biofortification? *Plant Soil* 302, 1–17. doi: 10.1007/s11104-007-9466-3

Cu, S. T., Guild, G., Nicolson, A., Velu, G., Singh, R., and Stangoulis, J. (2020). Genetic dissection of zinc, iron, copper, manganese and phosphorus in wheat (*Triticum aestivum* l.) grain and rachis at two developmental stages. *Plant Sci.* 291, 110338. doi: 10.1016/j.plantsci.2019.110338

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

Evenson, R. E., and Gollin, D. (2003). Assessing the impact of the green revolution 1960 to 2000. *Science* 300 (5620), 758–762. doi: 10.1126/science.1078710

Gupta, P. K., Balyan, H. S., Sharma, S., and Kumar, R. (2021). Biofortification and bioavailability of zn, fe and Se in wheat: present status and future prospects. *Theor. Appl. Genet.* 134 (1), 1–35. doi: 10.1007/s00122-020-03709-7

Hambidge, M. (2000). Human zinc deficiency. *J. Nutr.* 130 (5S Suppl), 1344S–1349S. doi: 10.1093/jn/130.5.1344S

Hamblin, M. T., Buckler, E. S., and Jannink, J. L. (2011). Population genetics of genomics-based crop improvement methods. *Trends Genet.* 27 (3), 98–106. doi: 10.1016/j.tig.2010.12.003

Hao, B., Ma, J., Chen, P., Jiang, L., Wang, X., Li, C., et al. (2021). Wheat breeding in China over the past 80 years has increased grain zinc but decreased grain iron concentration. *Field Crops Res.* 271, 108253. doi: 10.1016/j.fcr.2021.108253

I.W.G.S.C (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191. doi: 10.1126/science.aar7191

Juliana, P., Govindan, V., Crespo-Herrera, L., Mondal, S., Huerta-Espino, J., Shrestha, S., et al. (2022). Genome-wide association mapping identifies key genomic

regions for grain zinc and iron biofortification in bread wheat. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.903819

Khokhar, J. S., King, J., King, I. P., Young, S. D., Foulkes, M. J., De Silva, J., et al. (2020). Novel sources of variation in grain zinc (Zn) concentration in bread wheat germplasm derived from Watkins landraces. *PLoS One* 15 (2), e0229107. doi: 10.1371/journal.pone.0229107

Krishnappa, G., Khan, H., Krishna, H., Kumar, S., Mishra, C. N., Parkash, O., et al. (2022). Genetic dissection of grain iron and zinc, and thousand kernel weight in wheat (Triticum aestivum l.) using genome-wide association study. *Sci. Rep.* 12 (1), 12444. doi: 10.1038/s41598-022-15992-z

Liu, Y., Chen, Y., Yang, Y., Zhang, Q., Fu, B., Cai, J., et al. (2021). A thorough screening based on QTLs controlling zinc and copper accumulation in the grain of different wheat genotypes. *Environ. Sci. pollut. Res.* 28 (12), 15043–15054. doi: 10.1007/s11356-020-11690-3

Ludwig, Y., and Slamet-Loedin, I. H. (2019). Genetic biofortification to enrich rice and wheat grain iron: From genes to product. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00833

Ma, J., Qi, S., Yuan, M., Zhao, D., Zhang, D., Feng, J., et al. (2022). A genome-wide association study revealed the genetic variation and candidate genes for grain copper content in bread wheat (Triticum aestivum l.). *Food Funct.* 13 (9), 5177–5188. doi: 10.1039/d1fo04173h

Maryami, Z., Huertas-García, A. B., Azimi, M. R., Hernández-Espinosa, N., Payne, T., Cervantes, F., et al. (2020). Variability for glutenins, gluten quality, iron, zinc and phytic acid in a set of one hundred and fifty-eight common wheat landraces from Iran. *Agronomy* 10, 1797. doi: 10.3390/agronomy10111797

Pingali, P. L. (2012). Green revolution: impacts, limits, and the path ahead. *Proc. Natl. Acad. Sci. U.S.A.* 109 (31), 12302–12308. doi: 10.1073/pnas.0912953109

Poudel, R., and Bhatta, M. (2017). Review of nutraceuticals and functional properties of whole wheat. *J. Nutr. Food Sci.* 7, 1–6. doi: 10.4172/2155-9600.1000571

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155 (2), 945–959. doi: 10.1093/genetics/155.2.945

Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., et al. (2017). Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. *Mol. Plant* 10 (8), 1047–1064. doi: 10.1016/j.molp.2017.06.008

Rathan, N. D., Krishna, H., Ellur, R. K., Sehgal, D., Govindan, V., Ahlawat, A. K., et al. (2022). Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* l.). *Sci. Rep.* 12 (1), 7037. doi: 10.1038/s41598-022-10618-w

Rehman, A., Farooq, M., Nawaz, A., Al-Sadi, A. M., Al-Hashmi, K. S., Nadeem, F., et al. (2018). Characterizing bread wheat genotypes of Pakistani origin for grain zinc biofortification potential. *J. Sci. Food Agric.* 98 (13), 4824–4836. doi: 10.1002/jsfa.9010

Shah, D., and Sachdev, H. P. (2006). Zinc deficiency in pregnancy and fetal outcome. *Nutr. Rev.* 64 (1), 15–30. doi: 10.1111/j.1753-4887.2006.tb00169.x

Shewry, P. R., Lafiandra, D., and Bedo, Z. (2012). Improving the nutritional quality and health benefits of wheat. *Qual. Assur. Saf. Crop Foods* 4, 136–136. doi: 10.1111/j.1757-837X.2012.00138.x

Tong, J., Zhao, C., Sun, M., Fu, L., Song, J., Liu, D., et al. (2022). High resolution genome wide association studies reveal rich genetic architectures of grain zinc and iron in common wheat (*Triticum aestivum* l.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.840614

Uauy, C., Distelfeld, A., Fahima, T., Blechl, A., and Dubcovsky, J. (2006). A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* 314 (5803), 1298–1301. doi: 10.1126/science.1133649

Ul-Allah, S. (2018). Combating hidden hunger in agriculture perspective. *World Rev. Nutr. Diet.* 118, 161–166. doi: 10.1159/000484511

Velu, G., Ortiz-Monasterio, I., Cakmak, I., Hao, Y., and Singh, R. P. (2014). Biofortification strategies to increase grain zinc and iron concentrations in wheat. *J. Cereal Sci.* 59, 365–372. doi: 10.1016/j.jcs.2013.09.001

Velu, G., Singh, R. P., Crespo-Herrera, L., Juliana, P., Dreisigacker, S., Valluru, R., et al. (2018). Genetic dissection of grain zinc concentration in spring wheat for mainstreaming biofortification in CIMMYT wheat breeding. *Sci. Rep.* 8 (1), 13526. doi: 10.1038/s41598-018-31951-z

Wang, M., Kong, F., Liu, R., Fan, Q., and Zhang, X. (2020). Zinc in wheat grain, processing, and food. *Front. Nutr.* 7. doi: 10.3389/fnut.2020.00124

Wessells, K. R., and Brown, K. H. (2012). Estimating the global prevalence of zinc deficiency: results based on zinc availability in national food supplies and the prevalence of stunting. *PLoS One* 7 (11), e50568. doi: 10.1371/journal.pone.0050568

Wu, J., Yu, R., Wang, H., Zhou, C., Huang, S., Jiao, H., et al. (2021). A large-scale genomic association analysis identifies the candidate causal genes conferring stripe rust resistance under multiple field environments. *Plant Biotechnol. J.* 19 (1), 177–191. doi: 10.1111/pbi.13452

Zhou, Z., Shi, X., Zhao, G., Qin, M., Ibba, M. I., Wang, Y., et al. (2020). Identification of novel genomic regions and superior alleles associated with zn accumulation in wheat using a genome-wide association analysis method. *Int. J. Mol. Sci.* 21 (6), 1928. doi: 10.3390/ijms21061928

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 April 2023
DOI 10.3389/fpls.2023.1178624

OPEN ACCESS

EDITED BY
Yue Feng,
China National Rice Research Institute, Chinese Academy of Agricultural Sciences, China

REVIEWED BY
Jian Ma,
Triticeae Research Institute, Sichuan Agricultural University, China
Jirui Wang,
Sichuan Agricultural University, China
Fa Cui,
Ludong University, China

*CORRESPONDENCE
Tian Li
litian@caas.cn
Fang Wang
wangfang@gsau.edu.cn
Chenyang Hao
haochenyang@caas.cn

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 03 March 2023
ACCEPTED 22 March 2023
PUBLISHED 05 April 2023

CITATION
Cao L, Li T, Geng S, Zhang Y, Pan Y, Zhang X, Wang F and Hao C (2023) *TaSPL14-7A* is a conserved regulator controlling plant architecture and yield traits in common wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 14:1178624. doi: 10.3389/fpls.2023.1178624

TaSPL14-7A is a conserved regulator controlling plant architecture and yield traits in common wheat (*Triticum aestivum* L.)

Lina Cao[1,2,3,4], Tian Li[2]*, Shuaifeng Geng[2], Yinhui Zhang[2], Yuxue Pan[2], Xueyong Zhang[2], Fang Wang[1,3,4]* and Chenyang Hao[1,2]*

[1]College of Agronomy, Gansu Agricultural University, Lanzhou, China, [2]Key Laboratory of Crop Gene Resources and Germplasm Enhancement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [3]Gansu Provincial Key Laboratory of Aridland Crop Science, Gansu Agricultural University, Lanzhou, China, [4]Gansu Key Laboratory of Crop Improvement & Germplasm Enhancement, Gansu Agricultural University, Lanzhou, China

Plant architecture is a crucial influencing factor of wheat yield and adaptation. In this study, we cloned and characterized *TaSPL14*, a homologous gene of the rice ideal plant architecture gene *OsSPL14* in wheat. *TaSPL14* homoeologs (*TaSPL14-7A*, *TaSPL14-7B* and *TaSPL14-7D*) exhibited similar expression patterns, and they were all preferentially expressed in stems at the elongation stage and in young spikes. Moreover, the expression level of *TaSPL14-7A* was higher than that of *TaSPL14-7B* and *TaSPL14-7D*. Overexpression of *TaSPL14-7A* in wheat resulted in significant changes in plant architecture and yield traits, including decreased tiller number and increased kernel size and weight. Three *TaSPL14-7A* haplotypes were identified in Chinese wheat core collection, and haplotype-based association analysis showed that *TaSPL14-7A-Hap1/2* were significantly correlated with fewer tillers, larger kernels and higher kernel weights in modern cultivars. The haplotype effect resulted from a difference in *TaSPL14-7A* expression levels among genotypes, with *TaSPL14-7A-Hap1/2* leading to higher expression levels than *TaSPL14-7A-Hap3*. As favorable haplotypes, *TaSPL14-7A-Hap1/2* underwent positive selection during global wheat breeding over the last century. Together, the findings of our study provide insight into the function and genetic effects of *TaSPL14* and provide a useful molecular marker for wheat breeding.

KEYWORDS

***TaSPL14-7A*, tiller number, kernel weight, haplotype analysis, wheat**

Introduction

SPL (SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE) is a type of plant-specific transcription factor. The SPL family is characterized by a highly conserved SBP (SQUAMOSA-PROMOTER BINDING PROTEIN) domain consisting of approximately 76 amino acid residues and the presence of a zinc-binding domain containing two zinc-binding sites, and a conserved nuclear localization signal at the C-terminus (Yamasaki et al., 2004; Birkenbihl et al., 2005). This gene family was found to be widespread and evolutionarily conserved in plants (Preston and Hileman, 2013; Wang and Wang, 2015). There were 17 *SPL* genes in *Arabidopsis* (Yang et al., 2008), 31 in maize (Hultquist and Dorweiler, 2008), 19 in rice (Xie et al., 2006), and 56 in wheat (Zhu et al., 2020). Interestingly, many plant *SPL* genes possess the recognition target sites of miRNA156/157, and their expression is fine-tuned by miRNA156/157 at the transcriptional and posttranscriptional levels, generating an elaborative regulatory module during plant evolution (Gandikota et al., 2007). Increasing evidence indicates that *SPL* genes play an important role in plant growth and development, including leaf initiation (Wang and Li, 2008), shoot maturation (Shikata et al., 2009), fruit ripening (Manning et al., 2006), cell division and grain filling (Wang et al., 2012), and stress response and hormone signal transduction (Stone et al., 2005; Riese et al., 2008; Wang et al., 2009).

In rice, the *IPA1* (*Ideal Plant Architecture 1*) quantitative trait locus encodes OsSPL14 and has great value in rice yield improvement. A point mutation (C-to-A) in *OsSPL14* disrupts the OsmiR156-directed cleavage of its mRNA, thereby increasing *OsSPL14* transcription and resulting in ideal plant architecture traits for rice, such as reduced tiller number and increased lodging resistance and grain yield (Jiao et al., 2010). Furthermore, another epigenetic allele (*ipa1-2D*) also exhibits elevated expression of *OsSPL14* due to the presence of tandem repeat sequences in the promoter region, which is associated with reduced DNA methylation at the *IPA1* promoter, thus alleviating the epigenetic repression of *IPA1* (Zhang et al., 2017). This, in turn, results in a similar ideal plant architecture and enhanced yield. More interestingly, deleting a 54-bp cis-regulatory region in *IPA1 via* a tiling-deletion-based CRISPR-Cas9 screen resolves the trade-off between tiller number and grains per panicle, leading to substantially enhanced grain yield per plant (Song et al., 2022). The deleted fragment is a target site for An-1 transcription factor binding to inhibit *IPA1* expression in panicles and roots, offering a realistic solution for optimizing the expression levels of *IPA1* with a less deleterious effect on over-reduction of tiller number. Thus, *OsSPL14* plays an important role in regulating plant architecture and yield, and fine-tuning its expression level is key to achieving high yield in rice.

Plant architecture is a key determinant of crop productivity and adaptability. It is a complex trait involving multiple factors, such as plant height, branching pattern, and leaf shape (Wang et al., 2018; Hang et al., 2021). Common wheat (*Triticum aestivum* L., AABBDD) is a major staple crop and an important human food source. Optimizing plant architecture is an important goal of wheat breeding programs to improve varieties and increase yields. Recently, some key genes that control plant architecture and yield have been identified in wheat. *TaCol-B5*, which encodes a CONSTANS-like transcription factor, can modify spike architecture and enhance grain yield in wheat. Overexpression of *TaCol-B5* in wheat increases the numbers of spikelet nodes per spike, tillers, and spikes, thereby increasing grain yield by 11.9% under field conditions (Zhang et al., 2022a). Moreover, the miR319/TaGAMYB3 module also controls plant architecture and grain yield in wheat. Repressing miR319 or increasing *TaGAMYB3* expression can result in favorable plant architecture traits and enhance grain yield in field plot tests (Jian et al., 2022). Therefore, more important regulators of plant architecture and yield need to be identified and further studied in wheat.

In this study, we isolated and characterized three homoeologs of *TaSPL14* from wheat group 7 chromosomes. The expression patterns of *TaSPL14* homoeologs were investigated at various developmental stages in wheat. Overexpression of *TaSPL14-7A* in wheat led to a reduction in tillers as well as increases in kernel size and weight. We analyzed sequence variations in *TaSPL14-7A* among wheat accessions and performed an association analysis between phenotypes and haplotypes. *TaSPL14-7A-Hap1/2* was associated with multiple traits in modern cultivars, including lower tiller number, larger kernel size, and higher kernel weight. More importantly, *TaSPL14-7A-Hap1/2* underwent positive selection during global wheat breeding over the last century. Our findings provide evidence that *TaSPL14* is a conserved regulator of plant architecture and yield traits and provide a functional marker for molecular marker-assisted selection in wheat breeding.

Materials and methods

Plant materials and phenotype assessment

A total of 505 Chinese wheat accessions, including 157 landraces and 348 modern cultivars, were used for marker screening and association analyses (Liu et al., 2020). Agronomic trait data for these wheat accessions were collected from plants grown in three environments: Luoyang (112°26′E, 34°37′N), Henan Province, China, in 2002 and 2005, and Shunyi (116°65′W, 40°13′N), Beijing, China, in 2010. In addition, a total of 1,051 global wheat accessions, including 384 European, 480 North American, 83 former USSR, 53 CIMMYT, and 51 Australian modern wheat cultivars, were used to investigate the global distribution of *TaSPL14-7A* haplotypes. For spatiotemporal expression analysis, different tissue samples were collected from Chinese Spring (CS) wheat plants at different developmental stages. *TaSPL14* transgenic and wild-type wheat were grown in a greenhouse under long-day conditions (16 h light/8 h darkness) at 23°C. For each T_3-generation transgenic line, the phenotypes of at least 10 plants were analyzed.

RNA extraction and qRT-PCR analysis

Total RNA was extracted from various Chinese Spring tissues using Fruit-mate™ for RNA Purification (Takara Biomedical Technology, Beijing, China). cDNA was synthesized using the FastKing RT Kit (Tiangen Biotech) according to the

manufacturer's instructions. The coding sequences (CDSs) of *TaSPL14* homoeologs were cloned from Chinese Spring cDNA according to the Ensemblplants database (http://plants.ensembl.org/index.html). Genome-specific primers of *TaSPL14* homoeologs were designed using Primer Premier 5.0 software (http://www.premierbiosoft.com/) and validated by sequencing. Quantitative real-time PCR (qRT-PCR) assays were performed on a LightCycler 96 Real-Time PCR system (Roche Applied Science, Penzberg, Germany) using SYBR Premix Ex Taq (Takara Bio, Beijing, China). The wheat *Actin* gene was used as the internal reference (Liu et al., 2020). The relative expression of each gene was calculated according to the comparative CT method. All assays were performed three times in independent experiments. The primers used in this study were listed in Supplementary Table 1.

Subcellular localization

The full-length CDS of *TaSPL14-7A* was amplified from a Chinese Spring cDNA sample and subcloned into the pCAMBIA1300-GFP vector. The derived construct was transformed into *Agrobacterium tumefaciens* strain GV3101 and then co-infiltrated into tobacco (*Nicotiana benthamiana*) leaves. After a 48 h of incubation, the transformed tobacco leaves were stained with 4′,6-diamidino-2-phenylindole (DAPI), and then GFP and DAPI fluorescence were both observed using an LSM880 laser scanning confocal microscope (Carl Zeiss, Jena, Germany).

Vector construction and plant transformation

To construct the *TaSPL14-7A*-overexpression vector for wheat transformation, the full-length CDS of *TaSPL14-7A* was amplified and then cloned into a modified pCAMBIA3301 vector under the control of the maize *Ubiquitin* promoter as previously described (Zhang et al., 2022b). The resulting construct was mobilized into *Agrobacterium tumefaciens* strain EHA105 and then transformed into immature embryos of wheat cultivar Fielder *via* an Agrobacterium-mediated transformation method (Wang et al., 2017).

TaSPL14-7A haplotype discovery and association analysis

A total of 36 wheat accessions (13 modern cultivars and 23 landraces) were initially used for detection of *TaSPL14-7A* sequence variations (Supplementary Table 2). Specific primers were designed to amplify the 2.5-kb promoter region and the 4.5-kb coding region of *TaSPL14-7A*. The resulting DNA fragments were sequenced. Sequence variations, including single nucleotide polymorphisms (SNPs) and insertion/deletion (InDels), were identified using DNASTAR (http://www.dnastar.com/). The NEBcutter V3.0 (https://nc3.neb.com/NEBcutter/) was used to query the restriction enzyme recognition sites, and two cleaved amplified polymorphic site (CAPS) markers of the *TaSPL14-7A* haplotypes were developed based on two SNPs (T/C at 144 bp and G/A at 4111 bp). Haplotype association analysis was performed based on a previously described method (Hou et al., 2014; Ma et al., 2016). In brief, the variance analyses were carried out using SPSS 16.0 (IBM Corporation). Phenotypic differences among haplotypes were determined using one-way ANOVA and Tukey's test at a significance level of $P < 0.05$.

Differential expression and promoter activity analysis of *TaSPL14-7A* haplotypes

A total of 15 wheat varieties were used for differential expression analysis of three *TaSPL14-7A* haplotypes (Supplementary Table 3). Total RNA was extracted from 1-cm young spikes and stems at the elongation stage of these wheat varieties and subjected to qRT-PCR assays. For promoter activity assays, the 2.5-kb promoter sequence of each *TaSPL14-7A* haplotype was amplified and cloned into the pGreenII 0800-LUC vector (Hellens et al., 2005). As described previously (Liu et al., 2020), the promoter activity of each haplotype was measured using the Dual-Luciferase Reporter Assay System (Promega, USA), and calculated according to the relative luciferase activity (LUC/REN ratio).

Results

Cloning and expression analysis of *TaSPL14* homoeologs in wheat

To isolate *SPL14* homologous genes in wheat, the CDS of *OsSPL14* (Os08g0509600) was used as a query to search the Chinese Spring RefSeq v1.1 genome database (IWGSC, 2018). Three *SPL14* homoeologs located on chromosomes 7AS (*TraesCS7A02G246500*), 7BS (*TraesCS7B02G144900*), and 7DS (*TraesCS7D02G245200*) were obtained and designated *TaSPL14-7A*, *TaSPL14-7B*, and *TaSPL14-7D*, respectively. The phylogenetic analysis revealed that these three *TaSPL14* homoeologs have the closest homology to *OsSPL14* (Figure 1A). The genic regions of *TaSPL14-7A*, *TaSPL14-7B*, and *TaSPL14-7D* are 4,325, 3,817, and 3,863 bp in length, respectively, and each consists of 3 exons and 2 introns (Supplementary Figure 1). *TaSPL14-7A*, *TaSPL14-7B*, and *TaSPL14-7D* comprise 386, 386, and 385 amino acids, respectively, and share high sequence similarities with *OsSPL14* (~60%), especially in the conserved SBP domain (Figure 1B). These data suggest that *TaSPL14* homoeologs may have biological functions similar to those of *OsSPL14*.

To investigate the subcellular localization of *TaSPL14*, we transiently expressed the TaSPL14-7A-GFP (green fluorescent protein) fusion construct in tobacco (*Nicotiana benthamiana*) leaves. After 48 h of incubation and DAPI staining, the GFP and DAPI florescence signals were observed by laser confocal microscopy. As shown in Figure 2A, TaSPL14-7A-GFP fusion protein fluorescence was co-localized with DAPI-stained nucleus, whereas GFP fluorescence alone was present throughout the cell;

10.3389/fpls.2023.1178624

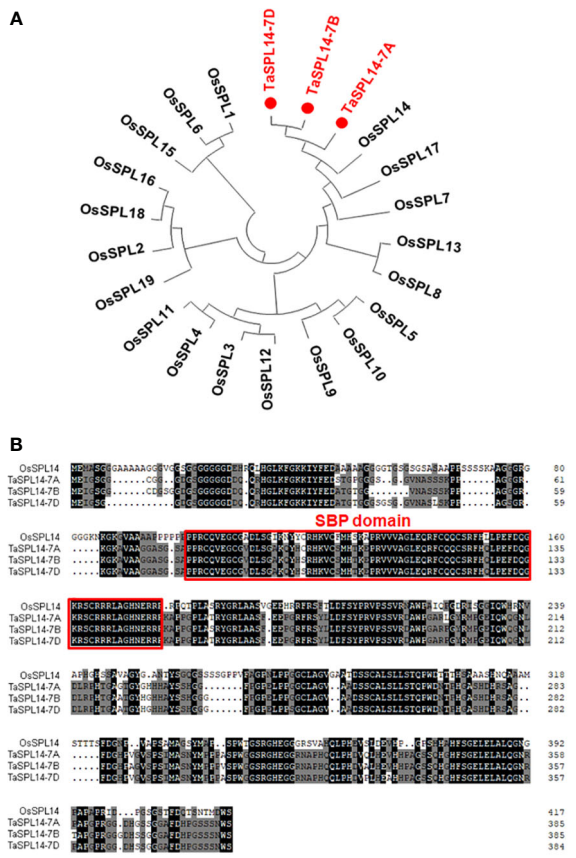


FIGURE 1
Phylogenetic tree and sequence alignment of TaSPL14 homoeologs. **(A)** Phylogenetic tree constructed from the 19 rice SPL proteins and three TaSPL14 homoeologs using the neighbor-joining method with a bootstrap value of 1000 by MEGA5.0. TaSPL14 homoeologs are marked in red. **(B)** Sequence alignment of OsSPL14 and TaSPL14 homoeologs. The conserved SBP domain is indicated by the red box.

therefore, *TaSPL14* is a nucleus-localized transcription factor similar to *OsSPL14*. Furthermore, genome-specific primers were used for qRT-PCR to explore the temporal and spatial expression of *TaSPL14* homoeologs. The results showed that the expression levels of *TaSPL14* homoeologs were significantly higher in stems at the elongation stage, and in young spikes (1 cm in length) and substantially lower in other tissues, including roots, leaves and grains (Figure 2B). Moreover, the expression level of *TaSPL14-7A* in all detected tissues was higher than that of *TaSPL14-7B* and *TaSPL14-7D*. Our results were consistent with the data in the wheat expression database (Supplementary Figure 2).

Overexpression of *TaSPL14-7A* leads to reduced tiller number and increased kernel weight and size in wheat

To explore the physiological function of *TaSPL14* in wheat, we generated *TaSPL14-7A* overexpression lines in the hexaploid wheat cultivar Fielder using Agrobacterium-mediated transformation. The total expression levels of *TaSPL14* were examined in the transgenic plants using qRT-PCR, and the results showed that the *TaSPL14-7A* expression was significantly increased in all three representative overexpression (OE) lines (Supplementary Figure 3). Compared with the wild-type (WT) control, all the OE lines showed dramatic changes in plant architecture and yield traits (Figures 3A, B). The tiller number of OE lines was approximately 47.5%-56.1% less than that of the WT (Figure 3C). In contrast, OE lines had larger kernel size and weight, and the thousand-kernel weight (TKW) was 4.86 g-5.29 g higher than that of the WT (Figure 3D). Notably, the increased TKW in the transgenic plants mainly resulted from a 0.7%-2.2% increase in kernel length (KL) and a 6.4%-7.7% increase in kernel width (KW) (Figures 3E, F). Therefore, our phenotypic data clearly indicated that *TaSPL14* is a conserved pleiotropic regulator of plant architecture and yield traits, consistent with the phenotypic effects of *OsSPL14* (Jiao et al., 2010).

TaSPL14-7A haplotypes are associated with tiller number and kernel weight and size

To detect natural variations in *TaSPL14* homoeologs, we re-sequenced the genome regions of 36 highly diversified wheat varieties, and the results showed that there were polymorphic loci in the genome regions of *TaSPL14-7A* and *TaSPL14-7B*, while no sequence differences were detected in *TaSPL14-7D*. Because *TaSPL14-7A* has the highest expression among *TaSPL14* homoeologs (Figure 2B), we developed two molecular markers, CAPS-144 and CAPS-4111, to distinguish the three *TaSPL14-7A* haplotypes (*TaSPL14-7A-Hap1/2/3*). CAPS-144 was developed based on the SNP (T/C) at the 144-bp position, which produced the recognition site of the restriction enzyme MseI at *TaSPL14-7A-Hap3*, but did not identify corresponding sites at *TaSPL14-7A-Hap1/2* (Figure 4). After two-step PCR amplification and enzyme digestion, the amplified fragments of *TaSPL14-7A-Hap3* were digested and separated (Figure 4). CAPS-4111 was developed based on the SNP (G/A) at the 4111-bp position, which produced the recognition sites of the restriction enzyme BglI at *TaSPL14-7A-Hap2/3* but did not identify the corresponding site at *TaSPL14-7A-Hap1* (Figure 4). After two-step PCR amplification and enzyme digestion, the amplified fragments of *TaSPL4-7A-Hap2/3* was digested and separated (Figure 4).

By using the newly developed molecular markers CAPS-144 and CAPS-4111, we genotyped 348 modern cultivars from Chinese wheat core collection (Supplementary Table 4). We then performed an association analysis between the *TaSPL14-7A* haplotypes and multiple agronomic traits measured in three environments (Luoyang, 2002; Luoyang, 2005; Shunyi, 2010). Significant differences in effective tiller number (ETN) were detected between *TaSPL14-7A-Hap1/2* and *TaSPL14-7A-Hap3* (Figure 5A; Supplementary Table 5). Moreover, the *TaSPL14-7A* haplotypes were significantly correlated with TKW, KL, and KW. The mean TKW of *TaSPL14-7A-Hap1/2* was 4.6 g-5.3 g higher than that of *TaSPL14-7A-Hap3* in all three environments (Figure 5B; Supplementary Table 5). Consistently, the mean KL and KW of *TaSPL14-7A-Hap1/2* were significantly higher than those of *TaSPL14-7A-Hap3* (Figures 5C, D; Supplementary Table 5).

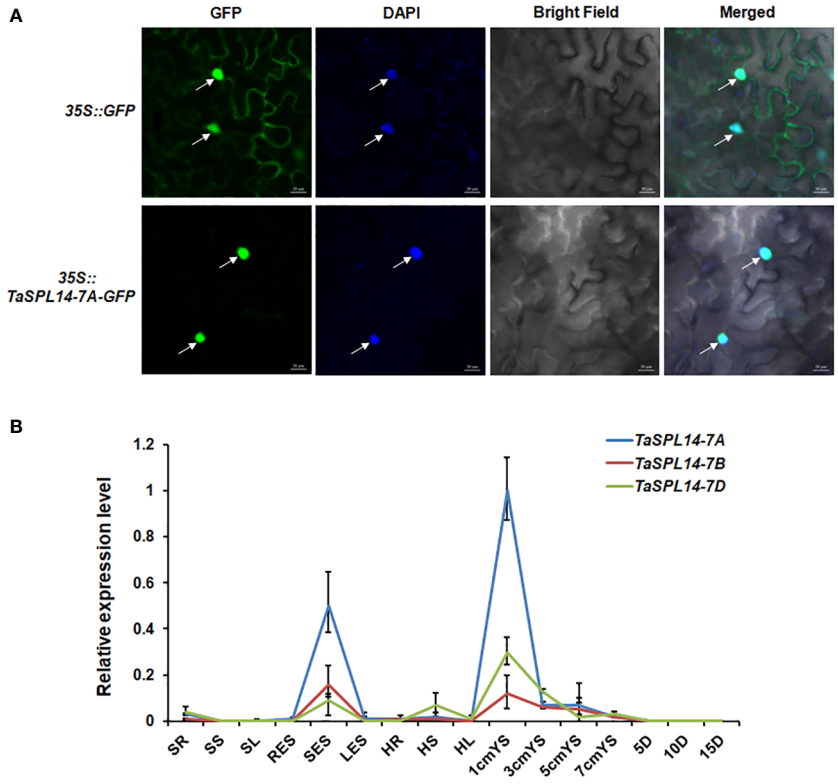


FIGURE 2
Characterization of TaSPL14 homoeologs. **(A)** Subcellular localization of TaSPL14-7A in tobacco leaves. GFP and TaSPL14-7A-GFP fusions under the control of the CaMV 35S promoter were transiently expressed in tobacco leaves. The arrows indicate nuclei stained with DAPI. Bar = 20 μm. **(B)** Spatial and temporal expression patterns of the three *TaSPL14* homoeologs in various Chinese Spring tissues. SR, seedling roots; SS, seedling stems; SL, seedling leaves; RES, roots at elongation stage; SES, stems at elongation stage; LES, leaves at elongation stage; HR, roots at heading stage; HS, stems at heading stage; HL, leaves at heading stage; 1cm–7cmYS, 1–7-cm young spikes; 5D–15D, grains at 5–15 days post-anthesis, respectively. Normalized values of *TaSPL14* expression relative to *Actin* were given as mean ± SD from three replicates.

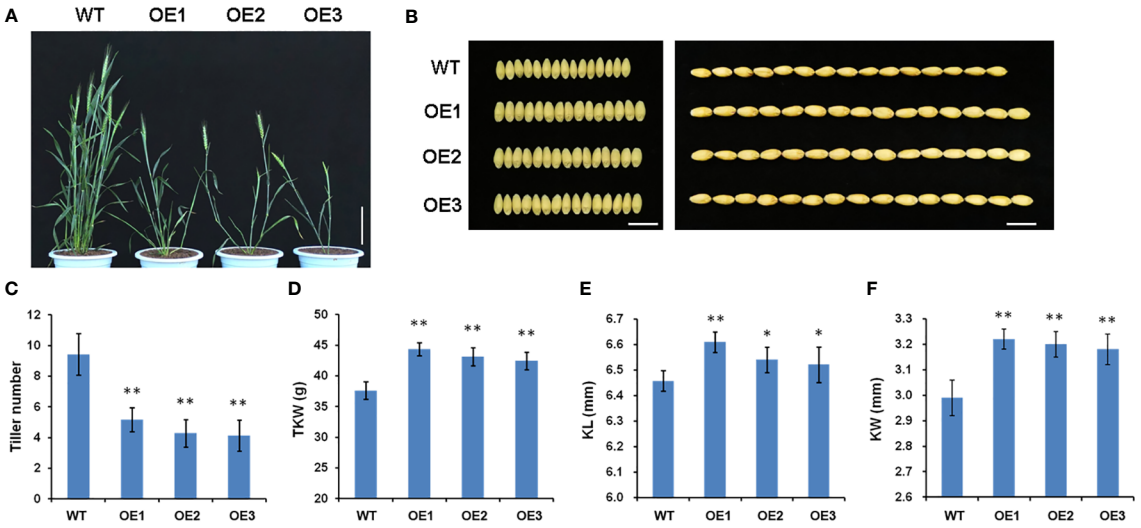


FIGURE 3
Phenotypic comparison of wild-type (WT) and three *TaSPL14-7A* overexpression (OE) lines. **(A)** Comparison of the plant architecture phenotype at heading stage among these genotypes. Bar = 20 cm. **(B)** Comparison of the kernel phenotypes among these genotypes. Bar = 1cm. **(C)** Tiller number; **(D)** Thousand kernel number (TKW); **(E)** Kernel length (KL); **(F)** Kernel width (KW). The values are presented as mean ± SD. $^{**}P<0.01$; $^{*}P<0.05$ (Student's *t*-test).

10.3389/fpls.2023.1178624

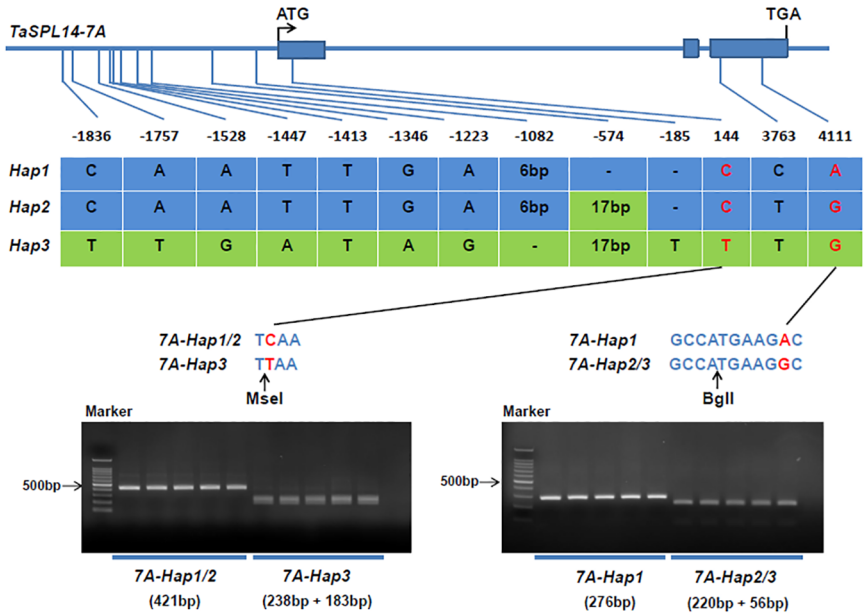


FIGURE 4
Identification of *TaSPL14-7A* haplotypes and development of molecular markers. The top panel shows a schematic diagram of the gene and 2.5 kb promoter structure of *TaSPL14-7A*. The ATG start codon was designated as position 1 bp. The bottom panel shows the polymorphic sites of *TaSPL14-7A*. A CAPS marker named CAPS-144 was developed based on the SNP (T/C) at 144 bp. Digestion of the amplified 421 bp fragment with MseI produced fragments of 238 bp and 183 bp for *TaSPL14-7A-Hap3* (T), whereas this fragment could not be digested for *TaSPL14-7A-Hap1/2* (C). A CAPS marker named CAPS-4111 was developed based on the SNP (G/A) at 4111 bp. Digestion of the amplified 276 bp fragment with BglI produced fragments of 220 bp and 56 bp for *TaSPL14-7A-Hap2/3* (G), whereas this fragment could not be digested for *TaSPL14-7A-Hap1* (A).

These results indicated that *TaSPL14-7A-Hap1/2* are favorable haplotypes associated with higher TKW and kernel size (KL and KW).

Significant expression differences of *TaSPL14-7A* exist among haplotypes

Most variations between *TaSPL14-7A-Hap1/2* and *TaSPL14-7A-Hap3* were located in the promoter regions and caused some changes in binding sites of important transcription factors such as WRKY, MYB and GRF (Supplementary Figure 4A). To test whether variations in the promoter region resulted in differential expression, we detected the expression of *TaSPL14-7A* in 1-cm young spikes and stems at the elongation stage in 15 wheat varieties with different haplotypes (Supplementary Table 3). As shown in Figure 6, *TaSPL14-7A* expression was significantly higher in the *TaSPL14-7A-Hap1/2* accessions than in the *TaSPL14-7A-Hap3* accessions in both 1-cm young spikes and stems, consistent with the hypothesis that *TaSPL14-7A-Hap1/2* with higher *TaSPL14-7A* expression is associated with lower tiller number and higher kernel weight and size. In addition, a minor difference in *TaSPL14-7A* expression was also observed between *TaSPL14-7A-Hap1* and *TaSPL14-7A-Hap2*, where *TaSPL14-7A-Hap2* accessions had the highest expression in 1-cm young spikes (Figure 6A), while *TaSPL14-7A-Hap1* accessions had the highest expression in stems (Figure 6B). In addition, we carried out promoter activity analysis through transient expression assay in *Nicotiana benthamiana* leaves. As shown in Supplementary Figure 4B, relative LUC activity (LUC/REN) of *TaSPL14-7A-Hap1* or *TaSPL14-7A-Hap2* promoter was about 2-fold greater than that of *TaSPL14-7A-Hap3* promoter, confirming that *TaSPL14-7A-Hap1/2* had higher promoter activity. These results suggest that *TaSPL14-7A-Hap1/2* are correlated with significantly increased expression levels of *TaSPL14-7A* due to their higher promoter activity.

Geographic distribution and frequency change of *TaSPL14-7A* haplotypes in global wheat breeding

Previous studies have shown that favorable haplotypes can be positively selected and accumulate during the wheat breeding (Paux et al., 2012). To determine the geographic distribution and artificial selection of different *TaSPL14-7A* haplotypes, we genotyped 157 landraces and 348 modern cultivars from ten wheat agro-ecological zones in China. From landraces to modern cultivars, the proportions of *TaSPL14-7A-Hap1/2* were significantly increased in the major production zones (I-IV), indicating a strong positive selection in the wheat breeding process (Figures 7A, B). From the 1940s to 1990s, the frequency of favorable *TaSPL14-7A-Hap2* increased from 12.5% to 68.5%, whereas the frequency of *TaSPL14-7A-Hap3* gradually declined from 49.5% to 4.5% (Figure 7C). Therefore, favorable *TaSPL14-7A* haplotypes were subjected to strong selection pressure and became the dominant haplotypes in the modern cultivars.

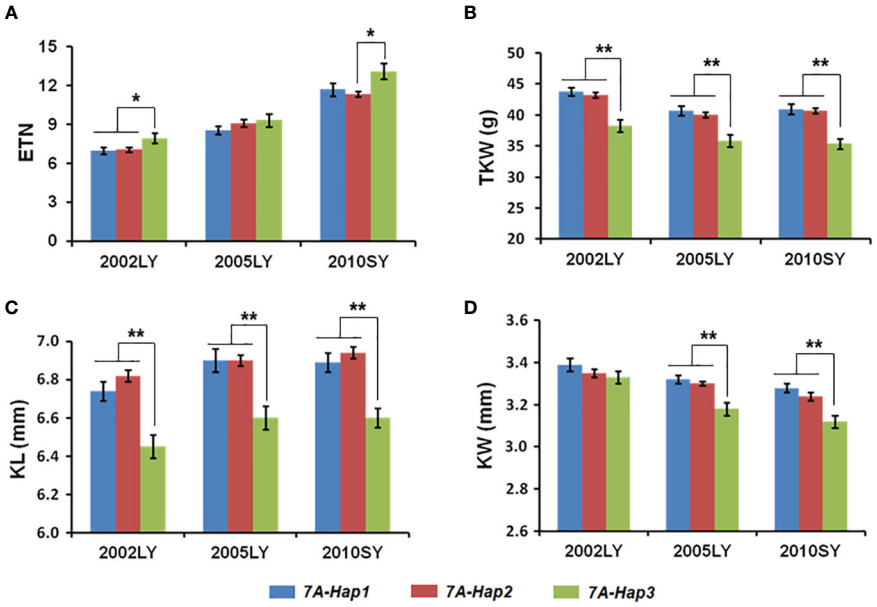


FIGURE 5
Association analysis of *TaSPL14-7A* haplotypes with tiller number and kernel traits in modern cultivars in three different environments. **(A)** Effective tiller number (ETN); **(B)** Thousand kernel weight (TKW); **(C)** Kernel length (KL); **(D)** Kernel width (KW) of modern cultivars with different *TaSPL14-7A* haplotypes. The x-axis represents different environments: Luoyang, 2002 (2002LY); Luoyang, 2005 (2005LY); Shunyi, 2010 (2010SY). The values are presented as mean ± SE. ***P*<0.01; **P*<0.05 (ANOVA).

Next, we performed a genotyping analysis of 1,051 modern cultivars from Europe, North America, the former USSR, CIMMYT, and Australia to assess the global geographic distribution pattern of the *TaSPL14-7A* haplotypes (Supplementary Table 6). Similar to the selection trend in China, the proportions of *TaSPL14-7A-Hap1/2* were also very high in the other five regions; especially in Australia, CIMMYT, Europe, and North America (Figure 7D). In these regions, the frequency of *TaSPL14-7A-Hap3* was very low, with the haplotype nearly eliminated. These results provide strong evidence that *TaSPL14-7A-Hap1/2* underwent positive selection as part of global wheat breeding.

Discussion

The SPL gene has a highly conserved DNA-binding domain (SBP domain), which represents a family of plant-specific transcription factors and plays a key role in regulating plant development in various plant species, such as *Arabidopsis*, rice, and maize (Xie et al., 2006; Hultquist and Dorweiler, 2008; Yang et al., 2008). *IPA1* (*OsSPL14*) is a core regulator of plant architecture and yield, having pleiotropic effects on multiple important yield traits, including tiller number, panicle primary branch number and TKW (Jiao et al., 2010). In the present study, we isolated *OsSPL14* homologues in wheat through a bioinformatics approach. *TaSPL14* was identified as a close

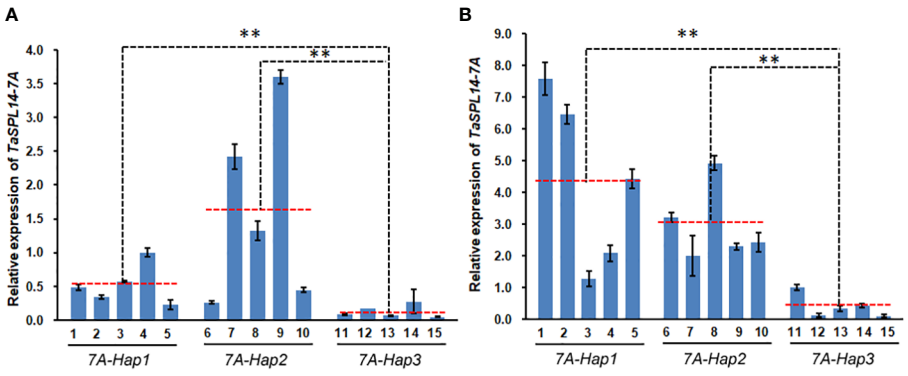


FIGURE 6
Differential expression of *TaSPL14-7A* in different haplotypes. **(A)** *TaSPL14-7A* expression was compared in the 1 cm young spikes. **(B)** *TaSPL14-7A* expression was compared in the stems at elongation stage. The red-dashed lines indicate the average expression levels of each haplotype accession. 1, Huoliaomai; 2, Dahongmai; 3, Baiyoumai; 4, Yangmai; 5, Dongnong 101; 6, Neimai 11; 7, Xiaohongpi; 8, Bihongsui; 9, Ganmai 46; 10, Jinchun 3; 11, Ganmai 6; 12, Xiangnong 3; 13, Xiaobaimai; 14, Xinkehan 9; 15, Lianglaiyou. ***P*<0.01 (ANOVA).

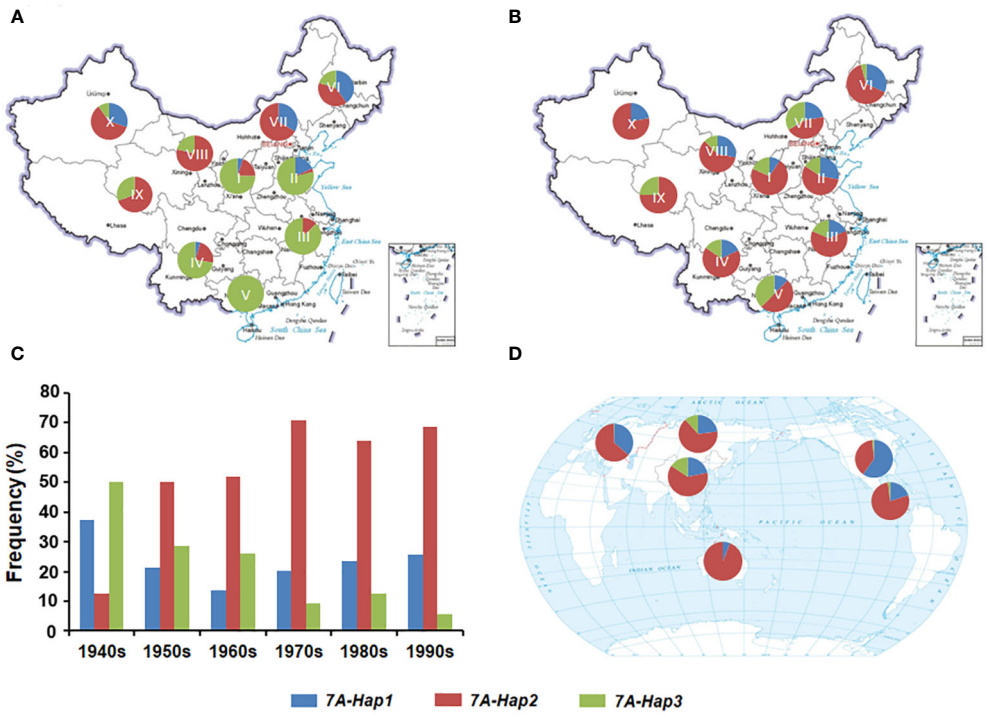


FIGURE 7
Geographic distribution and frequency change of *TaSPL14-7A* haplotypes. **(A)** Distribution of *TaSPL14-7A* haplotypes in Chinese landraces from ten agro-ecological zones (I, northern winter wheat region; II, Yellow and Huai River valley winter wheat region; III, low and middle Yangtze River valley winter wheat region; IV, southwestern winter wheat region; V, southern winter wheat region; VI, northeastern spring wheat region; VII, northern spring wheat region; VIII, northwestern spring wheat region; IX, Qinghai-Tibet spring-winter wheat region; X, Xinjiang winter-spring wheat region). **(B)** Distribution of *TaSPL14-7A* haplotypes in modern Chinese cultivars from ten agro-ecological zones. **(C)** Frequency changes of *TaSPL14-7A* haplotypes in different Chinese wheat breeding periods (from the 1940s to the 1990s). **(D)** Distribution of *TaSPL14-7A* haplotypes in major global wheat regions including China, North America, Europe, former USSR, CIMMYT, and Australia.

homologue of *OsSPL14*, and its homoeologs are located on chromosomes 7A, 7B, and 7D (Figure 1A). *TaSPL14* shares high sequence similarities with *OsSPL14* (~60%), particularly in the SBP protein domain (Figure 1B), suggesting that they have similar protein structures and functions. Furthermore, overexpression of *TaSPL14-7A* in common wheat cultivars resulted in significant changes in plant architecture and yield traits, such as a smaller tiller number and increased kernel weight and size (Figure 3), showing a phenotype similar to that of *IPA1* (*OsSPL14*) in rice (Jiao et al., 2010). Therefore, *TaSPL14* is a conserved negative regulator of tiller number, as well as a positive regulator of kernel weight and size. It should be noted that constitutive overexpression of *TaSPL14* cannot lead to an increase in grain yield because *TaSPL14* has opposite effects on tiller number and TKW, two key components of grain yield. Previous studies also showed that *IPA1* (*OsSPL14*) in rice regulates tiller number, stem diameter, and panicle primary branch number in a dose-dependent manner, but overexpression has a negative effect on grain yield (Zhang et al., 2017). These findings offer a realistic solution for achieving high-yield goals by optimizing the expression levels of *TaSPL14*. It is possible to balance tiller number and TKW by fine-tuning the tissue-specific expression of *TaSPL14*, e.g., weak alleles for the up-regulation of *TaSPL14* expression or gene editing of its cis-regulatory region.

In crops, variations in the coding regions account for only a small fraction of the large number of natural variations, many of which are deleterious and will be eliminated during evolution and selection due to their large impacts on protein functions. Unlike deleterious variations, most natural variations with phenotypic benefits are located in regulatory regions such as promoters and introns. In this study, we detected natural variations in *TaSPL14-7A* and found that most variations were located in the promoter region, including SNPs and InDels (Figure 4). Although three exonic SNPs were identified in the coding region, they did not cause obvious genetic effects according to haplotype-based association analysis results (Figure 5). In contrast, variations in the promoter region could result in differential expression of *TASPL14-7A* and significant alteration of promoter activity between *TaSPL14-7A-Hap1/2* and *TaSPL14-7A-Hap3*, providing supporting evidence for phenotypic differences and genetic effects of *TaSPL14-7A* haplotypes (Figure 6; Supplementary Figure 4). Similar regulatory patterns have been observed in many other yield-related genes in wheat, including *TaGW2-6A*, *TaGW2-6B*, *TaBT1-6B*, and *TaDA1-2A* (Su et al., 2011; Qin et al., 2014; Wang et al., 2019; Liu et al., 2020). Therefore, our present and previous studies supported that variations in regulatory regions are conducive to the rapid adaptation of crops and can allow gene expression to adapt to environmental changes. Therefore, these variations located in promoter regions can be used as potential target sites for wheat breeding and selection in the future.

Marker-assisted selection (MAS) is an efficient method for crop genetic improvement. The accumulation of favorable alleles or

haplotypes is mainly involved in wheat breeding, so MAS can be used to significantly accelerate the breeding process (Gupta et al., 2023). In this study, we developed two CAPS markers to distinguish three *TaSPL14-7A* haplotypes, and the association analysis indicated that the *TaSPL14-7A* haplotypes were significantly correlated with tiller number, TKW, kernel length, and kernel width in modern cultivars (Figure 5), among which *TaSPL14-7A-Hap1/2* were favorable haplotypes correlated with a smaller tiller number, higher TKW, and larger kernel size. According to the breeding process of three haplotypes in 10 different agro-ecological zones of China, *TaSPL14-7A-Hap2* was subjected to strong positive selection from landraces to modern cultivars (Figures 7A, B). Moreover, the frequency of *TaSPL14-7A-Hap2* increased significantly from the 1940s to 1990s in China (Figure 7C). As favorable haplotypes, *TaSPL14-7A-Hap1/2* had become dominant in the global modern cultivars (Figure 7D). Interestingly, *TaSPL14-7A-Hap1/2* have a strong positive effect on TKW but a weak negative effect on tiller number based on haplotype association analysis (Figure 5), possibly representing a superior plant architecture in modern wheat cultivar improvement. Balancing the tiller number and TKW is critical for achieving higher grain yield potential in wheat breeding. Therefore, *TaSPL14-7A-Hap1/2* and their functional markers will be applied in future wheat yield improvement.

Conclusion

In this study, we characterized *TaSPL14-7A* as a negative regulator of tiller number and a positive regulator of kernel weight and size in wheat. Haplotype analysis indicated that the favorable *TaSPL14-7A* haplotypes were associated with superior plant architecture and yield traits and underwent positive selection during global wheat breeding. Our findings provide insight into the genetic effects of *TaSPL14* and its potential application in wheat improvement.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

TL, FW and CH planned and designed the research. LC, TL, YZ, and YP performed the experiments. TL, SG, XZ, FW and CH collected and analyzed the data. LC, TL, and CH wrote the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (32272081 and 32172045).

Acknowledgments

The authors thank Mr. Kang Zhang (Genovo Biotechnology Co. Ltd., Tianjin, China) for his help with wheat transformation.

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.

The handling editor YF declared a shared affiliation with the authors TL, SG, YZ, YP, XZ, CH at the time of review.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1178624/full#supplementary-material

References

Birkenbihl, R. P., Jach, G., Saedler, H., and Huijser, P. (2005). Functional dissection of the plant specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. *J. Mol. Biol.* 352, 585–596. doi: 10.1016/j.jmb.2005.07.013

Gandikota, M., Birkenbihl, R. P., Höhmann, S., Cardon, G. H., Saedler, H., and Huijser, P. (2007). The miRNA156/157 recognition element in the 3' UTR of the arabidopsis SBP box gene *SPL3* prevents early flowering by translational inhibition in seedlings. *Plant J.* 49, 683–693. doi: 10.1111/j.1365-313X.2006.02983.x

Gupta, A., Hua, L., Zhang, Z., Yang, B., and Li, W. (2023). CRISPR-induced miRNA156-recognition element mutations in *TaSPL13* improve multiple agronomic traits in wheat. *Plant Biotechnol. J.* 21, 536–548. doi: 10.1111/pbi.13969

Hang, H., Bauer, M., Mio, W., and Mander, L. (2021). Geometric and topological approaches to shape variation in ginkgo leaves. *R. Soc. Open Sci.* 8, 210978. doi: 10.1098/rsos.210978

Hellens, R. P., Allan, A. C., Friel, E. N., Bolitho, K., Grafton, K., Templeton, M. D., et al. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* 1, 13. doi: 10.1186/1746-4811-1-13

Hou, J., Jiang, Q., Hao, C., Wang, Y., Zhang, H., and Zhang, X. (2014). Global selection on sucrose synthase haplotypes during a century of wheat breeding. *Plant Physiol.* 164, 1918–1929. doi: 10.1104/pp.113.232454

Hultquist, J. F., and Dorweiler, J. E. (2008). Feminized tassels of maize *mop1* and *ts1* mutants exhibit altered levels of miR156 and specific SBP-box genes. *Planta.* 229, 99–113. doi: 10.1007/s00425-008-0813-2

International Wheat Genome Sequencing Consortium (IWGSC) (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191. doi: 10.1126/science.aar7191

Jian, C., Hao, P., Hao, C., Liu, S., Mao, H., Song, Q., et al. (2022). The miR319/TaGAMYB3 module regulates plant architecture and improves grain yield in common wheat (*Triticum aestivum* l). *New Phytol.* 235, 1515–1530. doi: 10.1111/nph.18216

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* 42, 541–544. doi: 10.1038/ng.591

Liu, H., Li, H., Hao, C., Wang, K., Wang, Y., Qin, L., et al. (2020). *TaDA1*, a conserved negative regulator of kernel size, has an additive effect with *TaGW2* in common wheat (*Triticum aestivum* l.). *Plant Biotechnol. J.* 18, 1330–1342. doi: 10.1111/pbi.13298

Ma, L., Li, T., Hao, C., Wang, Y., Chen, X., and Zhang, X. (2016). *TaGS5-3A*, a grain size gene selected during wheat improvement for larger kernel and yield. *Plant Biotechnol. J.* 14, 1269–1280. doi: 10.1111/pbi.12492

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., et al. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* 38, 948–952. doi: 10.1038/ng1841

Paux, E., Sourdille, P., Mackay, I., and Feuillet, C. (2012). Sequence-based marker development in wheat: Advances and applications to breeding. *Biotechnol. Adv.* 30, 1071–1088. doi: 10.1016/j.biotechadv.2011.09.015

Preston, J. C., and Hileman, L. C. (2013). Functional evolution in the plant *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE* (*SPL*) gene family. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00080

Qin, L., Hao, C., Hou, J., Wang, Y., Li, T., Wang, L., et al. (2014). Homologous haplotypes, expression, genetic effects and geographic distribution of the wheat yield gene *TaGW2*. *BMC Plant Boil.* 14, 107. doi: 10.1186/1471-2229-14-107

Riese, M., Zobell, O., Saedler, H., and Huijser, P. (2008). SBP-domain transcription factors as possible effectors of cryptochrome-mediated blue light signalling in the moss physcomitrella patens. *Planta.* 227, 505–515. doi: 10.1007/s00425-007-0661-5

Shikata, M., Koyama, T., Mitsuda, N., and Ohme-Takagi, M. (2009). Arabidopsis SBP-box genes *SPL10*, *SPL11* and *SPL2* control morphological change in association with shoot maturation in the reproductive phase. *Plant Cell Physiol.* 50, 2133–2145. doi: 10.1093/pcp/pcp148

Song, X., Meng, X., Guo, H., Cheng, Q., Jing, Y., Chen, M., et al. (2022). Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. *Nat. Biotechnol.* 40, 1403–1411. doi: 10.1038/s41587-022-01281-7

Stone, J. M., Liang, X., Nekl, E. R., and Stiers, J. J. (2005). Arabidopsis *AtSPL14*, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. *Plant J.* 41, 744–754. doi: 10.1111/j.1365-313X.2005.02334.x

Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum* l.). *Theor. Appl. Genet.* 122, 211–223. doi: 10.1007/s00122-010-1437-z

Wang, K., Liu, H., Du, L., and Ye, X. (2017). Generation of marker-free transgenic hexaploid wheat *via* an Agrobacterium-mediated co-transformation strategy in commercial Chinese wheat varieties. *Plant Biotechnol. J.* 15, 614–623. doi: 0.1111/pbi.12660

Wang, Y., Hou, J., Liu, H., Li, T., Wang, K., Hao, C., et al. (2019). *TaBT1*, affecting starch synthesis and thousand kernel weight, underwent strong selection during wheat improvement. *J. Exp. Bot.* 70, 1497–1511. doi: 10.1093/jxb/erz032

Wang, Y., Hu, Z., Yang, Y., Chen, X., and Chen, G. (2009). Function annotation of an SBP-box gene in arabidopsis based on analysis of co-expression networks and promoters. *Int. J. Mol. Sci.* 10, 116–132. doi: 10.3390/ijms10010116

Wang, Y., and Li, J. (2008). Molecular basis of plant architecture. *Annu. Rev. Plant Bio.* 59, 253–279. doi: 10.1146/annurev.arplant.59.032607.092902

Wang, B., Smith, S. M., and Li, J. (2018). Genetic regulation of shoot architecture. *Annu. Rev. Plant Biol.* 69, 437–468. doi: 10.1146/annurev-arplant-042817-040422

Wang, H., and Wang, H. (2015). The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits. *Mol. Plant* 8, 677–688. doi: 10.1016/j.molp.2015.01.008

Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., et al. (2012). Control of grain size, shape and quality by *OsSPL16* in rice. *Nat. Genet.* 44, 950–954. doi: 10.1038/ng.2327

Xie, K., Wu, C., and Xiong, L. (2006). Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol.* 142, 280–293. doi: 10.1104/pp.106.084475

Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., et al. (2004). A novel zinc-binding motif revealed by solution structures of DNA-binding domains of arabidopsis SBP-family transcription factors. *J. Mol. Biol.* 337, 49–63. doi: 10.1016/j.jmb.2004.01.015

Yang, Z., Wang, X., Gu, S., Hu, Z., Xu, H., and Xu, C. (2008). Comparative study of SBP-box gene family in arabidopsis and rice. *Gene.* 407, 1–11. doi: 10.1016/j.gene.2007.02.034

Zhang, L., Yu, H., Ma, B., Liu, G., Wang, J., Wang, J., et al. (2017). A natural tandem array alleviates epigenetic repression of *IPA1* and leads to superior yielding rice. *Nat. Commun.* 8, 14789. doi: 10.1038/ncomms14789

Zhang, D., Zhang, X., Xu, W., Hu, T., Ma, J., Zhang, Y., et al. (2022a). TaGW2L, a GW2-like RING finger E3 ligase, positively regulates heading date in common wheat (*Triticum aestivum* l.). *Crop J.* 10, 972–979. doi: 10.1016/j.cj.2021.12.002

Zhang, X., Jia, H., Li, T., Wu, J., Nagarajan, R., Lei, L., et al. (2022b). *TaCol-B5* modifies spike architecture and enhances grain yield in wheat. *Science.* 376, 180–183. doi: 10.1126/science.abm0717

Zhu, T., Liu, Y., Ma, L., Wang, X., Zhang, D., Han, Y., et al. (2020). Genome-wide identification, phylogeny and expression analysis of the *SPL* gene family in wheat. *BMC Plant Biol.* 20, 420. doi: 10.1186/s12870-020-02576-0

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 April 2023
DOI 10.3389/fpls.2023.1184276

OPEN ACCESS

EDITED BY
Yue Feng,
China National Rice Research Institute (CAAS), China

REVIEWED BY
Jw Wu,
South China Agricultural University, China
Youxin Yang,
Jiangxi Agricultural University, China
Hao Chen,
Fujian Academy of Agricultural Sciences, China

*CORRESPONDENCE
Shaohong Zhang
szhanggz@tom.com
Wu Yang
yangwu@gdaas.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 11 March 2023
ACCEPTED 27 March 2023
PUBLISHED 14 April 2023

CITATION
Huo X, Wang J, Chen L, Fu H, Yang T, Dong J, Ma Y, Zhou L, Chen J, Liu D, Liu B, Zhao J, Zhang S and Yang W (2023) Genome-wide association mapping and gene expression analysis reveal candidate genes for grain chalkiness in rice. *Front. Plant Sci.* 14:1184276. doi: 10.3389/fpls.2023.1184276

Genome-wide association mapping and gene expression analysis reveal candidate genes for grain chalkiness in rice

Xing Huo†, Jian Wang†, Luo Chen, Hua Fu, Tifeng Yang, Jingfang Dong, Yamei Ma, Lian Zhou, Jiansong Chen, Dilin Liu, Bin Liu, Junliang Zhao, Shaohong Zhang* and Wu Yang*

Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangdong Key Laboratory of New Technology in Rice Breeding, Guangdong Rice Engineering Laboratory, Key Laboratory of Genetics and Breeding of High Quality Rice in Southern China (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Guangzhou, China

Grain chalkiness is the main factor determining the market value of rice. Reducing chalkiness is an important breeding goal for genetic improvement of high quality rice. Identification of QTLs or genes controlling chalkiness is the prerequisite for molecular breeding in rice. Here, we conducted a genome-wide association study to identify QTLs associated with grain chalkiness including percentage of grains with chalkiness (PGWC) and degree of endosperm chalkiness (DEC) in 450 rice accessions consisting of 300 *indica* and 150 *japonica* rice in two environments. A total of 34 QTLs were identified, including 14 QTLs for PGWC and 20 QTLs for DEC. Among them, seven QTLs were commonly identified in two environments, and eight QTLs were simultaneously related to two traits. Based on the haplotype analysis, LD decay analysis, RNA-sequencing, qRT-PCR confirmation and haplotype comparisons, four genes (*LOC_Os10g36170*, *LOC_Os10g36260*, *LOC_Os10g36340* and *LOC_Os10g36610*) were considered as the candidate genes for *qDEC-10c*^{1w,2wj}, which could be identified in both environments and had the most significant *p*-value among the newly identified QTLs. These results provided new insight into the genetic basis of grain chalkiness and gene resources for improving quality by molecular breeding in rice.

KEYWORDS

rice, grain chalkiness, genome-wide association study, quantitative trait loci (QTL), candidate gene

1 Introduction

Rice is one of the most important food crops in the world. Improving rice yield and quality is the main goal of rice breeding. At present, with the improvement of people's living standard, high quality rice is the primary concern of rice breeders and consumers. Rice quality mainly includes milling, appearance and cooking quality. Grain chalkiness, as

an important character of appearance quality, is highly undesirable in rice breeding and marketing (Zhao et al., 2022). Percentage of grains with chalkiness (PGWC) and degree of endosperm chalkiness (DEC) are two indexes to evaluate grain chalkiness. Rice with high PGWC and DEC is easily broken during milling, and thus reducing the yield of head milled rice. Furthermore, the taste of chalky grain declines after cooking (Wu et al., 2022). Therefore, reducing chalkiness has been the key trait selected for released rice cultivars.

Understanding the genetic basis of grain chalkiness can greatly improve breeding efficiency. Grain PGWC and DEC are quantitatively inherited, controlled by multiple QTLs or genes, and affected by environments (Wang et al., 2017; Qiu et al., 2021). In past decade, bi-parental populations, rice mutants and reverse genetics have been widely used in QTL mapping and gene cloning for the two traits, and many QTLs have been reported and some genes have been cloned. *Chalk5* is cloned from natural variants and encodes a vacuolar H^+-translocating pyrophosphatase. Elevated expression of *Chalk5* increases the endosperm chalkiness (Li et al., 2014). Recently, *WCR1* was identified by map-based cloning and up-regulated expression of *WCR1* significantly reduces white-core rate (Wu et al., 2022). Besides, some chalkiness-related genes are cloned from floury endosperm mutants, most of which are related to starch synthesis or starch quality. The *starch synthase IIIa* (*SSIIIa*) regulates short amylopectin elongation, and knockout of *SSIIIa* causes the formation of white core (Ryoo et al., 2007). The *GIF1* encodes a cell-wall invertase required for carbon partitioning during early grain-filling, and the *gif1* mutant shows loose starch grain structure and increased chalkiness (Wang et al., 2008). *OsLTPL36* is specially expressed in developing rice seed, and suppressed expression of this gene increases grain chalkiness rate (Wang et al., 2015). Fan et al. found that reduced expression of an autophagy gene *OsATG8b* results in chalky endosperm and poor seed quality (Fan et al., 2020).

In recent years, genome-wide association study (GWAS) based on linkage disequilibrium (LD) in a large germplasm resources has been widely used to identify QTLs that control complex traits, such as appearance quality in rice, on a large scale (Qiu et al., 2015; Wang et al., 2017; Qiu et al., 2021). Using a worldwide collection of *indica* rice germplasm, Qiu et al. identified two QTLs for PGWC and five QTLs for DEC, and two QTLs, *qPGWC5.1* and *qDEC5.2* on chromosome 5 were co-located with *GW5* (Qiu et al., 2015). Furthermore, they identified 12 QTLs for DEC in two years by using a larger panel mainly including *indica* and *japonica* from the 3K Rice Genomes Project, but none could be detected in both years (Qiu et al., 2021). Wang et al. detected four QTLs affecting DEC and four QTLs affecting PGWC, and haplotype analysis suggested that *Os07g0604500* was the candidate gene underlying *qDEC7* (Wang et al., 2017). Recently, *OsbZIP60* was identified as a vital regulator of grain chalkiness by GWAS, and knockout of *OsbZIP60* results in high grain chalkiness through regulating the expression of key genes related to grain chalkiness (Yang et al., 2022a). Although progress has been made in the identification and molecular mechanism of QTLs or genes associated with grain chalkiness in rice, but few of them have been applied to rice breeding. Therefore, it is necessary to further explore the stable QTLs or genes related to grain chalkiness in rice.

In the present study, PGWC and DEC in a diverse panel containing 450 accessions selected from the RDP2 (McCouch et al., 2016) were evaluated. GWAS was conducted to identify QTLs for the two traits. In total, 34 QTLs were identified including 11 QTLs co-localized with the reported QTLs or genes and 23 newly identified QTLs. Based on the linkage disequilibrium (LD) decay analysis, RNA-sequencing, qRT-PCR confirmation and haplotype comparisons, four genes were considered as candidate genes underlying $qDEC\text{-}10c^{1w,2wj}$, which could be identified in both environments and had the most significant *p*-value among the newly identified QTLs. This study provides new insight into the genetic basis of grain chalkiness and contributes to molecular breeding for high-quality rice.

2 Materials and methods

2.1 Plant materials

The 450 rice accessions (Table S1) were used for phenotype evaluation and GWAS in this study. These rice accessions were selected from RDP2 (McCouch et al., 2016) based on their origins and diversity, including 300 *indica* and 150 *japonica* rice.

2.2 Phylogenetic analysis of 450 rice accessions

The 450 rice accessions from RDP2 had been genotyped by 700K SNPs (McCouch et al., 2016). The phylogenetic tree was constructed by SNPhylo using SNP data. The figure of phylogenetic tree was drawn by ITOL (Interaction Tree Of Life, https://itol.embl.de/).

2.3 Phenotypic evaluation of PGWC and DEC

The 450 rice accessions were planted in the experimental fields of Guangzhou (2016GZ, 23°8' N and 113°17' E) and Yangjiang (2018YJ, 21°89' N and 111°90' E) in Guangdong Province, China, in the second cropping season in 2016 and 2018, respectively. The experiments were performed in a randomized complete block design with two replicates in both environments. The field management, including irrigation, fertilization, and disease and pest control, followed the conventional or local practice in rice production. At complete maturity, the grains were harvested and dried naturally, and then stored in storage at 15 °C. 20 g grains were de-husked by a huller (JLG-III, Chengdu, China) and milled by a polisher (JNM, Chengdu, China). PGWC and DEC of head rice were detected by the Rice Appearance Quality Determination Instrument (SC-E, Hangzhou, Chian). PGWC was the percentage of grains with chalkiness, and DEC was calculated as the product of PGWC and chalkiness size (the area of chalkiness divided by the area of whole grain). All measurements were conducted using three

independent samples from both two replicates, and the average values were used for subsequent GWAS analysis.

2.4 GWAS and QTL delimitation

GWAS was conducted as described in our previous study (Zhao et al., 2018; Yang et al., 2021) by using GAPIT version 2 software. Briefly, SNPs were filtered using the criteria of having less than 30% of missing data and minor allele frequency (MAF) > 0.05. The mixed linear model (MLM) with kinship matrix was used for GWAS, and the principal component was set to 3 in GAPIT. Manhattan and QQ plots were produced by R-package qqman (Turner, 2018). A region containing two or more than two significant SNPs ($p < 0.0001$) within 200 kb is considered as one QTL.

2.5 Linkage disequilibrium (LD) Decay analysis

The SNP information of about 900 kb (18.98-19.88 Mb) for *qDEC-10c*^{1w,2wj} was used for LD Decay analysis. The LD heatmap were drawn by the R package LD heatmap.

2.6 RNA-Sequencing

Based on haplotype analysis and consistency of flowering time, six accessions with consistent flowering time, including three accessions with High DEC (accessions 787, 806 and 854) and three accessions with low DEC (accessions 730, 937 and 1046) were selected to conduct RNA-sequencing. The spikes were sampled on the 15th day after flowering, and total RNA was extracted using Trizol reagent (Takara, Dalian, China). RNA-sequencing was conducted by the Annoroad Gene Technology (Beijing, China), and data analysis was conducted as described in our previous study (Yang et al., 2022b). Genes with read count less than 30 were regarded as no expression. The differentially expressed genes between two sets of contrasting accessions were identified according to the criteria of p-value ≤ 0.01 and fold change of pairwise comparison ≥2 or ≤0.5.

2.7 DNA Re-Sequencing

The leaves of rice seedlings were collected and DNA extraction was conducted by CTAB method. The experiment was conducted according to the standard protocol provided by Illumina. The qualified genomic DNA was fragmented by ultrasonication. These fragments were then for purification, terminal repair and addition of A at the 3' end, ligation and sequencing, and then fragmentation by agarosegel electrophoresis. Selection of appropriate DNA fragments and PCR amplification were performed to construct a sequencing library, and the library was sequenced using the Illumina NovaSeq6000 platform (BerryGenomics, Beijing, China). The raw sequence data have been reported in our previous study (Wang et al., 2023).

2.8 Real-time PCR analysis

The RNA samples for RNA-sequencing assays were used to confirm the expression level of candidate genes in this study. The cDNA synthesis was conducted using the PrimeScript^{TM} RT reagent kit (Takara, Japan). The qRT-PCR analysis was performed by qRT-PCR (Biorad CFX96, Pleasanton, CA, USA). The primers were designed by the Primer designing tool on NCBI. The gene-specific primers are listed in Supplementary Table S5. The *EF1α* was used as the normalized genes for mRNA. All reactions were repeated three times.

3 Results

3.1 Phenotypic variations of PGWC and DEC in 450 rice accessions

Phylogenetic analysis based on their genotypes determined by the 700 K SNPs (McCouch et al., 2016) indicated that the 450 rice accessions could be roughly clustered into two groups, consisting of 300 *indica* and 150 *japonica* rice (Table S1 and Figure 1A).

PGWC and DEC of the 450 rice accessions were evaluated in two environments (2016GZ and 2018YJ). The phenotype pairwise correlations between two traits were positively correlated in both environments, with correlation coefficients of 0.91 and 0.90 in 2016GZ and 2018YJ, respectively (Figure 1B). The panel showed a large variations for the two traits (Figure 1C). By comparing the phenotypes of *indica* and *japonica* rice, there were no significant differences between the two subpopulations in both environments (p>0.05) except that the DEC of *japonica* rice was significantly higher than that of *indica* rice in 2016GZ (p<0.01) (Figure 1D).

3.2 QTLs mapping for PGWC and DEC by GWAS

A total of 34 QTLs (14 for PGWC and 20 for DEC) were identified in the two environments (Figure 2; Figure S1 and Table 1). Among them, seven QTLs could be identified in both environments, while 15 and 12 QTLs were detected only in 2016GZ and 2018YJ, respectively. These QTLs were identified in different populations, including 11 QTLs in the whole population, 7 QTLs in the *indica* population, 3 QTLs in the *japonica* population, and 13 QTLs in more than one population.

Besides, it was found that some QTLs for PGWC and DEC were co-located on the same chromosomal regions. They were *qPGWC-1c*^{1wi} and *qDEC-1*^{1wi} on chromosome 1; *qPGWC-3b*^{1wj,2wj} and *qDEC-3*^{2w} on chromosome 3; *qPGWC-5a*^{1wij,2wij} and *qDEC-5a*^{1wij,2wij} on chromosome 5; *qPGWC-5c*^{1wi} and *qDEC-5b*^{1wi} on chromosome 5; *qPGWC-7*^{1wi} and *qDEC-7*^{1wi} on chromosome 7; *qPGWC-8*^{2w} and *qDEC-8a*^{2w} on chromosome 8; *qPGWC-9*^{1i} and

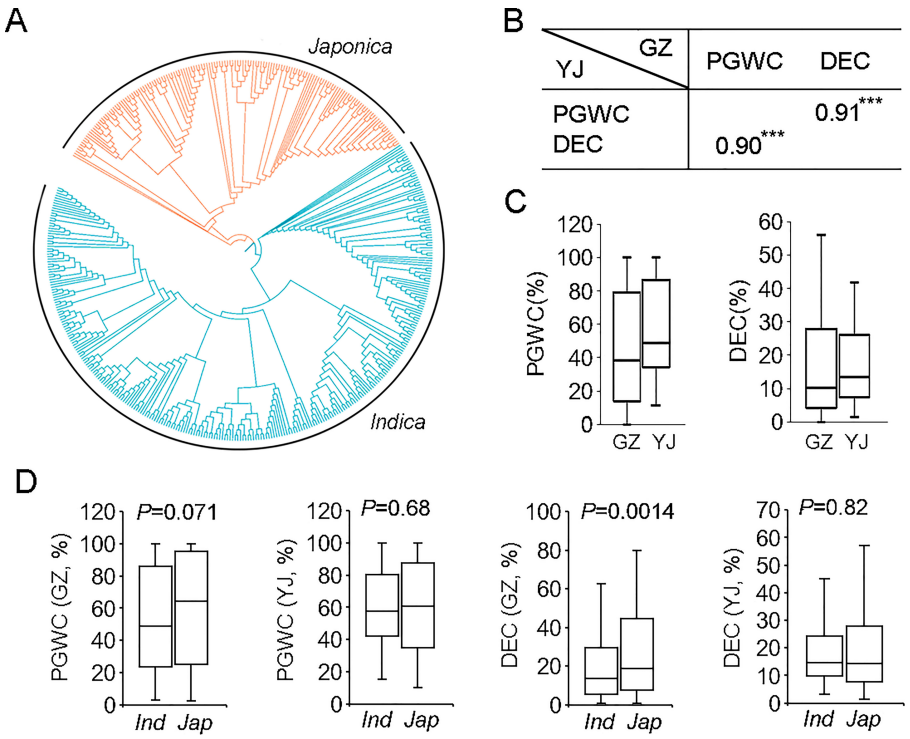


FIGURE 1
The phylogenetic tree and phenotypic comparison of 450 rice accessions used in this study. **(A)** The phylogenetic tree of 450 rice accessions. **(B)** Correlation analysis of PGWC and DEC in two environments. ***p<0.001. **(C)** Box plots of PGWC and DEC in two environments. **(D)** Phenotypic comparison between *indica* and *japonica* rice. Statistical comparison was performed by one-side *t*-test. GZ, Guangzhou; YJ, Yangjiang; PGWC, Percentage of grains with chalkiness; DEC, Degree of endosperm chalkiness.

qDEC-9a^{1i} on chromosome 9; *qPGWC-10*^{1w,2wj} and *qDEC-10c*^{1w,2wj} on chromosome 10. While the other 18 QTLs were associated with one trait (Figure 2).

Compared with the previous studies, 11 QTLs identified in this study were co-localized with the reported QTLs or cloned genes (Table 1). In particular, *qPGWC-3b*^{1wj,2wj} and *qDEC-3*^{2w} were co-located with *GS3*, a major gene regulating grain length (Mao et al., 2010); *qPGWC-5a*^{1wij,2wij} and *qDEC-5a*^{1wij,2wij} were co-located with *GW5*, a major gene regulating grain width (Weng et al., 2008) and several QTLs for chalky traits (Weng et al., 2008; Mao et al., 2010; Bian et al., 2014; Peng et al., 2014; Yun et al., 2016; Zhao et al., 2016; Wang et al., 2017; Misra et al., 2021; Qiu et al., 2021). The other 23 QTLs were newly identified in the present study.

3.3 Phenotype comparison and region analysis of *qDEC-10c*^{1w,2wj}

For the 34 QTLs identified in the present study, *qDEC-10c*^{1w,2wj} was co-located with *qPGWC-10*^{1w,2wj} and they could be identified in the whole populations in both environments (Figure 2). Besides, *qDEC-10c*^{1w,2wj} and *qPGWC-10*^{1w,2wj} had the most significant *p*-value among the newly identified QTLs (Table 1). For *qDEC-10c*^{1w,2wj} and *qPGWC-10*^{1w,2wj}loci, both PGWC and DEC were significantly different between accessions with different haplotypes (Figure 3A). Based on the SNP for GWAS, the LD decay analysis was used to narrow the region of *qDEC-10c*^{1w,2wj}. There was a significant disconnection at 19.28 Mb and 19.68 Mb of chromosome 10 (Figure 3B). Because the rice linkage disequilibrium block is about 100-200 kb (Wang et al., 2020). So the *qDEC-10c*^{1w,2wj} was delimited to an approximately 400 kb region (from 19.28 to 19.68 Mb) on chromosome 10.

3.4 Candidate gene analysis of *qDEC-10c*^{1w,2wj}

There are 51 annotated genes within the *qDEC-10c*^{1w,2wj} region (Table S2) based on release 7 of the MSU Rice Genome Annotation Project (Kawahara et al., 2013). Since the grain filling stage is the key period for seed development and chalkiness formation, and the differentially expressed genes during this stage may result in variant quality (Yang et al., 2022c; Zhao et al., 2022). To reduce the number of candidate genes, three accessions with high DEC haplotype and three accessions with low DEC haplotype based on haplotype analysis of *qDEC-10c*^{1w,2wj} were selected for gene differential expression analysis at grain filling stage (15 days after flowering). RNA-sequencing revealed that 28 genes were expressed, of which five genes (*LOC_Os10g36170*, *LOC_Os10g36260*, *LOC_Os10g36270*, *LOC_Os10g36340* and *LOC_Os10g36610*) were differentially expressed between the two sets of contrasting accessions (Table S2). qRT-PCR assays confirmed that the expression

10.3389/fpls.2023.1184276

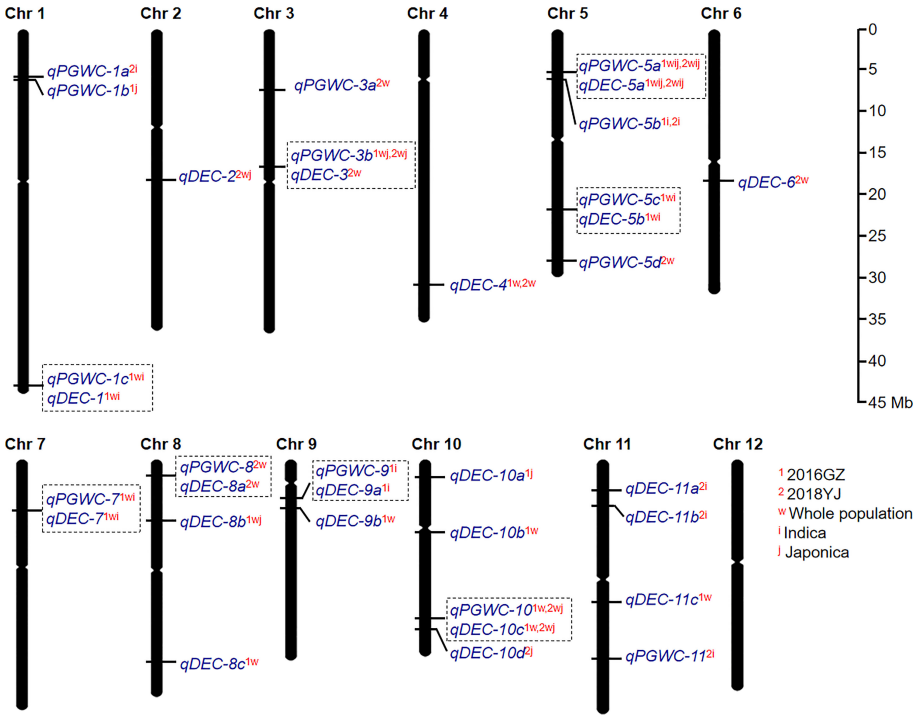


FIGURE 2
The global view of QTL mapping for PGWC and DEC by GWAS. The co-located QTLs were surrounded with dotted boxes. The superscript letters (1 and 2) refer to the two environments (2016GZ and 2018YJ, respectively). The superscript letters (w, i and j) refer to the three populations (whole, *indica* and *japonica*, respectively). For example, *qDEC-10c*^{1w,2wj} indicates that *qDEC-10c* can be identified in 2016GZ by the whole population and in 2018YJ by the whole and *japonica* population.

patterns of *LOC_Os10g36170*, *LOC_Os10g36260*, *LOC_Os10g36340* and *LOC_Os10g36610* were consistent with the RNA-sequencing result (Figure 4), while *LOC_Os10g36270* showed inconsistent differential expression compared with the RNA-sequencing results (Figure S2). The expression level of *LOC_Os10g36170* in the accessions with high DEC was lower than that in accessions with low DEC ($p < 0.01$), while expression levels of *LOC_Os10g36260*, *LOC_Os10g36340* and *LOC_Os10g36610* in the accessions with high DEC was higher than that in accessions with low DEC ($p < 0.01$) (Figure 4).

3.5 Phenotype comparisons among different haplotypes of four candidate genes

We further analyzed the promoter region (2 kb from ATG) of the four candidate genes using re-sequencing data of the 450 accessions, and conducted integration analysis between the four major haplotypes (Hap 1 to 4) of each gene and phenotypes. For *LOC_Os10g36170*, accessions with Hap 1 had higher PGWC and DEC than other haplotypes, and three sequence variations in the promoter region may be the difference between Hap 1 and the other haplotypes (Figure 5A; Table S3). For *LOC_Os10g36260*, accessions with Hap 1 had higher PGWC and DEC than other haplotypes, and 11 sequence variations in the promoter region may be the difference between Hap 1 and the other haplotypes (Figure 5B; Table S3). For *LOC_Os10g36340*, accessions with Hap 1 and Hap 2 had higher PGWC and DEC than Hap 3 and Hap 4, and the sequence deletion or existence in the promoter region may account for their differences (Figure 5C; Table S3). For *LOC_Os10g36610*, accessions with Hap 1 had higher PGWC and DEC than the other haplotypes, but no possible differences were found in the sequence comparison of their promoter regions (Figure 5D, Table S3).

4 Discussion

PGWC and DEC are two important indexes to evaluate rice appearance quality, and some previous studies have reported that there is a high correlation between them (Wang et al., 2017; Qiu et al., 2021). In the present study, the correlation between PGWC and DEC was high (Figure 1B). The co-located QTLs identified for PGWC and DEC further confirmed this (Figure 2). Therefore, PGWC and DEC could be simultaneously improved by molecular breeding without fear of genetic drag. However, 18 out of 34 QTLs were only associated with one trait (Figure 2), indicating that there were different regulatory mechanisms among the two traits. Research on the specific regulation mechanism between PGWC and DEC will further enrich the genetic mechanism of grain chalkiness.

Among the 34 QTLs identified in this study, 11 QTLs were co-localized with the reported QTLs or genes (Table 1), indicating the

10.3389/fpls.2023.1184276

TABLE 1 QTLs identified for PGWC and DEC in the present study.

QTL	Chromosome	Environment†	Population	The most significant SNP position (bp)#	*p*-value	Co-located QTL or cloned gene
PGWC						
qPGWC-1a^{2i}	1	2018YJ	*Indica*	6,083,694	2.82E-05	*qDEC1* (Qiu et al., 2021)
qPGWC-1b^{1j}	1	2016GZ	*Japonica*	6,241,628	3.47E-05	
qPGWC-1c^{1wi}	1	2016GZ	whole	43,184,323	5.60E-05	
	1	2016GZ	*Indica*	43,241,738	2.94E-05	
qPGWC-3a^{2w}	3	2018YJ	whole	7,795,320	1.97E-07	
qPGWC-3b^{1wj,2wj}	3	2016GZ	whole	16,709,942	1.16E-06	*GS3* (Mao et al., 2010)
	3	2016GZ	*Japonica*	16,856,475	3.01E-06	
	3	2018YJ	whole	16,876,766	1.24E-07	
	3	2018YJ	*Japonica*	16,876,766	9.59E-06	
qPGWC-5a^{1wij,2wij}	5	2016GZ	whole	5,371,772	7.91E-16	*GW5* (Weng et al., 2008)
	5	2016GZ	*Indica*	5,359,520	5.18E-11	
	5	2016GZ	*Japonica*	5,342,339	7.41E-09	
	5	2018YJ	whole	5,371,772	4.60E-10	
	5	2018YJ	*Indica*	5,359,520	6.95E-06	
	5	2018YJ	*Japonica*	5,367,184	2.71E-06	
qPGWC-5b^{1i,2i}	5	2016GZ	*Indica*	6,066,999	2.98E-05	
	5	2018YJ	*Indica*	6,005,763	2.11E-06	
qPGWC-5c^{1wi}	5	2016GZ	whole	21,924,400	1.69E-05	
	5	2016GZ	*Indica*	21,944,065	1.65E-05	
qPGWC-5d^{2w}	5	2018YJ	whole	27,961,292	4.14E-05	
qPGWC-7^{1wi}	7	2016GZ	whole	6,211,855	2.99E-06	*qTr7.1* (Wang et al., 2017)
	7	2016GZ	*Indica*	6,181,910	3.47E-05	
qPGWC-8^{2w}	8	2018YJ	whole	2,035,102	2.12E-05	

(Continued)

10.3389/fpls.2023.1184276

TABLE 1 Continued

QTL	Chromosome	Environment†	Population	The most significant SNP position (bp)#	*p*-value	Co-located QTL or cloned gene
qPGWC-9^{1i}	9	2016GZ	*Indica*	4,992,343	6.73E-06	
qPGWC-10^{1w,2wj}	10	2016GZ	whole	19,480,113	4.84E-06	
	10	2018YJ	whole	19,481,956	2.41E-08	
	10	2018YJ	*Japonica*	19,481,956	1.30E-06	
qPGWC-11^{2i}	11	2018YJ	*Indica*	24,399,195	2.01E-05	*qDEC11b* (Zhao et al., 2016); *qPGWC11c* (Zhao et al., 2016)
DEC						
qDEC-1^{1wi}	1	2016GZ	whole	43,139,043	4.83E-05	
	1	2016GZ	*Indica*	43,118,717	1.81E-06	
qDEC-2^{2wj}	2	2018YJ	whole	18,253,520	2.50E-05	
	2	2018YJ	*Japonica*	18,388,819	4.87E-05	
qDEC-3^{2w}	3	2018YJ	whole	16,876,766	2.43E-06	*GS3* (Mao et al., 2010)
qDCE-4^{1w,2w}	4	2016GZ	whole	30,708,157	1.54E-05	*PGC4.4* (Misra et al., 2021); *qWCR4* (Yun et al., 2016)
	4	2018YJ	whole	30,708,157	2.24E-05	
qDEC-5a^{1wij,2wij}	5	2016GZ	whole	5,371,772	3.22E-12	*GW5* (Weng et al., 2008)
	5	2016GZ	*Indica*	5,359,520	8.12E-08	
	5	2016GZ	*Japonica*	5,359,740	4.50E-05	
	5	2018YJ	whole	5,419,124	1.23E-09	
	5	2018YJ	*Indica*	5,431,662	4.57E-06	
	5	2018YJ	*Japonica*	5,419,124	6.38E-06	
qDEC-5b^{1wi}	5	2016GZ	whole	21,760,992	3.40E-05	
	5	2016GZ	*Indica*	21,873,455	1.10E-05	
qDEC-6^{2w}	6	2018YJ	whole	18,799,852	2.91E-06	
qDEC-7^{1wi}	7	2016GZ	whole	6,218,264	6.89E-10	*qTr7.1* (Wang et al., 2017)
	7	2016GZ	*Indica*	6,211,855	3.70E-08	

(Continued)

TABLE 1 Continued

QTL	Chromosome	Environment†	Population	The most significant SNP position (bp)#	*p*-value	Co-located QTL or cloned gene
qDEC-8a^{2w}	8	2018YJ	whole	2,035,102	6.91E-06	
qDEC-8b^{1wj}	8	2016GZ	whole	7,796,155	1.04E-05	
	8	2016GZ	*Japonica*	7,754,397	7.17E-06	
qDEC-8c^{1w}	8	2016GZ	whole	24,785,271	1.55E-05	*qWBA8-W+* (Peng et al., 2014); *qDEC8b* (Zhao et al., 2016); *qWBR8* (Yun et al., 2016)
qDEC-9a^{1i}	9	2016GZ	*Indica*	4,962,744	2.86E-05	
qDEC-9b^{1w}	9	2016GZ	whole	6,109,637	5.26E-06	*QPGWC.HN-9* (Bian et al., 2014)
qDEC-10a^{1j}	10	2016GZ	*Japonica*	2,319,127	4.47E-05	
qDEC-10b^{1w}	10	2016GZ	whole	9,320,097	3.97E-06	
qDEC-10c^{1w,2wj}	10	2016GZ	whole	19,480,113	3.04E-05	
	10	2018YJ	whole	19,481,956	1.47E-10	
	10	2018YJ	*Japonica*	19,481,956	1.06E-08	
qDEC-10d^{2j}	10	2018YJ	*Japonica*	20,500,122	2.25E-05	
qDEC-11a^{2i}	11	2018YJ	*Indica*	3,950,541	1.13E-05	
qDEC-11b^{2i}	11	2018YJ	*Indica*	5,770,569	7.68E-06	
qDEC-11c^{1w}	11	2016GZ	whole	17,446,555	3.51E-05	

† The experimental bases of Guangzhou in 2016 (2016GZ) and Yangjiang in 2018 (2018YJ).
# Position of the most significant SNP at the QTL region.

reliability of GWAS analysis and diversity of rice accessions used in this study. Among them, *qPGWC-3b*^{1wj,2wj} and *qDEC-3*^{2w} were co-located with *GS3*, and *qPGWC-5a*^{1wij,2wij} and *qDEC-5a*^{1wij,2wij} were co-located with *GW5* (Table 1). In several studies, *GS3* and *GW5* loci have been reported to control grain shape and chalkiness simultaneously (Qiu et al., 2015; Wang et al., 2017), indicating that the two genes have multiple effects. However, the regulation intensity and direction of different alleles of *GS3* and *GW5* on multiple quality traits are worthy of careful study, which can provide support for the cooperative improvement of grain shape and chalkiness.

In addition to the four QTLs mentioned above, the other seven QTLs were co-located with QTLs that had not yet been cloned. *qDCE-4*^{1w,2w} could be identified in the two environments (Figure 2) and co-located with *PGC4.4* (Misra et al., 2021)and *qWCR4* (Yun et al., 2016). Misra et al. demonstrated that *PGC4.4* is an important locus affecting chalkiness, and *LOC_Os04g52230* and *LOC_Os04g52240* may be the optimal candidate genes for this QTL by haplotype analysis (Misra et al., 2021). Other QTLs are co-located with some stable or unstable QTLs (Table 1). These information suggest that the stable expression of some QTLs is relative, greatly influenced by the environment, and may also be related to the background of research materials.

Because PGWC and DEC are highly positively correlated, it is possible to identify QTLs of PGWC and DEC in the similar region by using the same materials and environments. Wang et al. detected four QTLs affecting DEC and four QTLs affecting PGWC. Among them, *qDEC3* is co-located with *qPGWC3* and *qDEC8* is co-located with *qPGWC8* (Wang et al., 2017). Among the newly identified QTLs in this study, *qPGWC-10*^{1w,2wj} was co-located with *qDEC-10c*^{1w,2wj} on chromosome 10, and they were stably detected in both environments and had the most significant *p*-value (Table 1). Interestingly, a QTL controlling grain width (*qGW10.2*) (Wang et al., 2017) was also located in the *qDEC-10c*^{1w,2wj}region. Since

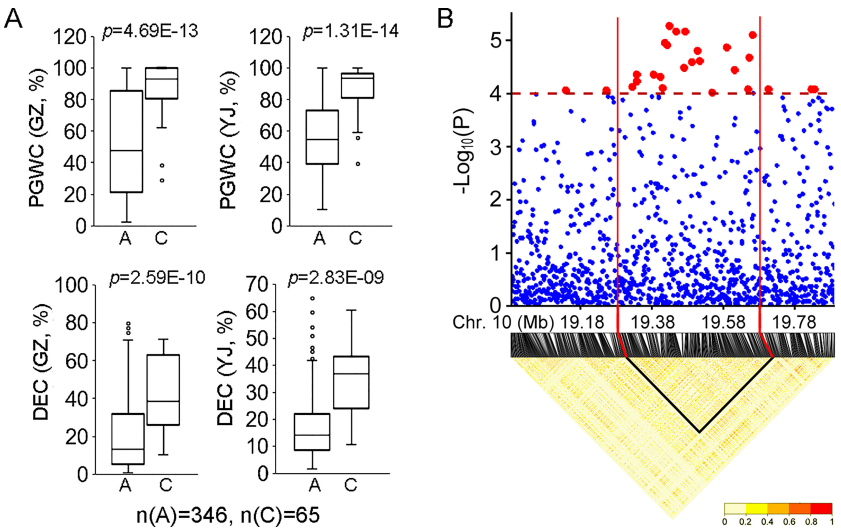


FIGURE 3
Phenotype and region analysis of *qDEC-10c*^{1w,2wj}. **(A)** Phenotypic comparison of PGWC and DEC between accessions with different haplotypes for *qDEC-10c*^{1w,2wj}. Statistical comparison was performed by one-side t-test. **(B)** Local manhattan plot of GWAS and linkage disequilibrium heatmap for *qDEC-10c*^{1w,2wj}.

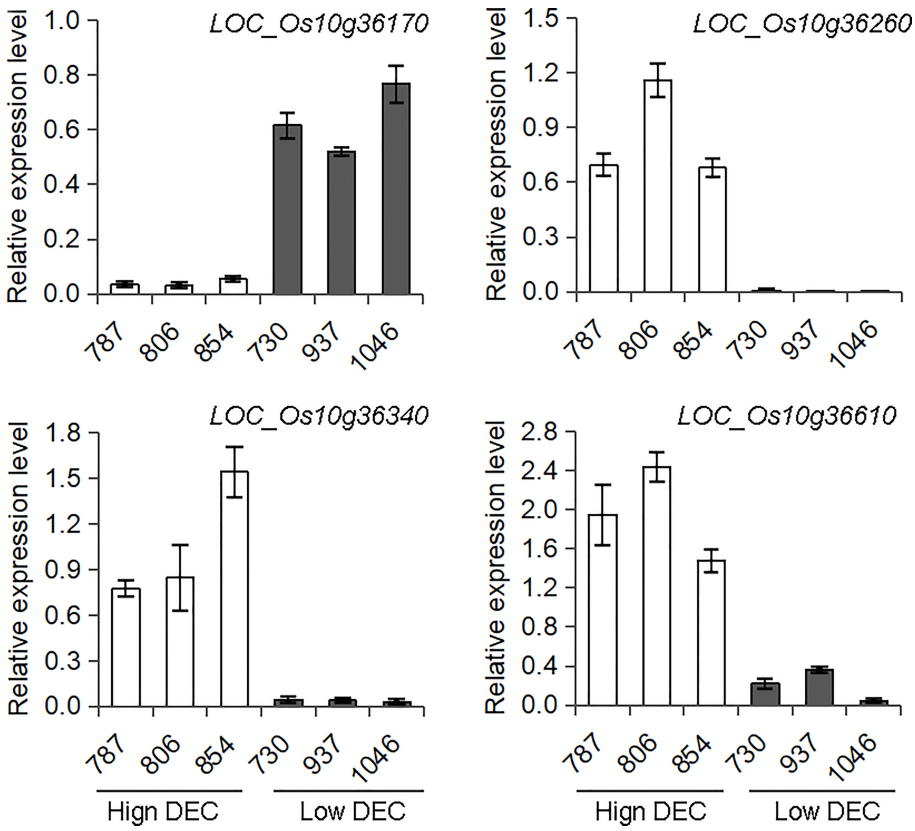


FIGURE 4
Expression analysis of candidate genes for *qDEC-10c*^{1w,2wj}. The spikes were sampled on the15th day after flowering of three accessions with high DEC (accessions 787, 806 and 854) and three accessions with low DEC (accessions 730, 937 and 1046) were selected based on haplotype analysis to conduct RNA-sequencing and qRT-PCR. Different letters indicate significant difference at $P < 0.01$.

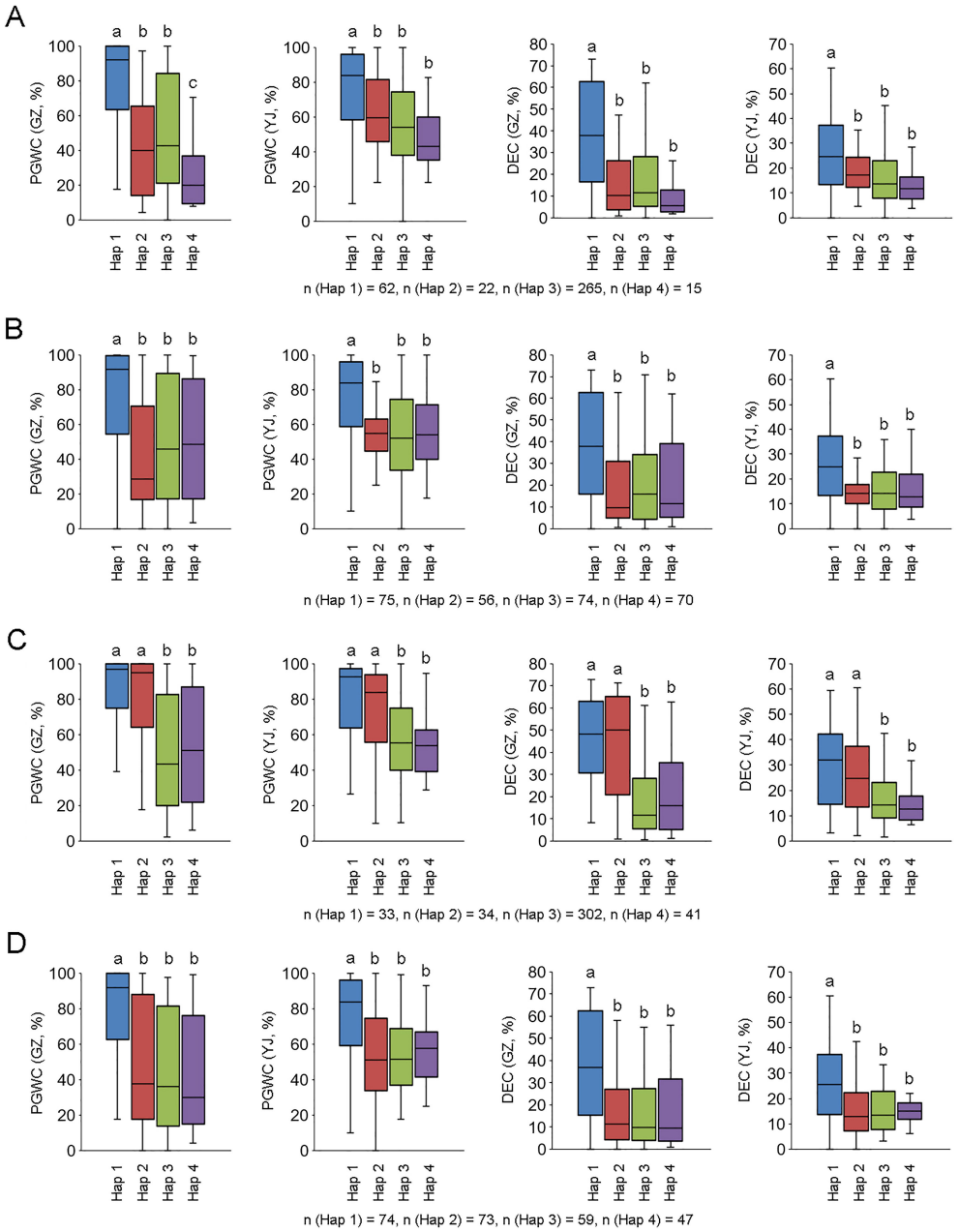


FIGURE 5
Phenotype comparisons among different haplotypes of four candidate genes. **(A–D)** are *LOC_Os10g36170*, *LOC_Os10g36260*, *LOC_Os10g36340* and *LOC_Os10g36610*, respectively. Different letters indicate significant difference at $P < 0.01$.

there is a high positive correlation between grain width and chalkiness (Qiu et al., 2015; Wang et al., 2017), whether *qDEC-10c*^{1w,2wj} regulates both grain width and chalkiness needs further investigation. We further mined the candidate genes for *qDEC-10c*^{1w,2wj}. By integrated analysis of interval definition, differential gene expression and haplotype comparison, four candidate genes were identified for *qDEC-10c*^{1w,2wj}. Three genes (*LOC_Os10g36260*, *LOC_Os10g36340*, *LOC_Os10g36610*) encoding unknown functional proteins and one gene (*LOC_Os10g36170*) encoding a lipid transfer protein (LTP) were considered as its candidate genes. LTPs play diverse roles in various biological processes, such as pollen development, stress response, plant signaling and seed maturation (Zhang et al., 2010; Wang et al., 2012). A previous study indicated that down-regulated expression of *OsLTPL36* leads to chalky endosperm and impaired seed germination in rice (Wang et al., 2015). Our results also exhibited that the expression level of *LOC_Os10g36170* in the accessions with high DEC was significantly lower than that in accessions with low DEC (Figure 4), which speculates that *LOC_Os10g36170* may functions on regulating the formation of grain chalkiness. In addition, we found some possible key sequence variants related to the expression and phenotype in the promoter regions of three genes by integrating phenotype, haplotype and sequence variation analysis (Table S3 and Figure 5). Further studies are needed to confirm the functions of

these candidate genes in grain chalkiness through gain or loss-of function analysis. The identification of novel QTLs and candidate genes for rice appearance quality provides new sources for molecular breeding and cloning of genes associated with rice quality.

5 Conclusion

A total of 34 QTLs for grain chalkiness were identified by GWAS in 450 rice accessions. Based on the haplotype analysis, LD decay analysis, RNA-sequencing and promoter sequences analysis, four genes were considered as the candidate genes underlying the potential QTL (*qDEC-10c*^{1w,2w}). These results will enrich the genetic bases of grain chalkiness and contributes to molecular breeding for high-quality rice.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA948558.

Author contributions

SZ, WY and JZ designed and supervised this work. XH, JW analyzed the data and wrote the article. LC, HF, TY and JD investigated the phenotype, and performed sample detection. YM and LZ performed the images. JC, DL and BL participated in the revision of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Research and Development Plan of Key Fields in Guangdong Province(2018B020202004), the National Natural Science Foundation of China (32272085), the Guangdong Basic and Applied Basic Research Foundation (2023A1515010400), the Special Fund for Scientific Innovation Strategy-construction of High Level Academy of Agriculture Science (R2018YJ-YB1001 and R2021PY-QF002), the Project of Collaborative Innovation Center of GDAAS (XT202218), and the Guangdong Key Laboratory of New Technology in Rice Breeding (2020B1212060047).

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/fpls.2023.1184276/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Manhattan plots of GWAS for PGWC and DEC using different populations in two environments.

SUPPLEMENTARY FIGURE 2
The expression level of *LOC_Os10g36270* measured by qRT-PCR.

SUPPLEMENTARY TABLE 1
Information of the 450 rice accessions used in this study.

SUPPLEMENTARY TABLE 2
The read count of genes within *qDEC-10c*^{1w,2wj} region measured by RNA-seq.

SUPPLEMENTARY TABLE 3
Sequence variations in promoter regions of the four candidate genes for *qDEC-10c*^{1w,2wj}.

SUPPLEMENTARY TABLE 4
Primers used in this study.

References

Bian, J., Li, C., He, H., Shi, H., and Yan, S. (2014). Identification and analysis of QTLs for grain quality traits in rice using an introgression lines population. *Euphytica* 195, 83–93. doi: 10.1007/s10681-013-0976-1

Fan, T., Yang, W., Zeng, X., Xu, X., Xu, Y., Fan, X., et al. (2020). A rice autophagy gene OsATG8b is involved in nitrogen remobilization and control of grain quality. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00588

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* 6, 4. doi: 10.1186/1939-8433-6-4

Li, Y., Fan, C., Xing, Y., Yun, P., Luo, L., Yan, B., et al. (2014). Chalk5 encodes a vacuolar h(+)-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* 46, 398–404. doi: 10.1038/ng.2923

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. U S* 107, 19579–19584. doi: 10.1073/pnas.1014419107

McCouch, S. R., Wright, M. H., Tung, C. W., Maron, L. G., McNally, K. L., Fitzgerald, M., et al. (2016). Open access resources for genome-wide association mapping in rice. *Nat. Commun.* 7, 10532. doi: 10.1038/ncomms10532

Misra, G., Badoni, S., Parween, S., Singh, R. K., Leung, H., Ladejobi, O., et al. (2021). Genome-wide association coupled gene to gene interaction studies unveil novel epistatic targets among major effect loci impacting rice grain chalkiness. *Plant Biotechnol. J.* 19, 910–925. doi: 10.1111/pbi.13516

Peng, B., Wang, L., Fan, C., Jiang, G., Luo, L., Li, Y., et al. (2014). Comparative mapping of chalkiness components in rice using five populations across two environments. *BMC Genet.* 15, 49. doi: 10.1186/1471-2156-15-49

Qiu, X., Pang, Y., Yuan, Z., Xing, D., Xu, J., Dingkuhn, M., et al. (2015). Genome-wide association study of grain appearance and milling quality in a worldwide collection of indica rice germplasm. *PloS One* 10, e0145577. doi: 10.1371/journal.pone.0145577

Qiu, X., Yang, J., Zhang, F., Niu, Y., Zhao, X., Shen, C., et al. (2021). Genetic dissection of rice appearance quality and cooked rice elongation by genome-wide association study. *Crop J.* 9, 1470–1480. doi: 10.1016/j.cj.2020.12.010

Ryoo, N., Yu, C., Park, C. S., Baik, M. Y., Park, I. M., Cho, M. H., et al. (2007). Knockout of a starch synthase gene OsSSIIIa/Flo5 causes white-core floury endosperm in rice (Oryza sativa l.). *Plant Cell Rep.* 26, 1083–1095. doi: 10.1007/s00299-007-0309-8

Turner, S. D. (2018). Qqman: An r package for visualizing GWAS results using q-q and manhattan plots. *J. Open Source Software* 3, 731. doi: 10.21105/joss.00731

Wang, H. W., Hwang, S. G., Karuppanapandian, T., Liu, A., Kim, W., Jang, C. S., et al. (2012). Insight into the molecular evolution of non-specific lipid transfer proteins *via* comparative analysis between rice and sorghum. *DNA Res.* 19, 179–194. doi: 10.1093/dnares/dss003

Wang, X., Pang, Y., Wang, C., Chen, K., Zhu, Y., Shen, C., et al. (2017). New candidate genes affecting rice grain appearance and milling quality detected by genome-wide and gene-based association analyses. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01998

Wang, Q., Tang, J. L., Han, B., and Huang, X. H. (2020). Advances in genome-wide association studies of complex traits in rice. *Theor. Appl. Genet.* 133, 1415–1425. doi: 10.1007/s00122-019-03473-3

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat. Genet.* 40, 1370–1374. doi: 10.1038/ng.220

Wang, J., Yang, W., Zhang, S., Hu, H., Yuan, Y., Dong, J., et al. (2023). A pangenome analysis pipeline provides insights into functional gene identification in rice. *Genome Biol.* 24, 19. doi: 10.1186/s13059-023-02861-9

Wang, X., Zhou, W., Lu, Z., Ouyang, Y., O, C. S., and Yao, J. (2015). A lipid transfer protein, OsLTPL36, is essential for seed development and seed quality in rice. *Plant Sci.* 239, 200–208. doi: 10.1016/j.plantsci.2015.07.016

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., et al. (2008). Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* 18, 1199–1209. doi: 10.1038/cr.2008.307

Wu, B., Yun, P., Zhou, H., Xia, D., Gu, Y., Li, P., et al. (2022). Natural variation in WHITE-CORE RATE 1 regulates redox homeostasis in rice endosperm to affect grain quality. *Plant Cell* 34, 1912–1932. doi: 10.1093/plcell/koac057

Yang, W., Chen, L., Zhao, J., Wang, J., Li, W., Yang, T., et al. (2022b). Genome-wide association study of pericarp color in rice using different germplasm and phenotyping methods reveals different genetic architectures. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.841191

Yang, W., Jiang, X. Y., Xie, Y. L., Chen, L., Zhao, J. L., Lui, B., et al. (2022c). Transcriptome and metabolome analyses reveal new insights into the regulatory mechanism of head milled rice rate. *Plants-Basel* 11, 2838. doi: 10.3390/plants11212838

Yang, W., Xu, P., Zhang, J., Zhang, S., Li, Z., Yang, K., et al. (2022a). OsbZIP60-mediated unfolded protein response regulates grain chalkiness in rice. *J. Genet. Genomics* 49, 414–426. doi: 10.1016/j.jgg.2022.02.002

Yang, W., Zhao, J., Zhang, S., Chen, L., Yang, T., Dong, J., et al. (2021). Genome-wide association mapping and gene expression analysis reveal the negative role of OsMYB21 in regulating bacterial blight resistance in rice. *Rice (N Y)* 14, 58. doi: 10.1186/s12284-021-00501-z

Yun, P., Zhu, Y., Wu, B., Gao, G., Sun, P., Zhang, Q., et al. (2016). Genetic mapping and confirmation of quantitative trait loci for grain chalkiness in rice. *Mol. Breed.* 36, 162. doi: 10.1007/s11032-016-0600-x

Zhang, D., Liang, W., Yin, C., Zong, J., Gu, F., and Zhang, D. (2010). OsC6, encoding a lipid transfer protein, is required for postmeiotic anther development in rice. *Plant Physiol.* 154, 149–162. doi: 10.1104/pp.110.158865

Zhao, X., Daygon, V. D., McNally, K. L., Hamilton, R. S., Xie, F., Reinke, R. F., et al. (2016). Identification of stable QTLs causing chalk in rice grains in nine environments. *Theor. Appl. Genet.* 129, 141–153. doi: 10.1007/s00122-015-2616-8

Zhao, J., Yang, W., Zhang, S., Yang, T., Liu, Q., Dong, J., et al. (2018). Genome-wide association study and candidate gene analysis of rice cadmium accumulation in grain in a diverse rice collection. *Rice (New York N.Y.)* 11, 61. doi: 10.1186/s12284-018-0254-x

Zhao, D., Zhang, C., Li, Q., and Liu, Q. (2022). Genetic control of grain appearance quality in rice. *Biotechnol. Adv.* 60, 108014. doi: 10.1016/j.biotechadv.2022.108014

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 June 2023
DOI 10.3389/fpls.2023.1173861

OPEN ACCESS

EDITED BY
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

REVIEWED BY
Tian Li,
Chinese Academy of Agricultural Sciences, China
Yanhao Xu,
Hubei Academy of Agricultural Sciences, China

*CORRESPONDENCE
Jirui Wang
isphsc@sicau.edu.cn;
wangjirui@gmail.com

SPECIALTY SECTION
RECEIVED 25 February 2023
ACCEPTED 03 May 2023
PUBLISHED 05 June 2023

CITATION
Tang H, Dong H, Guo X, Cheng M, Li M, Chen Q, Yuan Z, Pu Z and Wang J (2023) Identification of candidate gene for the defective kernel phenotype using bulked segregant RNA and exome capture sequencing methods in wheat. *Front. Plant Sci.* 14:1173861. doi: 10.3389/fpls.2023.1173861

Identification of candidate gene for the defective kernel phenotype using bulked segregant RNA and exome capture sequencing methods in wheat

Hao Tang[1], Huixue Dong[1], Xiaojiang Guo[2], Mengping Cheng[1], Maolian Li[1], Qian Chen[1], Zhongwei Yuan[1], Zhien Pu[2] and Jirui Wang[1,2]*

[1]State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu, China, [2]Ministry of Education Key Laboratory for Crop Genetic Resources and Improvement in Southwest China, Sichuan Agricultural University, Chengdu, China

Wheat is a significant source of protein and starch worldwide. The defective kernel (Dek) mutant *AK-3537*, displaying a large hollow area in the endosperm and shrunken grain, was obtained through ethyl methane sulfonate (EMS) treatment of the wheat cultivar Aikang 58 (AK58). The mode of inheritance of the *AK-3537* grain Dek phenotype was determined to be recessive with a specific statistical significance level. We used bulked segregant RNA-seq (BSR-seq), BSA-based exome capture sequencing (BSE-seq), and the ΔSNP-index algorithm to identify candidate regions for the grain Dek phenotype. Two major candidate regions, DCR1 (Dek candidate region 1) and DCR2, were identified on chromosome 7A between 279.98 and 287.93 Mb and 565.34 and 568.59 Mb, respectively. Based on transcriptome analysis and previous reports, we designed KASP genotyping assays based on SNP variations in the candidate regions and speculated that the candidate gene is *TraesCS7A03G0625900* (*HMGS-7A*), which encodes a 3-hydroxy-3-methylglutaryl-CoA synthase. One SNP variation located at position 1,049 in the coding sequence (G>A) causes an amino acid change from Gly to Asp. The research suggests that functional changes in *HMGS-7A* may affect the expression of key enzyme genes involved in wheat starch syntheses, such as *GBSSII* and *SSIIIa*.

KEYWORDS

BSE-seq, BSR-seq, AK58, exome capture, *HMGS-7A*

Introduction

Wheat (*Triticum aestivum* L.) is a major global crop, providing approximately 20% of the total caloric intake for the world's population (Ogbonnaya et al., 2013; Ma et al., 2022). Therefore, maintaining high grain quality and yield of wheat is essential for food security (Shiferaw et al., 2013). Wheat grain, the reproductive and storage organ, plays a crucial role in wheat propagation, spread, and yield and primarily consists of the embryo, endosperm, and seed coat (Ebrahimnejad and Rameeh, 2016). In wheat, grain filling refers to the process of starch biosynthesis and accumulation in the endosperm (Ahmed et al., 2015). Starch is the main component of wheat grain, accounting for 65%–70% of dry grain weight (Housley et al., 1981), and it significantly impacts wheat flour quality (Mancebo et al., 2015).

Starch synthesis in wheat grain starts after fertilization and continues until approximately 35 days, when the grain matures and dries (Khatun and Ahmed, 2015). When starch is synthesized in the endosperm, sucrose produces through leaf photosynthesis and enters the cytoplasm, serving as the carbon source for starch synthesis in wheat (Guo et al., 2015). Multiple factors influence wheat grain development, including wheat tissue organs (Barneix, 2007), starch synthesis-related enzymes, plant hormones, and environmental factors. Relevant tissue organs include leaves and stem sheaths. At the same time, starch synthesis-related enzymes encompass sucrose synthase (SuSy), ADP-glucose pyrophosphorylase (AGPase), granule-bound starch synthase (GBSS), starch synthase (SS), starch branching enzyme (SBE), starch debranching enzyme (DEB), starch phosphorylase (SP), and sucrose convertase (SC), among others (Wang et al., 2014). Plant hormones affecting wheat grain development include ethylene (ET), brassinosteroid (BR), gibberellin (GA), and abscisic acid (ABA), among others (Liu et al., 2013; Xiong et al., 2022). Environmental factors such as temperature, light, and soil moisture also play a role (Liu et al., 2016; Jiang et al., 2017; Djanaguiraman et al., 2020; Mirosavljević et al., 2021).

Defective kernel (Dek) mutants exhibit shrunken grains, and grain filling in crop Dek mutants is drastically impaired (Wang et al., 2017). Generally, the Dek phenotype reduces grain weight and significantly affects grain appearance and seed vigor (Li et al., 2017; Fu et al., 2019; Qi et al., 2019). Many Dek mutants have been identified in maize and rice, and numerous genetic loci regulating grain fullness have also been discovered. Dek mutants, such as *Dek10*, *Dek35*, *Dek36*, *Dek37*, *Dek39*, *Dek40*, and *Dek42* (Chen et al., 2017; Qi et al., 2017; Wang et al., 2017; Dai et al., 2018; Li et al., 2018; Ren et al., 2019; Zuo et al., 2019), display germinated mutant kernels that are lethal during the seedling stage; *Dek15*, *Dek38*, *Dek41*, and *Dek44* (Lid et al., 2002; Garcia et al., 2017; He et al., 2019; Qi et al., 2019; Zhu et al., 2019) seeds cannot germinate at all, resulting in lethal embryo mutations. For example, the *Dek15* gene encodes sister chromatid cohesion protein 4 (SCC4), and mutation of this gene disrupts the cell cycle and nuclear replication, leading to the complete failure of seed germination (He et al., 2019). The *Dek38* gene encodes TEL2-interaction protein 2 (TTI2) molecular chaperone protein, which affects the development of male germ cells (Garcia et al., 2017). The *Dek1* gene, located in the 47.1 to 47.4 Mb region on chromosome 1 in maize, is involved in the differentiation and development of maize aleurone cells. A mutation of this gene leads to embryo lethality and affects the development of the aleurone layer and the accumulation of endosperm gliadin content (Lid et al., 2002; Song et al., 2020). At present, only a few Dek-related studies have been reported in Triticeae. Three QTLs, *QDek.Caas-3BS.1*, *QDek.Caas-3BS.2*, and *QDek.Caas-4AL*, associated with wheat grain Dek were identified using wheat mutant groups, explaining 14.78%–18.17%, 16.61%–21.83%, and 19.08%–28.19% of phenotypic variances, respectively (Fu et al., 2019). The loss-of-function mutation of the *sex6* (*SSIIa*) gene on chromosome 7H in barley causes amylopectin synthesis to decrease to less than 20% of the wild-type level. Simultaneously, the mutation also affects the binding of starch synthetases SSI, SBEIIa, and SBEIIb to starch granules and ultimately causes barley grain to become shrunken (Dek) (Morell et al., 2003). The *barley's sex6 (SSIIa) mutant* was crossed with the *amo1* (*SSIIIa*) mutant to generate the *sex6amo1* double mutant, which produces high-amylose starch. The level of granule-bound starch synthase I (GBSSI) protein in starch granules increased, and starch synthase I (SSI), SSIIa, starch branching enzyme IIa (SBEIIa), and SBEIIb also significantly increased in the starch granules. The double mutant's Dek phenotype was restored to a normal grain phenotype, indicating that changes in starch synthase function in cereal crops can also lead to shrunken grains (Li et al., 2011). These genes are crucial for synthesizing and accumulating starch and protein in the endosperm. Therefore, excavating DeK-related genes will be conducive to improving crop grain yield and quality.

This study analyzed Aikang58 (AK58) and its Dek mutant line *AK-3537* to investigate grain characteristics in different environments over several years. *AK-3537* exhibited poor grain filling, collapsed abdominal grooves, and shrunken grains. Transcriptome analysis using RNA-seq was performed on the grains of AK58 and *AK-3537*, and an F_2 population with 130 individuals was constructed using *AK-3537* and AK58. Two high-confidence candidate regions, DCR1 (Dek candidate region 1) and DCR2, regulating wheat grain Dek, were mapped using BSE-seq (Bulked Segregant Exome Capture Sequencing) and BSR-seq (Bulked Segregant RNA-seq). We identified *TraesCS7A03G0625900* in the DCR1 region as a candidate gene for the wheat grain Dek phenotype using KASP (Kompetitive Allele-Specific Polymerase Chain Reaction) markers, the Chinese Spring reference genome (RefSeq 2.1), and the AK58 genome. This study will contribute to a deeper understanding of the regulatory mechanisms underlying wheat grain morphology and provide new insights to improve wheat yield through breeding.

Materials and methods

Plant materials

The EMS (ethyl methane sulfonate) mutant Dek line *AK-3537* originated from the wheat variety AK58. The Dek phenotype was stably inherited after eight generations of self-pollination. For genetic analysis and mapping, *AK-3537* was crossed with AK58 to

produce an F_2 population of 130 individuals. Wheat materials were cultivated in experimental fields and greenhouses at Sichuan Agricultural University, Chengdu Chongzhou (103° 38' E, 30° 32' N), Wenjiang (103° 51' E, 30° 43' N), and Xishuangbanna (99° 56' E, 21° 08' N), China. All field trials were well irrigated and managed following local standard practices, and all *AK-3537* × AK58 F_2 plants were grown in a greenhouse at 20°C with a 16-h/8-h light/dark cycle.

In the field, a total of 20 plants from AK58 (10 individuals) and *AK-3537* (10 individuals) were randomly selected to investigate agronomic traits, including plant height (PH), tiller number (TN), heading date (HD), flag leaf length (FLL), and flag leaf width (FLW), using the method reported in previous studies. Thousand kernel weight (TKW), grain length (GL), and grain width (GW) were also measured using previously reported methods (Liu et al., 2017). Excel 2019 (Microsoft, Redmond, WA, USA) was used to calculate the phenotypic data. Analysis of variance was conducted, and individuals were ranked through Duncan's test and plotted using GraphPad Prism V9.0.0. R software (version 4.2.1) was used as a plotting tool to calculate the wheat Dek phenotypic data.

Identification of Dek phenotype

After maturity, wheat grains were harvested and threshed by hand. One hundred grains were randomly sampled from each individual, and the sampling was repeated three times. The wheat grains were visually examined for the Dek phenotype, and the incidence (percentage of grain Dek phenotype) was calculated. To better observe the wheat grain Dek phenotype, the grains of AK58 and *AK-3537* were stored at 4°C in FAA (Formalin-Aceto-Alcohol) fixative (ensuring the kernels did not float on the surface of the fixative solution) during the wheat grain filling stage. Subsequently, the cell structures of the normal phenotype and Dek phenotype were observed using frozen and free-hand sections (Leica CM1860) and x-ray computed tomography (Micro-CT).

The BSE-seq and BSR-seq for rapid map Dek gene

(*AK-3537* × AK58) F_1 grains were observed at the mature grain period in the natural field. F_2 populations were used for genetic analysis. We then performed a chi-square test (χ^2) to test phenotypic data (grain Dek phenotype) for a goodness of fit to the ratio of 3:1 expected for a single gene (or semi-dominant) genetic basis in Excel 2019 by CHISQ.TEST function ($p > 0.05$ means no deviation from expectations of 3:1). Using the combined approaches of BSE-seq and BSR-seq, AK58, *AK-3537*, and the F_2 population (grains with normal phenotype and Dek phenotype, with a mix of 30 random individuals each) were selected for DNA and RNA segregant pools. Leaves were collected for DNA extraction using the Plant Genomics DNA Extraction Kit (BIOFIT®, DN32-100, Chengdu, China), and grains (10–20 days post-anthesis, DPA) were sampled to extract total RNA using the Plant RNA Extraction Kit (BIOFIT®, RN34050, Chengdu, China). RNA sequencing generating 150 bp paired-end reads was performed on the Illumina HiSeq™ × platform. Clean RNA-seq data were mapped onto the Chinese Spring reference genome (RefSeq 2.1) and the AK58 genome (Jia J et al., 2021) using the software Bowtie2, and SNP calling was performed with the SAMtools software. The newly designed exome capture probe panel (Dong et al., 2020) and ΔSNP-index algorithm (Abe et al., 2012) were used to map the grain Dek gene in wheat rapidly. High-quality reads were aligned to the Chinese Spring reference genome (RefSeq 2.1) and the AK58 genome (Jia J et al., 2021) with default parameters. The parental AK58 sequencing data (DNA and RNA sequencing data) were used as a "background" to identify the causal mutation based on the assumption. Calculations were analyzed on PlantGmap (Zhang et al., 2021) (http://183.223.252.63:3333/).

Lastly, considering the characteristics of EMS mutagenesis, certain variations were filtered (Dong et al., 2020). Only the candidate regions identified in BSE-seq and BSR-seq were considered. Then, the Chinese Spring reference genome (RefSeq 2.1) was used to obtain the candidate gene, gene sequence, and gene annotation. Homologous analysis and gene expression patterns were evaluated on WheatOmics 1.0 (http://202.194.139.32/). Arabidopsis (TAIR10) and rice (IRGSP-1.0) genomes were used for comparative genomics analyses. A phylogenetic tree was constructed using MEGA11, and Geneious was employed to assemble high-quality reads.

Differentially expressed gene analysis

Grains of *AK-3537* and AK58 during the grain filling stage (10 DPA) were collected. RNA extraction, library construction, and RNA sequencing were performed as described previously. Differentially expressed gene (DEG) analysis, Gene Ontology (GO) annotation, and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were adopted to confirm the putative biological functions and biochemical pathways of DEGs (*AK-3537* and AK58) on the OmicsShare Tools (https://www.omicshare.com/).

Validation of KASP markers and quantitative reverse transcription polymerase chain reaction

DNA extraction was performed as previously described, and KASP markers were used for genotyping in parents and the F_2 population. The experimental method for KASP markers was referenced from previous reports (Li et al., 2020). The quantitative reverse transcription polymerase chain reaction (qRT-PCR) experiment was conducted to detect the expression levels of key genes involved in the wheat starch synthesis pathway, including *GBSSI*, *GBSSII*, *SBEI, SBEII, SSI, SSII, SSIIIa, SSIV, TaAGPL1, TaAGPS1a, TaBEI, TaBEIIa, TaBEIIb*, and *TaBEIII*. The experimental method for qRT-PCR was referenced from previous reports (Li et al., 2022). *Actin* was selected as the reference gene (Wang et al., 2014), and the relative quantification formula ($2^{-\Delta\Delta}$Ct) ± standard error of the mean (SEM) was used to evaluate quantitative variation further. Three biological replicates were tested for each sample. All primers used in this study are listed in Supplementary Table 1.

Results

Characterization of the Dek mutant phenotype

In a screening of wheat grain, we identified a Dek mutant, *AK-3537*, with grain shrunken from an ethyl methyl sulfide (EMS) mutant library in the AK58 background (Figure 1A). The grains of the wild-type AK58 were plump and normal (AK58 type), whereas *AK-3537* showed a 100% grain Dek phenotype rate across all environments (*AK-3537* type) (Figure 1). To further establish the genetic basis of this phenotype, a cross was performed between *AK-3537* and AK58. All F_1 plants exhibited a normal grain phenotype, suggesting the presence of a recessive gene controlling the wheat grain Dek phenotype. To assess *AK-3537* mutation segregation, (*AK-3537* × AK58) F_1 plants were self-pollinated and an F_2 segregating population (130 individuals) was developed. The F_2 population was segregated into two categories: 95 plants exhibited an AK58-type phenotype, and 35 plants showed an *AK-3537*-type phenotype (Table 1). The results were in agreement with a 3:1 segregation ratio [$\chi^2 = 0.26 < \chi^2(0.05, 1) = 3.84$, $p = 0.61$], suggesting that a single gene regulates grain Dek.

Compared with that in AK58, a large hollow area was observed in the endosperm of *AK-3537* by frozen and free-hand section, and the hollow phenotype was also observed in the Dek grain of *AK-3537* by X-ray 3D tomography (Figure 1B). The results showed that *AK-3537* grain filling was significantly affected. Finally, we evaluated the agronomic phenotypes of AK58 and *AK-3537* in the field. Compared with AK58, FLW, TN, TKW, GL, and GW were all significantly decreased in *AK-3537*, whereas HD increased considerably (Figures 2A, B).

Dek gene mapping by BSE-seq and BSR-seq

For Dek gene mapping, two pooled samples, each comprising 30 normal and Dek phenotype F_2 segregants, were constructed. Approximately 20 Gb of sequence data for the DNA pool and approximately 6 Gb for the RNA pool were generated and compared with the Chinese Spring reference genome (RefSeq V2.1) and AK58 genome. Then, we used the ΔSNP-index algorithm (Abe et al., 2012) for Dek gene mapping. ΔSNP-index higher than 0.7 in BSE-seq and BSR-seq Dek gene candidate region (DCR) was defined conservatively as the union of BSA-seq and BSR-seq credible intervals for candidate gene identification. In the AK58 genome, two DCRs, DCR1 and DCR2, were mapped on chromosome 7A by BSE-seq and BSR-seq (Table 2), among which DCR1 was located between 279.98 and 287.93 Mb on Chr7A with two SVs (structural variation) in two genes, and DCR2 was located between the range of 565.34 and 568.58 Mb on Chr7A with two SVs in two genes (Table 2). In the Chinese Spring reference genome (RefSeq V2.1), five DCRs, namely, DCR3, DCR4, DCR5, DCR6, and DCR7, were identified on chromosomes 1B, 7A, and 7B. DCR6 and DCR2 are the same candidate region and contain the same variant genes (Table 2). Because AK58 is a wheat–rye 1B/1R translocation line material, the genetic background of the 1B chromosome is very different between the Chinese Spring reference genome (RefSeq V2.1) and the AK58 genome (Jia J et al., 2021). It has also been reported that wheat chromosome 7B underwent structural rearrangement (Chen et al., 2020). Therefore, we speculated that DCR1 and DCR2 located on chromosome 7A were candidate regions for the wheat grain Dek phenotype, which was predicted to be a moderate functional effect (e.g., missense mutation).

Since a single gene regulates the Dek phenotype, we annotated the genes in the two candidate regions based on the wheat Chinese Spring RefSeq v2.1 genome. In DCR1, a gene encoding 3-hydroxy-3-methylglutaryl-CoA synthase (*TraesCS7A03G0625900*) and a gene encoding a kinesin-like protein (*TraesCS7A03G0631200*) were annotated. In DCR2, an FBD-associated F-box protein (*TraesCS7A03G0922700*) and an S-adenosyl-L-methionine-dependent methyltransferase superfamily protein (*TraesCS7A03G0929200*) were annotated (Figure 3; Table 2). Expression patterns of the candidate genes in the WheatOmics 1.0 (http://202.194.139.32/) database revealed that all four candidate genes were expressed in seeds.

To further exclude SNP variations caused by sequence assembly errors, exome capture and RNA-seq data were used to assemble the sequences of the mutated genes in the DCR1 and DCR2 regions. The results showed that the SNP variations at *TraesCS7A03G0631200* and

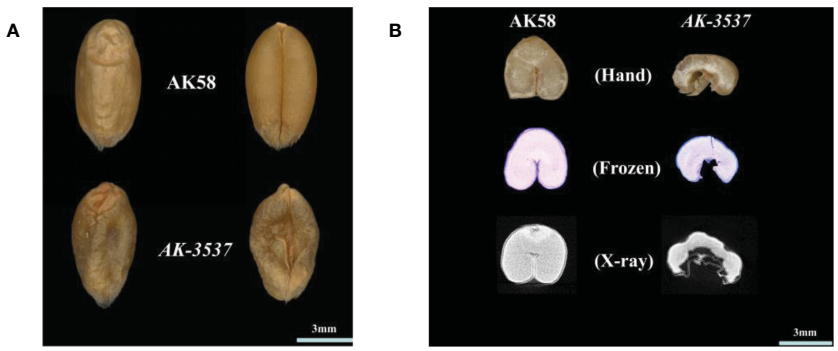


FIGURE 1
Wheat grain Dek phenotype and the internal structure of AK58 and *AK-3537*. **(A)** Dek phenotype of *AK-3537* mature grain compared to AK58 normal grain. **(B)** The internal structure of AK58 and *AK-3537* mature grains was observed by free-hand section (Hand), frozen section (Frozen), and x-ray tomography (X-ray).

TABLE 1 Grain phenotype statistics of AK58 and *AK-3537* parents and F_2 population.

	Name	Phenotype		Rate
Parents		AK58 type	*AK-3537* type	
	AK58	10		100%
	AK-3537		10	100%
F_2 populatoin	*AK-3537* × AK58	95	35	3:01

TraesCS7A03G0922700 could be detected in both exome capture and RNA-seq data. However, the SNP of *TraesCS7A03G0929200* could only be detected in exome capture data, while the SNP of *TraesCS7A03G0625900* could only be detected in RNA-seq data (Figure 4; Supplementary Figure 1). To validate the segregation of these SNP variations in the F_2 population, KASP markers were developed based on these four SNP variations in this study and validated in the F_2 population. The results showed that only the SNP

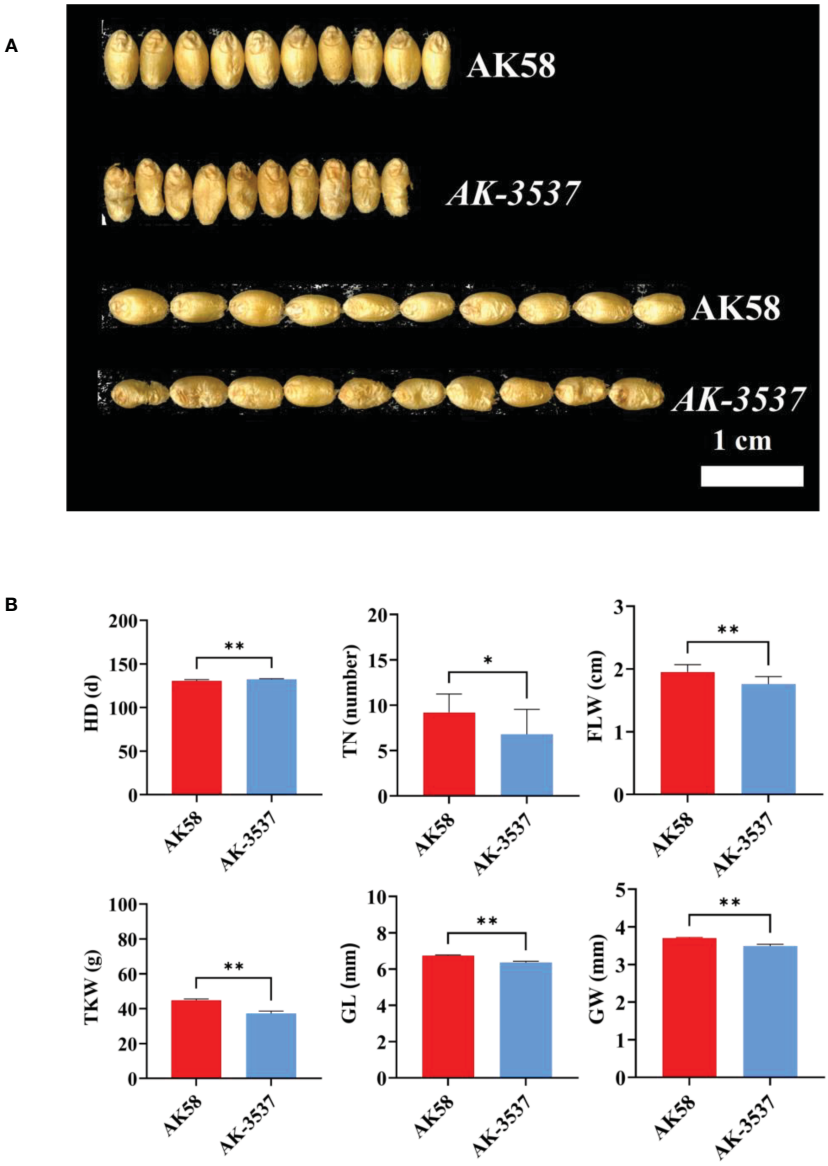


FIGURE 2
Comparison of grain and field phenotypes between AK58 and *AK-3537*. **(A)** Comparison of length and width phenotypes of *AK-3537* defective kernel and AK58 normal grain. **(B)** Statistical analysis of field phenotypes and grain phenotypes of AK58 and *AK-3537* including tiller number (TN), heading date (HD), flag leaf width (FLW), thousand kernel weight (TKW), grain length (GL), and grain width (GW). "*" means $p < 0.05$ and "**" means $p < 0.01$.

10.3389/fpls.2023.1173861

TABLE 2 The wheat grain Dek candidate region in the AK58 genome and the Chinese Spring reference genome (RefSeq 2.1).

Region	Type	Chr	Pos	Gene	ΔSNP-index	Variant	Annotation	Genome
DCR1	BSR-seq	7A	279.98	*TraesCS7A03G0625900*	0.733	c.1049G>A	3-hydroxy-3-methylglutaryl-CoA synthase	AK58
	BSE-seq	7A	287.92	*TraesCS7A03G0631200*	0.733	c.1766C>T	Kinesin, motor region domain containing protein	
DCR2	BSR-seq	7A	565.34	*TraesCS7A03G0922700*	0.789	c.163G>A	FBD-associated F-box protein	
	BSE-seq	7A	568.58	*TraesCS7A03G0929200*	0.789	c.535G>A	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	
DCR3	BSE-seq	1B	0.88	*TraesCS1B03G0001300*	0.714	c.271G>A	Disease resistance protein (NBS-LRR class) family	Chinese spring
	BSE-seq	1B	1.80	*TraesCS1B03G0005100*	0.763	c.1553G>A	NBS-LRR-like resistance protein	
	BSE-seq	1B	4.14	*TraesCS1B03G0014000*	0.833	c.689G>A	Receptor protein kinase	
	BSE-seq	1B			0.7	c.723G>A	Receptor protein kinase	
	BSR-seq	1B	27.65	*TraesCS1B03G0099800LC*	0.783	c.106C>T	Polycystic kidney disease protein 1-like 2	
DCR4	BSE-seq	1B	125.71	*TraesCS1B03G0289600*	0.857	c.712G>A	Agmatine coumaroyltransferase-1	
	BSE-seq	1B	127.61	*TraesCS1B03G0293800*	0.875	c.1373C>T	Kinase family protein	
	BSE-seq	1B	137.88	*TraesCS1B03G0305800*	0.727	c.950C>T	NAC domain protein	
	BSE-seq	1B			0.727	c.949C>T	NAC domain protein	
	BSR-seq	1B	151.77	*TraesCS1B03G0332400*	0.739	c.569G>A	Translocase of chloroplast 159	
	BSR-seq	1B			0.739	c.-86G>A		
	BSR-seq	1B			0.739	c.568G>A		
	BSR-seq	1B			0.739	c.-87G>A		
	BSR-seq	1B			0.75	c.386C>T		
	BSR-seq	1B			0.75	c.-269C>T		
	BSR-seq	1B			0.75	c.376C>T		
	BSR-seq	1B			0.75	c.-279C>T		
DCR5	BSR-seq	1B	207.08	*TraesCS1B03G0399700LC*	0.812	c.209G>A	Transposon protein	
	BSE-seq	1B	218.79	*TraesCS1B03G0407700LC*	0.75	c.403G>A	Gag-pol polyprotein	
DCR6	BSR-seq	7A	559.38	*TraesCS7A03G0922700*	0.765	c.163G>A	FBD-associated F-box protein	

(Continued)

TABLE 2 Continued

Region	Type	Chr	Pos	Gene	ΔSNP-index	Variant	Annotation	Genome
	BSE-seq	7A	562.69	*TraesCS7A03G0929200*	0.742	c.535G>A	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	
DCR7	BSE-seq	7B	761.91	*TraesCS7B03G1333300LC*	0.857	c.505C>T	Aspartate–tRNA(Asp/Asn) ligase	
	BSE-seq	7B			0.875	c.499G>A		
	BSR-seq	7B	763.24	*TraesCS7B03G1339800*	0.8	c.2653G>A	Rp1-like protein	

variation at position 1,049 of the *TraesCS7A03G0625900* coding region co-segregated with the grain normal and Dek phenotype in the F_2 population (Figure 5B). A query of the Chinese Spring RefSeq v2.1 genome found that the coding sequence of the *TraesCS7A03G0625900* gene was 1,410 bp in length, and the full-length genome contained 12 exons and 11 introns. Sequencing analysis of the *TraesCS7A03G0625900* coding region in *AK-3537* revealed a G-to-A mutation at position 1,049, the 11th exon of the coding region, which resulted in an amino acid substitution from glycine (GGC, Gly, G) to aspartic acid (GAC, Asp, D) (Figure 5A). The target gene was tentatively termed *HMGS-7A*.

Differentially expressed genes between AK58 and *AK-3537* grains

In the transcriptome analysis of wheat grains of AK58 and *AK-3537*, a total of 12,655 DEGs were identified. Enrichment analysis of these DEGs in the wheat grain Dek regulatory pathway showed significant differences in gene expression levels between the mutant *AK-3537* and the wild-type AK58, with 6,618 genes downregulated and 6,037 genes upregulated. GO and KEGG analyses were further performed on the screened DEGs to understand the functions and pathways of these DEGs. GO analysis showed that these DEGs were mainly concentrated in processes involved in carbohydrate metabolism (GO:0005975 and GO:0044723) (Figure 6A). KEGG analysis showed significant enrichment ($p \leq 0.05$) of energy metabolism and starch synthesis pathway (Ko00500) (Figure 6B). This indicates that, compared to AK58 grains, *AK-3537* grains have significant differences in wheat carbon metabolism, photosynthesis product synthesis, and starch synthesis pathways. These DEGs may cause the *AK-3537* grain Dek phenotype. In previous reports, starch synthase also regulated similar phenotypes (Morell et al., 2003; Li et al., 2011). Therefore, this study also found that the expression levels of *GBSSII* and *SSIIIa* in the Ko00500 pathway were significantly increased in *AK-3537* (Figure 7).

Discussion

Wheat grain is the most important component of wheat yield and quality. In this study, we identified a candidate gene,

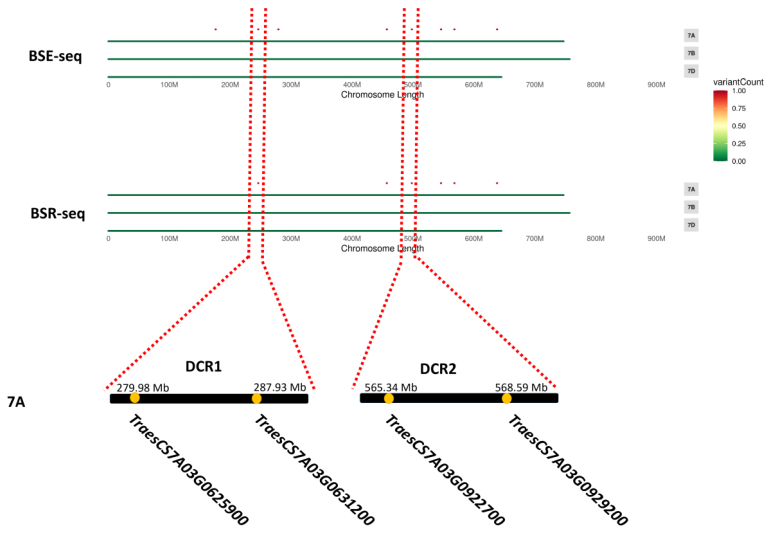


FIGURE 3
Most likely candidate region for wheat Dek detected via BSE-seq and BSR-seq. "." means the SNPs associated with wheat grain Dek phenotype after the screening.

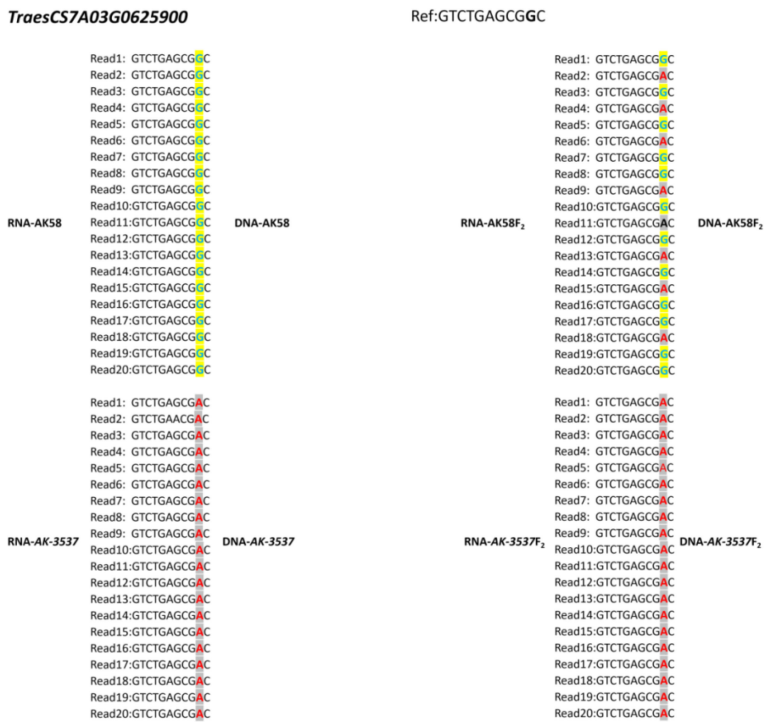


FIGURE 4
Candidate genes influencing wheat grain Dek phenotypes by the assembly of exome capture and RNA-seq sequencing data.

TraesCS7A03G0625900 (*HMGS-7A*), encoding a 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS), in AK58 and *AK-3537* by BSE-seq and BSR-seq, responsible for the Dek phenotype in wheat grain. The G-to-A mutation at position 1,049 of the *HMGS-7A* coding region leads to amino acid change from glycine (Gly) to aspartic acid (Asp). In previous studies, several genes related to grain size and weight have been mapped and cloned on wheat chromosome 7A, including *TaTPP-7A* (7A, 135.0 Mb; Liu et al., 2023), *TaGASR7* (7A, 176.0 Mb; Zhang et al., 2015), *TaTGW-7A* (7A, 211.6 Mb; Hu et al., 2016), *TaGW8* (7A, 257.3 Mb; Yan et al., 2019), and *TaIAA21* (7A, 488.5 Mb; Jia M et al., 2021). The *TraesCS7A03G0625900* (*HMGS-7A*) gene identified in this study

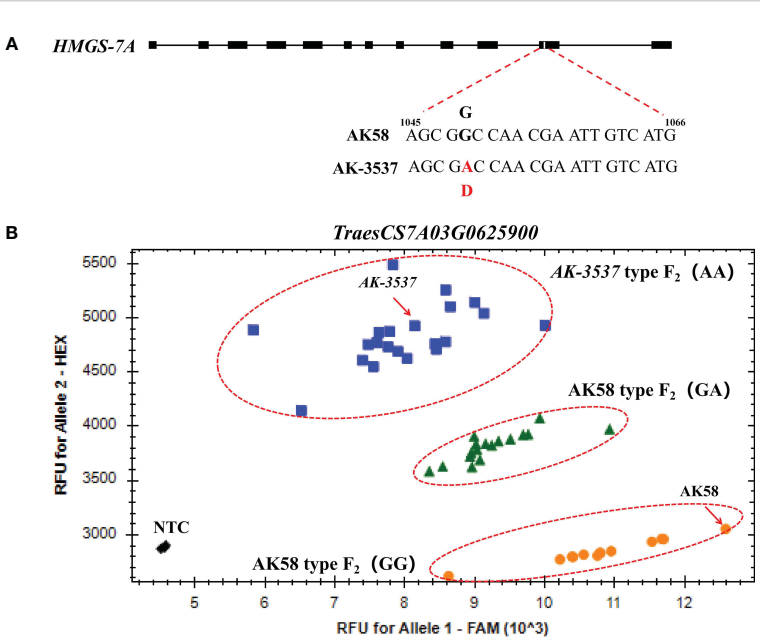


FIGURE 5
Mutation and marker verification of *HMGS-7A* gene. **(A)** Structure and mutation site of *HMGS-7A*. **(B)** *HMGS-7A*^{1049} was validated by the KASP marker.

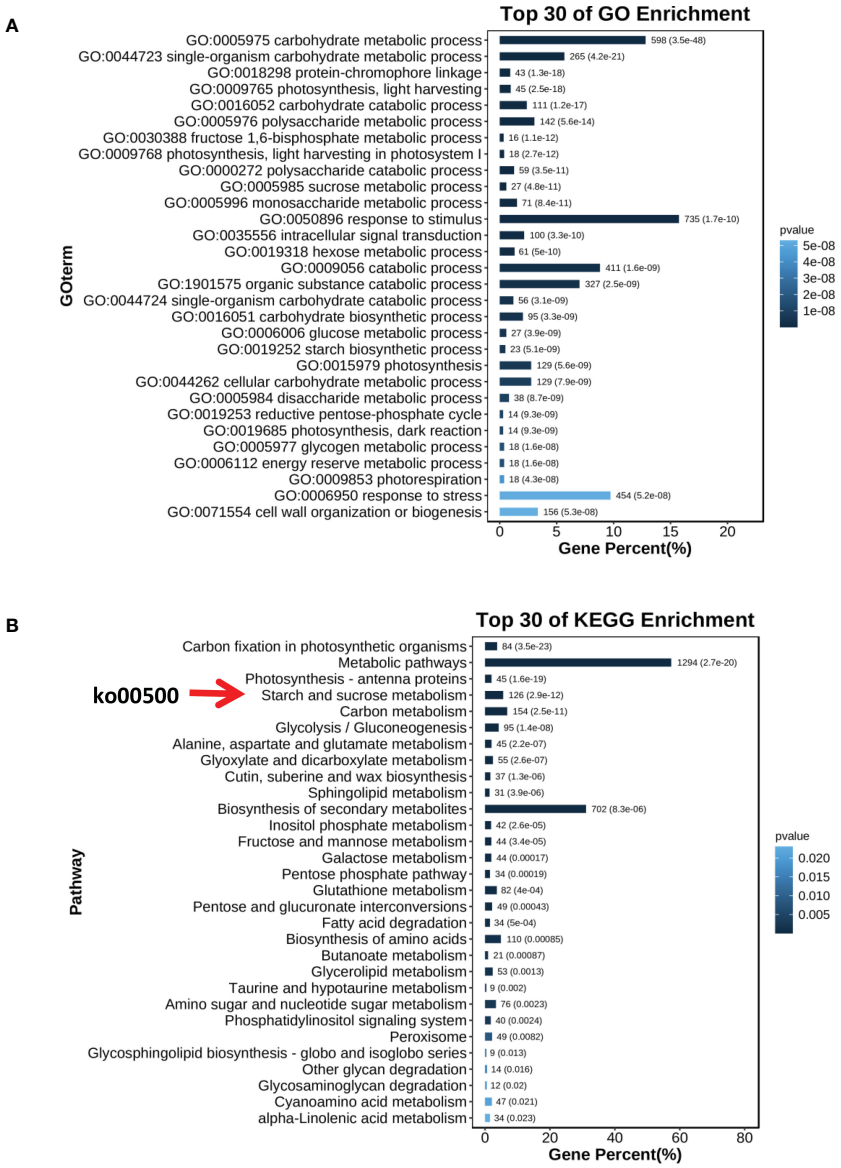


FIGURE 6
Differential expression analysis for *AK58* and *AK-3537* grain transcriptome. **(A)** GO analysis for DEGs. **(B)** KEGG analysis for DEGs.

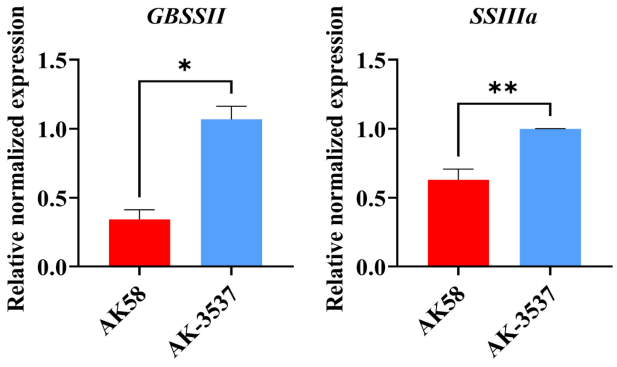


FIGURE 7
Statistical analysis of gene expression of starch synthetase *GBSSII* and *SSIIIa*.* means $p < 0.05$ and ** means $p < 0.01$.

does not overlap with these reported grain morphology genes. Therefore, it is speculated to be a novel gene controlling wheat grain Dek phenotype.

The coding sequences of *TraesCS7A03G0625900* (*HMGS-7A*), and its two homologous genes, *TraesCS7B03G0451200* (*HMGS-7B*) and *TraesCS7D02G269600* (*HMGS-7D*), were obtained from Ensembl plants (https://plants.ensembl.org/index.html) and used to construct a phylogenetic tree with *HMGS* coding sequences from rice, Arabidopsis, and Brassica (Figure 8). The results showed that the *HMGS* in wheat is more closely related to that in rice. The importance of *HMGS* in Arabidopsis has been demonstrated in steroid biosynthesis, pollen fertility, and seed weight (Ishiguro et al., 2010; Bhangu-Uhlmann, 2011; Lange et al., 2015). HMGS is the second key enzyme in the mevalonate (MVA) pathway, significantly affecting plant sterol biosynthesis (Wang et al., 2012; Liao et al., 2014a; Lange et al., 2015; Pérez et al., 2022). Brassinosteroids (BR) is

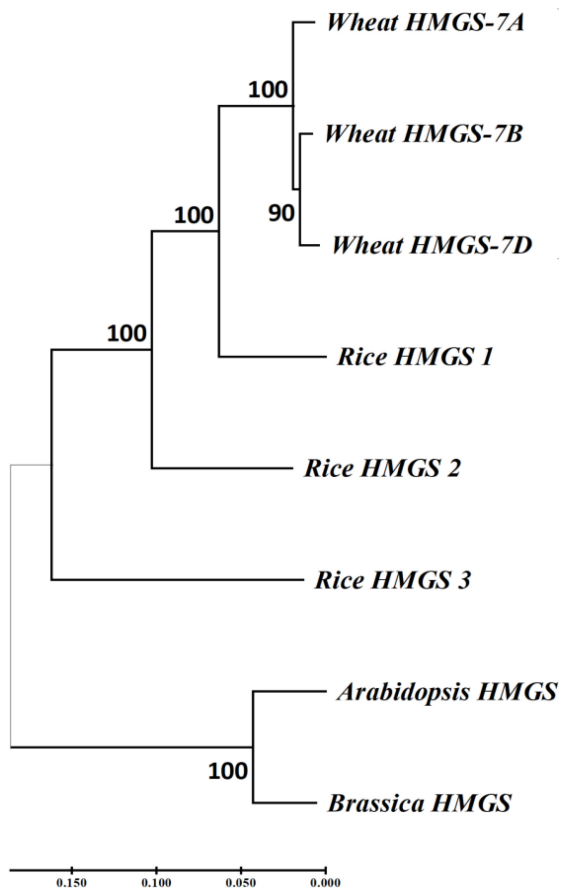


FIGURE 8
Phylogenetic analysis of *HMGS* homologous genes in wheat and *HMGS* in other plants.

one of the main types of sterols (Zhang et al., 2020), which plays a vital role in the grain-filling process of plants. Transferring the gene encoding C-22 hydroxylase, an enzyme involved in sterol biosynthesis, into rice significantly increased the BR hormone content and the TKW, and the increase in TKW resulted from BR stimulating the transport of photosynthates in rice (Wu et al., 2008). Knockout of *TaD11-2A* results in dwarfism, a significant decrease in endogenous BR content, and smaller grains in wheat (Xu et al., 2022). *TaBRI1* is the BR receptor gene in wheat, and *TaBRI1* knockout mutants were insensitive to exogenous BR and significantly reduced TKW (Fang et al., 2020). *GW5* is a positive regulator of BR signaling, expressed in various rice organs, considerably affecting the width and weight of rice grains, and is a feasible target for increasing grain yield in rice and other cereal crops through gene editing (Liu et al., 2017). *SMG3* and *DGS1* regulate the size and weight of rice grains through the BR signaling pathway. Loss of *SMG3* or *DGS1* function results in smaller grains, while overexpression of *SMG3* or *DGS1* leads to longer grains (Li et al., 2023).

The sterol content in plant grains may also affect the starch synthesis pathway. A study of *sbeIIb* mutants in rice showed that many starch synthesis enzyme genes were upregulated, except for genes encoding granule-bound starch synthase, branching enzyme, pullulanase, and starch phosphorylase, which were downregulated. This increased amylose and resistant starch content, in addition to an increase in many other substances such as sugar, fatty acids, amino acids, and plant sterols in the endosperm (Baysal et al., 2020), and the wheat mutant *SM482gs*, with increased grain size, TKW, and protein content with BR biosynthesis and signal transduction, were significantly upregulated, but *AGP-S1*, *AGP-L2*, *SSI*, *SSIIa*, *SSIIIa*, *SBEIIa*, *SBEIIb*, and *GBSSIa* show the lower expression on *SM482gs* (Zhong et al., 2021), which indicated that plant sterols might be involved in the synthesis of amylose in plant grains. In rice, overexpression of *HMGS* significantly increased fatty acids, abscisic acid, gibberellins, and lutein in transgenic rice (Pérez et al., 2022), while overexpression of *HMGS* in mustard significantly increased grain weight (Liao et al., 2014b). In barley, mutations in the starch synthase genes *SSIIa* and *SSIIIa* result in grain phenotypes similar to those observed in this study with *AK-3537* (Li et al., 2011). Therefore, in this study, we detected the expression levels of key genes involved in starch synthesis in seeds and found that the key gene *SSIIIa*, which regulates the content of amylose and amylopectin in plants, was highly expressed in *AK-3537*, indicating that the functional changes of *HMGS-7A* may affect the expression of key enzyme genes involved in wheat starch synthesis. In the future, we will further analyze *HMGS-7A* and verify the role of *HMGS-7A* in wheat grain filling.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

JW and HT designed research; HT performed experiments; HT and MC analyzed the data, ML, ZY, and ZP prepared the plant materials; HT, XG, HD, QC, and JW wrote and revised the paper; JW supervised the project. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Major Program of National Agricultural Science and Technology of China (NK20220607), the National Natural Science Foundation of China (U22A20472), the National Key Research and Development Program of China (2018YFE0112000), the Sichuan Science and Technology Support Project (2021YFH0077, 2021YFYZ0027, and 23NSFSC0770), the Science and Technology Support Project of Chengdu (2021-GH03-

00002-HZ), and the open research fund of SKL-CGEUSC (SKL-ZD202212).

Acknowledgments

The authors thank the reviewers and editors for their suggestions to improve this manuscript. The authors thank Professor Jizeng Jia from the Institute of Crop Science, Chinese Academy of Agricultural Science, for sharing the wheat mutation lines.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1173861/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Candidate genes influencing wheat grain Dek phenotypes by assembly of exon capture and RNA-seq sequencing data.

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., et al. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* 30, 174–178. doi: 10.1038/nbt.2095

Ahmed, N., Tetlow, I. J., Nawaz, S., Iqbal, A., Mubin, M., Rehman, S., et al. (2015). Effect of high temperature on grain filling period, yield, amylose content and activity of starch biosynthesis enzymes in endosperm of basmati rice. *J. Sci. Food Agric.* 95, 2237–2243. doi: 10.1002/jsfa.6941

Barneix, A. J. (2007). Physiology and biochemistry of source-regulated protein accumulation in the wheat grain. *J. Plant Physiol.* 164, 581–590. doi: 10.1016/j.jplph.2006.03.009

Baysal, C., He, W., Drapal, M., Villorbina, G., Medina, V., Capell, T., et al. (2020). Inactivation of rice starch branching enzyme IIb triggers broad and unexpected changes in metabolism by transcriptional reprogramming. *Proc. Natl. Acad. Sci. U. S. A.* 117, 26503–26512. doi: 10.1073/pnas.2014860117

Bhangu-Uhlmann, A. (2011). "The mevalonate pathway: a monitoring approach in plants by systems biology tools," in *Doctoral dissertation, ETH Zurich* (Zurich, Swiss Confederation: ETH Zurich Research Collection). doi: 10.3929/ethz-a-006690303

Chen, X., Feng, F., Qi, W., Xu, L., Yao, D., Wang, Q., et al. (2017). *Dek35* encodes a PPR protein that affects cis-splicing of mitochondrial *nad4* intron 1 and seed development in maize. *Mol. Plant* 10, 427–441. doi: 10.1016/j.molp.2016.08.008

Chen, Y., Song, W., Xie, X., Wang, Z., Guan, P., Peng, H., et al. (2020). A collinearity-incorporating homology inference strategy for connecting emerging assemblies in the *triticeae tribe* as a pilot practice in the plant pangenomic era. *Mol. Plant* 13, 1694–1708. doi: 10.1016/j.molp.2020.09.019

Dai, D., Luan, S., Chen, X., Wang, Q., Feng, Y., Zhu, C., et al. (2018). Maize *Dek37* encodes a p-type PPR protein that affects cis-splicing of mitochondrial nad2 intron 1 and seed development. *Genetics* 208, 1069–1082. doi: 10.1534/genetics.117.300602

Djanaguiraman, M., Narayanan, S., Erdayani, E., and Prasad, P. V. (2020). Effects of high temperature stress during anthesis and grain filling periods on photosynthesis, lipids and grain yield in wheat. *BMC Plant Biol.* 20, 1–12. doi: 10.1186/s12870-020-02479-0

Dong, C., Zhang, L., Chen, Z., Xia, C., Gu., Y., Wang, J., et al. (2020). Combining a new exome capture panel with an effective varbscore algorithm accelerates bsa-based gene cloning in wheat. *Plant Sci.* 11. doi: 10.3389/fpls.2020.01249

Ebrahimnejad, S., and Rameeh, V. (2016). Correlation and factor analysis of grain yield and some important component characters in spring bread wheat genotypes. *Cerce. Agron. Mol.* 49, 5–15. doi: 10.1515/cerce-2016-0001

Fang, J., Zhu, W., and Tong, Y. (2020). Knock-down the expression of brassinosteroid receptor *TaBRI1* reduces photosynthesis, tolerance to high light and high temperature stresses and grain yield in wheat. *Plants* 9, 840. doi: 10.3390/plants9070840

Fu, C., Du, J., Tian, X., He, Z., Fu, L., Wang, Y., et al. (2019). Rapid identification and characterization of genetic loci for defective kernel in bread wheat. *BMC Plant Biol.* 19, 483. doi: 10.1186/s12870-019-2102-6

Garcia, N., Li, Y., Dooner, H. K., and Messing, J. (2017). Maize defective kernel mutant generated by insertion of a *ds* element in a gene encoding a highly conserved *TTI2* cochaperone. *Proc. Natl. Acad. Sci. U. S. A.* 114, 5165–5170. doi: 10.1073/pnas.1703498114

Guo, M., Zhao, G., Guo, W., Chang, X., Wang, D., Yang, Y., et al. (2015). Effect of different sowing dates on characteristics of flag leaf photosynthesis and grain filling of wheat cultivars with different gluten. *J. Triticeae Crops.* 35, 192–197. doi: 10.7606/j.issn.1009-1041.2015.02.07

He, Y., Wang, J., Qi, W., and Song, R. (2019). Maize *dek15* encodes the cohesin-loading complex subunit *SCC4* and is essential for chromosome segregation and kernel development. *Plant Cell* 31, 465–485. doi: 10.1105/tpc.18.00921

Housley, T. L., Kirleis, A. W., Ohm, H. W., and Patterson, F. L. (1981). An evaluation of seed growth in soft red winter wheat. *Can. J. Plant Sci.* 61, 525–534. doi: 10.4141/cjps81-075

Hu, M. J., Zhang, H. P., Liu, K., Cao, J. J., Wang, S. X., Jiang, H., et al. (2016). Cloning and characterization of TaTGW-7A gene associated with grain weight in wheat via SLAF-seq-BSA. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01902

Ishiguro, S., Nishimori, Y., Yamada, M., Saito, H., Suzuki, T., Nakagawa, T., et al. (2010). The arabidopsis *FLAKY POLLEN* 1 gene encodes a 3-hydroxy-3-methylglutaryl-coenzyme a synthase required for development of tapetum-specific organelles and fertility of pollen grains. *Plant Cell Physiol.* 51, 896–911. doi: 10.1093/pcp/pcq068

Jia, M., Li, Y., Wang, Z., Tao, S., Sun, G., Kong, X., et al. (2021). *TaIAA21* represses TaARF25-mediated expression of *TaERFs* required for grain size and weight development in wheat. *Plant J.* 108 (6), 1754–1767. doi: 10.1111/tpj.15541

Jia, J., Xie, Y., Cheng, J., Kong, C., Wang, M., Gao, L., et al. (2021). Homology-mediated inter-chromosomal interactions in hexaploid wheat lead to specific subgenome territories following polyploidization and introgression. *Genome Biol.* 22, 1–21. doi: 10.1186/s13059-020-02225-7

Jiang, X., Cheng, H., Jiang, L., Yang, X., Lv, R., Hua, M., et al. (2017). Effect of increasing diffuse radiation fraction under low light condition on the grain-filling process of winter wheat (*Triticum aestivum* l.). *Chin. J. Agrom.* 38, 753–760. doi: 10.3969/j.issn.1000-6362.2017.12.001

Khatun, S., and Ahmed, J.U (2015). Response of elevated temperature on carbohydrate accumulation and grain yield in different wheat cultivars. *Bangladesh J. Agril. Res.* 40, 205–215. doi: 10.3329/bjar.v40i2.24558

Lange, I., Poirier, B. C., Herron, B. K., and Lange, B. M. (2015). Comprehensive assessment of transcriptional regulation facilitates metabolic engineering of isoprenoid accumulation in arabidopsis. *Plant Physiol.* 169, 1595–1606. doi: 10.1104/pp.15.00573

Li, J., Fu, J., Chen, Y., Fan, K., He, C., Zhang, Z., et al. (2017). The *U6 biogenesis-like* 1 plays an important role in maize kernel and seedling development by affecting the 3′ end processing of *U6* snRNA. *Mol. Plant* 10, 470–482. doi: 10.1016/j.molp.2016.10.016

Li, X., Gu, W., Sun, S., Chen, Z., Chen, J., Song, W., et al. (2018). *Defective kernel* 39 encodes a PPR protein required for seed development in maize. *J. Integr. Plant Biol.* 60, 45–64. doi: 10.1111/jipb.12602

Li, Z., Li, D., Du, X., Wang, H., Larroque, O., Jenkins, C. L., et al. (2011). The barley amo1 locus is tightly linked to the starch synthase IIIa gene and negatively regulates expression of granule-bound starch synthetic genes. *J. Exp. Bot.* 62, 5217–5231. doi: 10.1093/jxb/err239

Li, C., Liu, H., Wang, J., Pan, Q., Wang, Y., Wu, K., et al. (2022). Characterization and fine mapping of a lesion mimic mutant (*Lm5*) with enhanced stripe rust and

powdery mildew resistance in bread wheat (*Triticum aestivum* l.). *Theor. Appl. Genet.* 135, 1–18. doi: 10.1007/s00122-021-03973-1

Li, C., Tang, H., Luo, W., Zhang, X., Mu, Y., Deng, M., et al. (2020). A novel, validated, and plant height-independent QTL for spike extension length is associated with yield-related traits in wheat. *Theor. Appl. Genet.* 133, 3381–3393. doi: 10.1007/s00122-020-03675-0

Li, J., Zhang, B., Duan, P., Yan, L., Yu, H., Zhang, L., et al. (2023). An endoplasmic reticulum-associated degradation–related E2–E3 enzyme pair controls grain size and weight through the brassinosteroid signaling pathway in rice. *Plant Cell* 35, 1076–1091. doi: 10.1093/plcell/koac364

Liao, P., Wang, H., Hemmerlin, A., Nagegowda, D. A., Bach, T. J., and Wang, M.. (2014a). Past achievements, current status and future perspectives of studies on 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) in the mevalonate (MVA) pathway. *Plant Cell Rep.* 33, 1005–1022. doi: 10.1007/s00299-014-1592-9

Liao, P., Wang, H., Wang, M., Hsiao, A. S., Bach, T. J., Chye, M. L., et al. (2014b). Transgenic tobacco overexpressing *Brassica juncea HMG-CoA synthase* 1 shows increased plant growth, pod size and seed yield. *PloS One* 9, e98264. doi: 10.1371/journal.pone.0108026

Lid, S. E., Gruis, D., Jung, R., Lorentzen, J. A., Ananiev, E., Chamberlin, M., et al. (2002). The *defective kernel* 1 (*dek1*) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5460–5465. doi: 10.1073/pnas.042098799

Liu, J., Chen, J., Zheng, X., Wu, F., Lin, Q., Heng, Y., et al. (2017). *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat. Plants* 3, 1–7. doi: 10.1038/nplants.2017.43

Liu, Y., Gu, D., Wu, W., Wen, X., and Liao, Y. (2013). The relationship between polyamines and hormones in the regulation of wheat grain filling. *PloS One* 8, e78196. doi: 10.1371/journal.pone.0078196

Liu, Y., Liang, H., Lv, X., Liu, D., Wen, X., and Liao, Y. (2016). Effect of polyamines on the grain filling of wheat under drought stress. *Plant Physiol. Biochem.* 100, 113–129. doi: 10.1016/j.plaphy.2016.01.003

Liu, Y., Lin, Y., Gao, S., Li, Z., Ma, J., Deng, M., et al. (2017). A genome-wide association study of 23 agronomic traits in Chinese wheat landraces. *Plant J.* 91, 861–873. doi: 10.1111/tpj.13614

Liu, H., Si, X., Wang, Z., Cao, L., Gao, L., Zhou, X., et al. (2023). *TaTPP-7A* positively feedback regulates grain filling and wheat grain yield through T6P-SnRK1 signalling pathway and sugar–ABA interaction. *Plant Biotechnol. J.*, 1–17. doi: 10.1111/pbi.14025

Ma, J., Liu, Y., Zhang, P., Chen, T., Tian, T., and Wang, P. (2022). Identification of quantitative trait loci (QTL) and meta-QTL analysis for kernel size-related traits in wheat (*Triticum aestivum* l.). *BMC Plant Biol.* 22, 1–18. doi: 10.1186/s12870-022-03989-9

Mancebo, C. M., Merino, C., Martínez, M. M., and Gómez, M. (2015). Mixture design of rice flour, maize starch and wheat starch for optimization of gluten free bread quality. *J. Food Sci. Technol.* 52, 6323–6333. doi: 10.1007/s13197-015-1769-4

Mirosavljević, M., Mikić, S., Župunski, V., Kondić Špika, A., Trkulja, D., Ottosen, C. O., et al. (2021). Effects of high temperature during anthesis and grain filling on physiological characteristics of winter wheat cultivars. *J. Agron. Crop Sci.* 207, 823–832. doi: 10.1111/jac.12546

Morell, M. K., Kosar-Hashemi, B., Cmiel, M., Samuel, M. S., Chandler, P., Rahman, S., et al. (2003). Barley sex6 mutants lack starch synthase IIa activity and contain a starch with novel properties. *Plant J.* 34, 173–185. doi: 10.1046/j.1365-313X.2003.01712.x

Ogbonnaya, F. C., Abdalla, O., Mujeebkazi, A., Kazi, A. G., Xu, S., Gosman, N., et al. (2013). Synthetic hexaploids: harnessing species of the primary gene pool for wheat improvement. *Plant Breed. Rev.* 37, 35–122. doi: 10.1002/9781118497869.ch2

Pérez, L., Alves, R., Perez-Fons, L., Albacete, A., Farré, G., Soto, E., et al. (2022). Multilevel interactions between native and ectopic isoprenoid pathways affect global metabolism in rice. *Transgenic Res.* 31, 249–268. doi: 10.1007/s11248-022-00299-6

Qi, W., Lu, L., Huang, S., and Song, R. (2019). Maize *dek44* encodes mitochondrial ribosomal protein l9 and is required for seed development. *Plant Physiol.* 180, 2106–2119. doi: 10.1104/pp.19.00546

Qi, W., Tian, Z., Lu, L., Chen, X., Chen, X., Zhang, W., et al. (2017). Editing of mitochondrial transcripts nad3 and cox2 by *Dek10* is essential for mitochondrial function and maize plant development. *Genetics* 205, 1489–1501. doi: 10.1534/genetics.116.199331

Ren, R., Lu, X., Zhao, Y., Wei, Y., Wang, L., Zhang, L., et al. (2019). Pentatricopeptide repeat protein *DEK40* is required for mitochondrial function and kernel development in maize. *J. Exp. Bot.* 70, 6163–6179. doi: 10.1093/jxb/erz391

Shiferaw, B., Smale, M., Braun, H. J., Duveiller, E., Reynolds, M., and Muricho, G. (2013). Crops that feed the world 10. past successes and future challenges to the role played by wheat in global food security. *Food Secur.* 5, 291–317. doi: 10.1007/s12571-013-0263-y

Song, X., Hu, S., Zhang, K., Cui, Z., Li, J., Yang, X., et al. (2020). Phenotypic analysis and fine mapping of *dek101* in maize. *Acta Agron. Sin.* 46, 1831–1838. doi: 10.3724/SP.J.1006.2020.03017

Wang, Z., Li, W., Qi, J., Shi, P., and Yin, Y. (2014). Starch accumulation, activities of key enzyme and gene expression in starch synthesis of wheat endosperm with different starch contents. *J. Food Sci. Technol.* 51, 419–429. doi: 10.1007/s13197-011-0520-z

Wang, H., Nagegowda, D. A., Rawat, R., Bouvier-Navé, P., Guo, D., Bach, T. J., et al. (2012). Overexpression of brassica juncea wild-type and mutant *HMG-CoA synthase* 1 in arabidopsis up-regulates genes in sterol biosynthesis and enhances sterol production and stress tolerance. *Plant Biotechnol. J.* 10, 31–42. doi: 10.1111/j.1467-7652.2011.00631.

Wang, G., Zhong, M., Shuai, B., Song, J., Zhang, J., Han, L., et al. (2017). E+ subgroup PPR protein defective kernel 36 is required for multiple mitochondrial transcripts editing and seed development in maize and arabidopsis. *New Phytol.* 214, 1563–1578. doi: 10.1111/nph.14507

Wu, C. Y., Trieu, A., Radhakrishnan, P., Kwok, S. F., Harris, S., Zhang, K., et al. (2008). Brassinosteroids regulate grain filling in rice. *Plant Cell* 20, 2130–2145. doi: 10.1105/tpc.107.055087

Xiong, M., Feng, G., Gao, Q., Zhang, C., Li, Q., and Liu, Q. (2022). Brassinosteroid regulation in rice seed biology. *Seed Biol.* 1, 1–9. doi: 10.48130/SeedBio-2022-0002

Xu, H., Sun, H., Dong, J., Ma, C., Li, J., Li, Z., et al. (2022). The brassinosteroid biosynthesis gene *TaD11-2A* controls grain size and its elite haplotype improves wheat grain yields. *Theor. Appl. Genet.* 135, 2907–2923. doi: 10.1007/s00122-022-04158-0

Yan, X., Zhao, L., Ren, Y., Dong, Z., Cui, D., and Chen, F. (2019). Genome-wide association study revealed that the *TaGW8* gene was associated with kernel size in Chinese bread wheat. *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-38570-2

Zhang, L., Dong, C., Chen, Z., Gui, L., Chen, C., Li, D., et al. (2021). WheatGmap: a comprehensive platform for wheat gene mapping and genomic studies. *Mol. Plant* 14, 187–190. doi: 10.1016/j.molp.2020.11.018

Zhang, X., Lin, K., and Li, Y. (2020). Highlights to phytosterols accumulation and equilibrium in plants: biosynthetic pathway and feedback regulation. *Plant Physiol. Biochem.* 155, 637–649. doi: 10.1016/j.plaphy.2020.08.021

Zhang, D., Wang, B., Zhao, J., Zhao, X., Zhang, L., Liu, D., et al. (2015). Divergence in homoeolog expression of the grain length-associated gene *GASR7* during wheat allohexaploidization. *Crop J.* 3, 1–9. doi: 10.1016/j.cj.2014.08.005

Zhong, X., Lin, N., Ding, J., Yang, Q., Lan, J., Tang, H., et al. (2021). Genome-wide transcriptome profiling indicates the putative mechanism underlying enhanced grain size in a wheat mutant. *3 Biotech.* 11, 1–16. doi: 10.1007/s13205-020-02579-6

Zhu, C., Jin, G., Fang, P., Zhang, Y., Feng, X., Tang, Y., et al. (2019). Maize pentatricopeptide repeat protein DEK41 affects cis-splicing of mitochondrial *nad4* intron 3 and is required for normal seed development. *J. Exp. Bot.* 70, 3795–3808. doi: 10.1093/jxb/erz193

Zuo, Y., Feng, F., Qi, W., and Song, R. (2019). *Dek42* encodes an RNA-binding protein that affects alternative pre-mRNA splicing and maize kernel development. *J. Integr. Plant Biol.* 61, 728–748. doi: 10.1111/jipb.12798

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 June 2023
DOI 10.3389/fpls.2023.1148658

OPEN ACCESS

EDITED BY
Shouvik Das,
Regional Centre for Biotechnology (RCB), India

REVIEWED BY
Revathi Ponnuswamy,
Indian Institute of Rice Research (ICAR), India
Kumar Paritosh,
University of Delhi, India

*CORRESPONDENCE
Amit Kumar Singh
amit.singh5@icar.gov.in
Mahesh C. Yadav
mahesh.yadav1@icar.gov.in

†These authors have contributed equally to this work

RECEIVED 20 January 2023
ACCEPTED 11 April 2023
PUBLISHED 28 June 2023

CITATION
Kumari J, Lakhwani D, Jakhar P, Sharma S, Tiwari S, Mittal S, Avashthi H, Shekhawat N, Singh K, Mishra KK, Singh R, Yadav MC, Singh GP and Singh AK (2023) Association mapping reveals novel genes and genomic regions controlling grain size architecture in mini core accessions of Indian National Genebank wheat germplasm collection. *Front. Plant Sci.* 14:1148658. doi: 10.3389/fpls.2023.1148658

Association mapping reveals novel genes and genomic regions controlling grain size architecture in mini core accessions of Indian National Genebank wheat germplasm collection

Jyoti Kumari[1†], Deepika Lakhwani[1†], Preeti Jakhar[1], Shivani Sharma[1], Shailesh Tiwari[1], Shikha Mittal[1,2], Himanshu Avashthi[1], Neelam Shekhawat[3], Kartar Singh[3], Kaushlesh Kumar Mishra[4], Rakesh Singh[1], Mahesh C. Yadav[1*], Gyanendra Pratap Singh[1] and Amit Kumar Singh[1*]

[1]ICAR-National Bureau of Plant Genetic Resources, New Delhi, India, [2]Jaypee University of Information Technology, Solan, India, [3]ICAR-National Bureau of Plant Genetic Resources, Regional Station Jodhpur, Jodhpur, India, [4]Zonal Agricultural Research Station, Powarkheda, India

Wheat (*Triticum aestivum L.*) is a staple food crop for the global human population, and thus wheat breeders are consistently working to enhance its yield worldwide. In this study, we utilized a sub-set of Indian wheat mini core germplasm to underpin the genetic architecture for seed shape-associated traits. The wheat mini core subset (125 accessions) was genotyped using 35K SNP array and evaluated for grain shape traits such as grain length (GL), grain width (GW), grain length, width ratio (GLWR), and thousand grain weight (TGW) across the seven different environments (E_1, E_2, E_3, E_4, E_5, E_5, E_6, and E_7). Marker-trait associations were determined using a multi-locus random-SNP-effect Mixed Linear Model (mrMLM) program. A total of 160 non-redundant quantitative trait nucleotides (QTNs) were identified for four grain shape traits using two or more GWAS models. Among these 160 QTNs, 27, 36, 38, and 35 QTNs were associated for GL, GW, GLWR, and TGW respectively while 24 QTNs were associated with more than one trait. Of these 160 QTNs, 73 were detected in two or more environments and were considered reliable QTLs for the respective traits. A total of 135 associated QTNs were annotated and located within the genes, including ABC transporter, Cytochrome450, Thioredoxin_M-type, and hypothetical proteins. Furthermore, the expression pattern of annotated QTNs demonstrated that only 122 were differentially expressed, suggesting these could potentially be related to seed development. The genomic regions/candidate genes for grain size traits identified in the present study represent valuable genomic resources that can potentially be utilized in the markers-assisted breeding programs to develop high-yielding varieties.

KEYWORDS

wheat, QTN, genome wide association studies, thousand grain weight, mrMLM

Introduction

Bread wheat (*Triticum aestivum* L.) is an important staple food crop, serving as the main source of energy, protein, and fiber for much of the world's human population (Ling et al., 2013; Ji et al., 2022). However, with the current rate of yearly increment in wheat yield, feeding the ever-increasing world population, which is expected to reach 9.3 billion by 2050, will be a daunting task. Further, depletion of natural resources such as land and water and a rise in the mean earth surface temperature exacerbates the problem and poses a challenge to producing sufficient wheat to feed the human population in the future (Nehe et al., 2019). To increase wheat yield, it is important to understand the genetic basis of traits that contribute to grain yield (GY). Grain yield and its contributing traits are complex in nature, highly influenced by environmental conditions, and regulated by multiple genes (Kato et al., 2000; Li et al., 2019; Ji et al., 2022). Quantitative trait loci (QTL) associated with GY has been extensively studied and reported on all 21 wheat chromosomes (Bennett et al., 2012; Sun et al., 2020). Several studies have identified numerous QTL for GY and productivity (Bennett et al., 2012; Zegeye et al., 2014; Malik et al., 2021; Ji et al., 2022). However, there is very limited information on the marker-assisted improvement of GY traits in wheat. This is primarily due to the non-availability of tightly linked robust markers with the GY-associated traits. The conventional QTL mapping approach has been extensively used for gene mapping and has enabled genetic dissection of seed traits in wheat (Breseghello and Sorrells, 2007; Duan et al., 2020). However, this approach does not allow detection of all the possible allelic variants of the target gene that might exist in the natural populations of wheat. Another downside of the QTL mapping approach is its poor resolution. Availability of gold standard wheat reference genome sequence and high-density SNP arrays is expected to accelerate high-resolution mapping of complex traits using both conventional and association mapping approaches (IWGSC et al., 2018; Chaurasia et al., 2021).

In the past few years, the genome-wide association study (GWAS) has become a popular approach to identify the QTL associated with complex traits in crops. In contrast to QTL mapping, this approach enables the exploration of a large number of alleles for any locus from a natural population of diverse individuals. This approach facilitates high-resolution mapping of traits because the individuals used for the association analysis might have undergone several rounds of historical recombination (Yu and Buckler, 2006; Li et al., 2019). Several GWAS studies have been performed on major crops such as *Oryza sativa* (Spindel et al., 2016), *Zea mays* (Xu et al., 2017; Xu et al., 2018), *Hordeum vulgare* (Visioni et al., 2013), *Avena sativa* (Newell et al., 2011), *Brassica napus* (Zhou et al., 2017), *Glycine max* (Zhang et al., 2015), *Sorghum bicolor* (Morris et al., 2013), and in wheat for the genetic dissection of various desirable traits (Peng et al., 2018; Chaurasia et al., 2021; Malik et al., 2021). Over the past few years, efforts have been made to develop GWAS models that are more suited to investigating genetics of simple as well as complex traits in plants. These GWAS models are broadly grouped into single-locus GWAS (SL-GWAS) and multi-locus GWAS (ML-GWAS) methods. SL-GWAS methods have been widely used to detect genetic variants for various traits, but one main limitation of this model is that the p-values of the markers identified to be associated with the target trait need to be subjected to multiple rounds of testing to avoid false positive associations. To overcome this limitation, Zhang et al. (2020) developed a mrMLM package that contains the following six ML-GWAS methodologies: mrMLM (multilocus random-SNP-effect MLM) (Wang et al., 2016), FASTmrMLM (fast mrMLM) (Tamba et al., 2017), ISIS EM-BLASSO (iterative modified-sure independence screening expectation-maximization Bayesian least absolute shrinkage and selection operator) (Wen et al., 2017), pKWmEB (integration of Kruskal-Wallis test with empirical Bayes) (Ren et al., 2018), FASTmrEMMA (fast multi-locus random-SNP-effect efficient mixed model analysis) (Wen et al., 2017), and pLARmEB (polygenic-background-control-based least angle regression plus empirical Bayes) (Zhang et al., 2017).

Among the GY-associated traits, grain size contributes the most, making it a key selection target in wheat breeding programs for developing high yielding varieties. Thousand grain weight (TGW) is the main component of GY and is determined by grain size traits such as grain length (GL), grain width (GW), and grain length width ratio (GLWR) (Sun et al., 2009; Li et al., 2022). Thus, it is important to understand the genetic and molecular basis of the mechanisms governing grain size in wheat genotypes and to identify the superior novel alleles governing this trait from germplasm collections for exploitation in the breeding program. Therefore, the main aim of this study was to dissect the genetic control of grain size traits such as GL, GW, GLWR, and TGW in wheat germplasm employing 35K SNP array using multi-locus GWAS approaches.

Materials and methods

Experimental materials and phenotyping

The experimental material for the GWAS study comprised of 125 diverse wheat accessions, a subset of a mini core developed from the composite wheat core set (Phogat et al., 2020) of the National Genebank of India. These accessions were comprised of 85 indigenous and 40 exotic collections, which included released varieties, landraces, genetic stocks, and elite genotypes (Supplementary Table 1). These accessions were evaluated following augmented block design in five blocks using four checks (HD2967 and C-306) randomized in each block over five years (*Rabi* 2015-16 to *Rabi* 2019-20). The GWAS panel was evaluated at the ICAR-National Bureau of Plant Genetic Resources (NBPGR), Issapur Farm, Delhi located 28.3748° N, 77.0902° E, 228.6 m AMSL, for five consecutive years; during the fifth year, the trial was also taken up at the ICAR-NBPGR, Regional Station, Jodhpur located at 26.2389° N, 73.0243° E, 263 m AMSL) and the Zonal Agricultural Research Station, Powarkheda, Madhya Pradesh located at 22.4154° N, 77.4442° E, 229 m AMSL. In total, these made up seven environments: Delhi (2015-16)-E_1, Delhi (2016-17)-E_2, Delhi (2017-18)-E_3, Delhi (2018-19)-E_4, Delhi (2019-20)-E_5, Powarkheda (2019-20)-E_6, and Jodhpur (2019–20) - E_7.

Each genotype was grown in a three-rows plot of 2 m length each, with a row-to-row distance of 0.25 m. Pests and diseases were controlled chemically, whereas weeds were controlled manually.

The wheat GWAS panel was evaluated for GL (mm), GW (mm), GLWR, and TGW (gm) from the harvested grain samples. The measurements of these grain parameters were performed by selecting the main spike of five random individual plants in the middle of the row for each accession. Grains per spike were estimated by hand-threshing the mature spike. TGW of each genotype was recorded by weighing all the seeds from a sample, dividing it by the total seed number measured, and multiplying the result by 1000. For GL and GW analysis, ten seeds from each five spikes were measured using digital vernier caliper and average value of the plot accessions was taken up for analysis. Grain length/width ratio (GLWR) was calculated by dividing the grain length mean by the grain width mean for each genotype.

Phenotypic data analyses

The phenotypic data was analyzed using ACBD-R (Augmented Complete Block Design with R) version 4.0 software (Rodriguez et al., 2018). The mean, coefficient of variation (CV), least significant difference (LSD), genotypic variance, and heritability were estimated. In ACBD-R v4.0, the best linear unbiased predictors (BLUPs) of each genotype were calculated for each environment and across environments along with four checks varieties. The calculated BLUPs were then used in the GWAS analysis. The frequency distribution graphs, correlation coefficients of the recorded traits, and principal component analysis were obtained through SAS JMP Version 14 software (https://www.jmp.com/en_in/software/data-analysis-software.html).

Genomic DNA isolation and SNP genotyping

Genomic DNA was isolated from one-week-old wheat seedlings using the CTAB method (Murray and Thompson, 1980) with a few modifications and then treated with RNase to remove any RNA contamination. The integrity of DNA samples was checked on 0.8% agarose gel and concentration was determined by using a NanoDrop1000 (Thermo Scientific). Genotyping of isolated DNA samples was done using Breeder's 35K Axiom® array (Allen et al., 2017). The SNPs with a genotyping call rate < 97% and minor allele frequency (MAF) <5% were removed while performing genomic data analysis.

Clustering, population structure, and linkage disequilibrium analysis

A total of 23,874 SNPs were used to perform principal component analysis (PCA) and generate kinship matrix using TASSEL 5.2 program (https://www.maizegenetics.net/tassel). STRUCTURE software was used to estimate the level of genetic differentiation in the population using the Bayesian model-based approach; the parameter burn-in period and Monte Carlo Markov Chain (MCMC) replication number were set to 10,000 and 20,000 respectively for ten independent runs to estimate the number of subpopulations (k) in a putative range of k = 1 to 5. The optimal subpopulation number was estimated using an *ad hoc* statistic delta k (Evanno et al., 2005). The squared allele frequency correlation (r^2) between SNP markers was used to estimate linkage disequilibrium (LD) using TASSEL v5.2 (https://www.maizegenetics.net/tassel).

Genome-wide association analysis

All 125 accessions were genotyped using 35K SNPs array. We used five ML-GWAS methods which were included in the R package mrMLM v4.0.2 (https://cran.r-project.org/web/packages/mrMLM/index.html). These five models are mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, and ISISEM-BLASSO. All the parameters were set at default values. The critical thresholds of significant association for all the five methods were set as logarithm of the odds (LOD) score ≥3.00. The most significant SNPs, detected in at least two methods, were considered as reliable SNPs.

Differential gene expression analysis

We utilized RNA-seq data of two wheat genotypes with contrasting seed size i.e., IC111905 (large-seeded)and EC575981 (small-seeded) at 15 days and 30 days post anthesis (DPA) with three biological replicates to check the expression profile of putative candidate genes located in the identified genomic regions. Illumina sequencing was performed, which generated approximately 177 Gb raw data. Approximately 98.5% of reads passed the quality control and clean reads were mapped back on to the reference genome (IWGSC v2.0) (https://plants.ensembl.org/Triticum_aestivum/Info/Index) using bwa-mem software (https://sourceforge.net/projects/bio-bwa/files/). The differential gene expression analysis was performed using edge R package and genes with p-value <0.05 were considered as significantly differentially expressed genes. Heat maps of differentially expressed genes were generated by MeV software (https://sourceforge.net/projects/mev-tm4/).

Result

Phenotypic evaluation and variability

All the genotypes of the wheat association panel were phenotyped for grain size parameters (GL, GW, GLWR, and TGW) in seven different environments (E_1-E_7). The descriptive statistics of the investigated traits in seven environments are presented in Supplementary Table 2, and revealed wide variability for all the traits. GL ranged from 5.24 to 8.20 mm in E_1, 5.09 to 8.17 mm in E_2, 4.83 to 8.15 mm in E_3, 5.04 to 8.09 mm in E_4, 4.43 to 7.79 mm in E_5, 4.43 to 8.41 mm in E_6, and 4.47 to 7.87 mm in E_7. GW ranged from 2.29 to 3.95 mm in E_1, 2.26 to 3.71 mm in E_2, 1.89 to 3.73 mm in E_3, 1.87 to 3.69 mm in E_4, 1.41 to 3.67 mm in E_5, 1.41 to 3.58 mm in E_6, and 1.48 to 3.81 mm in E_7. GLWR ranged from

1.49 to 2.65 in E_1, 1.29 to 2.70 in E_2, 1.64 to 3.12 in E_3, 1.64 to 3.52 in E_4, 1.62 to 3.12 in E_5, 1.69 to 3.12 in E_6, and 1.61 to 3.02 in E_7. TGW ranged from 18.00 to 64.54 in E_1, 13.51 to 67.56 in E_2, 7.91 to 51.93 in E_3, 9.87 to 61.39 in E_4, 12.17 to 56.66 in E_5, 13.75 to 67.21 in E_6, and 14.26 to 55.51 in E_7. The coefficients of variation for GL, GW, GLWR, and TGW ranged from 8.32% to 24.00%, indicating considerable variability for these traits. The CV percent was highest for TGW in E_6 (24.00%) followed by E_4 (21.62%) and E_2 (20.39%). The frequency distribution of four traits (GL, GW, GLWR, and TGW) (Supplementary Figure 1) showed near normal distribution in all environments, indicating the quantitative nature of these traits except for GW under environments E_3 and E_4.

Based on BLUP analysis (Supplementary Table 3), genotypes recorded an overall grand mean of 6.33 ± 0.40 mm, 6.34 ± 0.39 mm, 6.28 ± 0.36 mm, 6.23 ± 0.44 mm, 6.27 ± 0.34mm, 6.52 ± 0.28 mm, and 6.64 ± 0.63 mm respectively under different environments for grain length, whereas for grain width, overall means of 3.19 mm ± 0.06, 3.19 ± 0.07 mm, 3.10 ± 0.18 mm, 3.02 ± 0.39 mm, 3.22 ± 0.22 mm, 3.08 ± 0.41 mm, and 3.20 ± 0.41 mm respectively were recorded. GLWR recorded overall mean of 2.02 ± 0.40, 2.02 ± 0.40, 2.06 ± 0.15, 2.08 ± 0.16, 1.95 ± 0.15, 2.12 ± 0.14, and 2.13 ± 0.18 in different environments, whereas TGW recorded grand means of 42.60 ± 5.24 g, 44.30 ± 6.73 g, 37.66 ± 3.00 g, 39.85 ± 6.84 g, 43.07 ± 5.08 g, 38.35 ± 5.30 g, and 42.00 ± 3.12g respectively. Promising accessions were identified based on adjusted BLUP mean. The top ten accessions for GL were EC578134 (7.457mm), IC539313 (7.148 mm), EC339611 (7.106 mm), EC464070 (7.085 mm), C697725 (7.036 mm), IC252928 (7.033 mm), IC535217 (7.016 mm), EC542279 (7.014 mm), EC313710(7.005 mm), and EC578152(6.969 mm). For GW, the top ten accessions were IC252429 (3.753 mm), IC335683 (3.441 mm), IC252954 (3.432 mm), IC335715 (3.418 mm), IC252772 (3.38 mm), IC574476 (3.378 mm), IC252422 (3.369 mm), IC75240 (3.355 mm), IC122726(3.349 mm), IC116274(3.34 mm), and IC539314(3.328 mm). Similarly, promising accessions with more than 50 g thousand grain weight were identified as IC539313 (55.03 g), EC578134 (53.56 g), IC542076 (50.95 g), IC335715 (50.36 g), and EC578152 (50.0g).

The environment-wise heritability and variance components based on BLUP value are presented in Supplementary Table 3. Heritability for GL ranged from 22.5% (E_7) to 82.4% (E_4), whereas heritability for GW ranged from 22.3% (E_7) to 60.2% (E_4). Similarly, heritability for GLWR ranged from 22.0% (E_1) to 62.3% (E_5), whereas TGW was found to be heritable in the range of 21.8% (E_6) to 74.1% (E_3).

Multivariate analysis

Correlation between traits in different environments

Pearson's correlation coefficients were estimated among grain traits for diverse wheat accessions under each environment separately (Supplementary Table 4). GL was found to have a consistently significant positive correlation with GW (0.368, 0.406, 0.444), GLWR (0.585, 0.562, 0.335), and TGW (0.471, 0.394, 0.538) under E_1, E_2, and E_3 respectively. Contrarily, GW showed a negative correlation with GLWR (-0.363, -0.438, and -0.684) and significant positive correlation with TGW (0.371, 0.380, and 0.604) under E_1, E_2, and E_3 respectively. Under the environment E_4, GL showed significant positive correlations with GW, GLWR, and TGW that ranged between 0.225 (GW) to 0.457 (GLWR), while GW showed highly a significant negative correlation with GLWR (-0.748) and significant positive correlation with TGW (0.488). Similarly, under E_5, a significant positive correlation was observed with GW (0.406), GLWR (0.361), and TGW (0.643), while GW showed a significant negative correlation with GLWR (-0.686) and significant positive correlation with TGW (0.723). A similar correlation pattern among seed traits was also observed under E_6 and E_7 (Supplementary Table 4).

Phenotypic correlation between different environments for traits

The magnitudes of correlation between environments were assessed for knowing behavior of genotypes for trait expression. A total of twenty-one combinations of correlations were observed between different pairs of environments for all the traits (Supplementary Table 5). Here, significant positive associations were revealed for response of traits by genotypes for all environment pairs except five for grain length, and TGW and three for grain width and GLWR. For GL, these correlation values ranged from 0.084 between E_3 to E_7 to 0.985 between E_1 to both E_2 and E_4. For GW, these correlation values ranged from 0.061 between E_4 to E_7 to 0.953 between E_1 and E_2. Similarly, for GLWR, the lowest association was observed as 0.089 between E_4 to E_6 and highest as 0.961 between E_1 and E_2. For TGW, lowest association was observed as 0.021 between E_4 to E_6 and highest as 0.954 between E_2 and E_4.

Principal component analysis and correlation study based on pooled analysis

PCA was performed on the basis of pooled data for seven environments for grain parameters. Genotype by trait biplot depicted two-dimensional spatial diversity of accessions as well as trait variability (Figure 1A). High trait variability was observed for traits like GLWR, GL, GW, and TGW, which was evidenced by the larger length of the characters and high positive correlations between traits such as GL and TGW and GLWR and GL, and a negative correlation between TGW and GW, which is evident by the narrow angle between them. Here, the first principal component, PC1, explained 57.4% of the cumulative variation. The major contributing traits for PC1 were GW (0.90), TGW (0.90), and GL (0.79) in positive direction. Second principal component, PC2, explained 34.5% of the cumulative variation. The major contributing traits for PC2 were GLWR (0.97) and GL (0.56). Further, we analyzed correlations between traits using the BLUP values where GL showed a highly significant positive correlation with GW (0.562), GLWR (0.368), and TGW (0.667), whereas GW showed a highly significant negative correlation with GLWR (-0.508) and significant positive correlation with TGW (0.680). GLWR was not significantly related to TGW (Figure 1B).

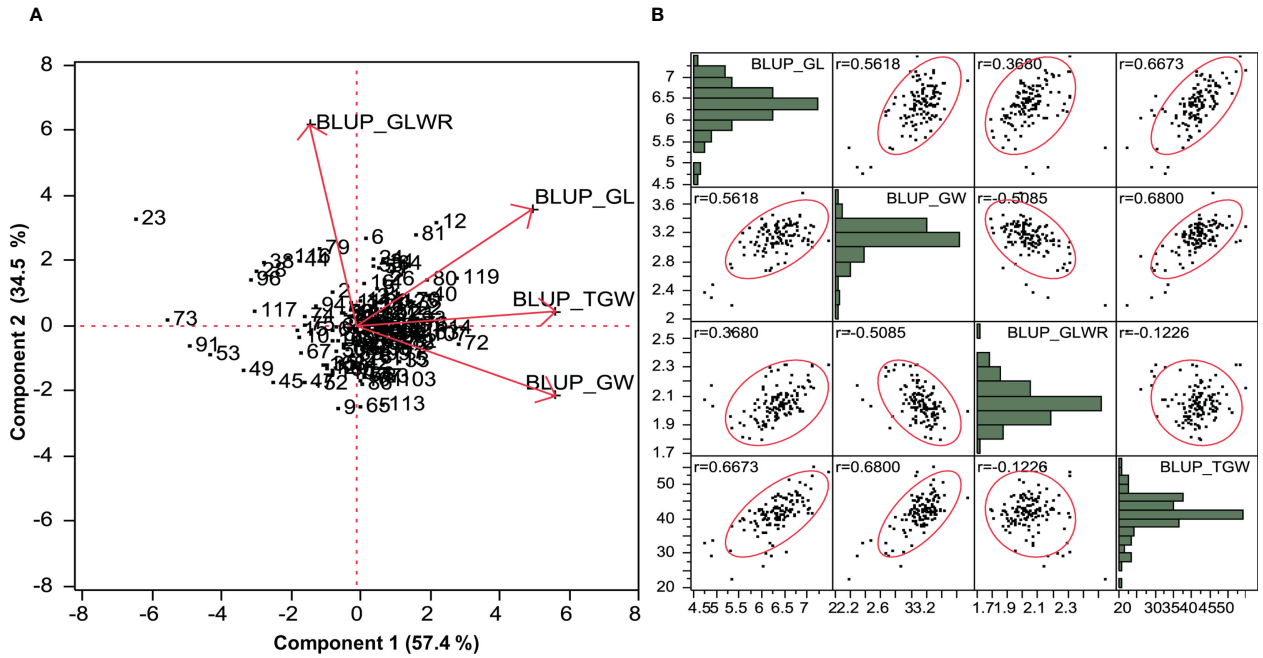


FIGURE 1
(A) Principal component biplot based on BLUP value of grain parameters over the environments. **(B)** Scatter plot showing correlation matrix between grain parameters based on BLUP value.

Genotyping

A total of 125 wheat accessions representing a subset of the Indian National Genbank mini core germplasm were genotyped using 35K wheat SNP array that contains 35,143 genome-wide single nucleotide polymorphism (SNP) markers. The SNP probe sequences of wheat array were BLASTn (https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html) against wheat genome to find out their physical location, which revealed only 31,926 SNPs with known positions. Furthermore, SNPs were also filtered on the basis of minor allele frequency (≥0.05), missing threshold of < 10%, and call rate ≥97%. Finally, a total of 23,874 SNPs were retained for genetic diversity, population structure, and GWAS analysis. Mapping of 23,874 filtered SNP markers provided a whole genome-wide coverage along the 21 chromosomes of wheat (Figure 2). Further, distribution analysis of SNPs on wheat chromosomes revealed that the maximum number of SNPs was positioned on 1B (1594), followed by 2D (1575) and 1D (1516), while the lowest number was positioned on 4D (508), followed by 4B (800) and 4A (808). We also compared the distribution of SNPs on the three wheat sub genomes; it was found that 7,291 SNPs belonged to A sub-genome, 8,784 SNPs were found on B sub-genome, and 7,799 SNPs belonged to D sub-genome (Table 1).

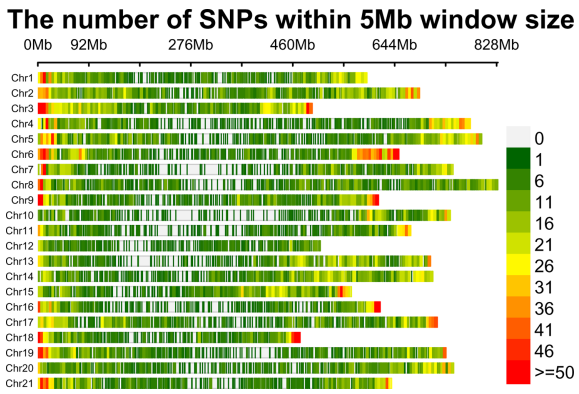


FIGURE 2
SNP density plot of 21 wheat chromosomes displaying distribution of SNPs within 5 Mb window size. The horizontal axis shows chromosome length (Mb); Different colors depict SNP density.

Population structure, kinship, and linkage disequilibrium decay analyses

We used 23, 874 SNP markers to ascertain the population structure in the wheat mini core set using STRUCTURE and PCA analysis. The most probable number of populations were estimated using delta K method implemented in the STRUCTURE HARVESTOR program. The value of ΔK peaked at K=2 and revealed two sub-populations in the wheat mini core germplasm. Sub-population 1 represented 82% of the individuals; out of that, 62% were pure and 38% admixtures. Whereas sub-population 2 had 18% of the individuals of the AM panel, and contained 75% pure and 25% admixtures. PCA also detected the two sub-populations indicated by two significant components, explaining the maximum variation of the population. Further, kinship matrix was also created to explore the relationship among the individuals using the genome association and prediction integrated tool (GAPIT) which demonstrated the presence of two sub-groups within the association panel (Figure 3).

The LD decay in the wheat mini core set was estimated by calculating the squared correlation coefficient (r^2) for all the SNPs. The LD decay for the whole genome was 1.9 Mb. Further, it was found that the decay was most rapid in the A sub-genome (1.63 Mb), followed by the D sub-genome (1.93 Mb) and B sub-genome (2.28 Mb) (Figure 4).

GWAS for grain size traits

GWAS was performed using 23,874 SNPs filtered on various parameters to identify genomic loci associated to four different grain yield traits (GL, GW, GLWR, and TGW) independently for the seven environments and also based on the BLUP values derived from data of grain size traits of all the seven environments. Here, we have used five multi-locus models (mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, and ISIS EM-BLASSO) to conduct GWAS. A total of 752 significant SNPs were predicted for four grain size traits using ML-GWAS models with LOD score ≥ 3. Manhattan plots for GL drawn using various ML-GWAS models that depict marker trait associations are presented in Figure 5. Of these 752 SNPs, 72 were identified using BLUP values derived from the data of all the environments and other SNPs were identified by analyzing data of each location separately. We classified 752 SNPs according to trait, which demonstrated that 156, 179, 250, and 167 SNPs were associated with GL, GW, GLWR and TGW traits, respectively (Figure 6). In addition, comparative study on the basis of the

TABLE 1 Chromosome-wise distribution of 23,874 SNPs and the intra-chromosomal estimated LD among 125 wheat genotypes.

Chrom	Size (Mb)	No. of SNP	SNP/Mb Density	SNP pair in LD (p <0.05)	SNP pair In LD (r2 = 1)	Average LD (r2)	D prime
1A	594.1	1138	1.92	102600	5766	0.220567146	0.71973128
1B	689.85	1594	2.31	236002	22116	0.271350626	0.77589076
1D	495.45	1516	3.06	163630	19467	0.314246556	0.77693029
2A	780.8	1187	1.52	101708	1958	0.189312876	0.66675889
2B	801.26	1513	1.89	155261	2648	0.169167472	0.64845702
2D	651.85	1575	2.42	146171	6944	0.249166202	0.73375966
3A	750.84	1024	1.36	62572	1342	0.17548539	0.66747653
3B	830.83	1195	1.44	101455	1207	0.163443076	0.64468071
3D	615.55	1158	1.88	62878	3762	0.239489328	0.71345541
4A	744.59	808	1.09	39406	685	0.177018377	0.65255529
4B	673.62	800	1.19	40467	1107	0.218309804	0.68785027
4D	509.86	508	1	12770	1324	0.308622143	0.78044746
5A	709.77	1132	1.59	76483	1872	0.171701946	0.64123409
5B	713.15	1278	1.79	130839	2647	0.163887549	0.63806885
5D	566.08	1113	1.97	72678	4467	0.232972965	0.7229823
6A	618.08	845	1.37	43714	1187	0.191258087	0.66474072
6B	720.99	1259	1.75	126269	1264	0.171017301	0.65574685
6D	473.59	851	1.8	42753	2005	0.217027812	0.70667145
7A	736.71	1157	1.57	94614	1230	0.165787476	0.65314744
7B	750.62	1145	1.53	78061	1132	0.179186058	0.66074305
7D	638.69	1078	1.69	61139	2465	0.237260494	0.72331864

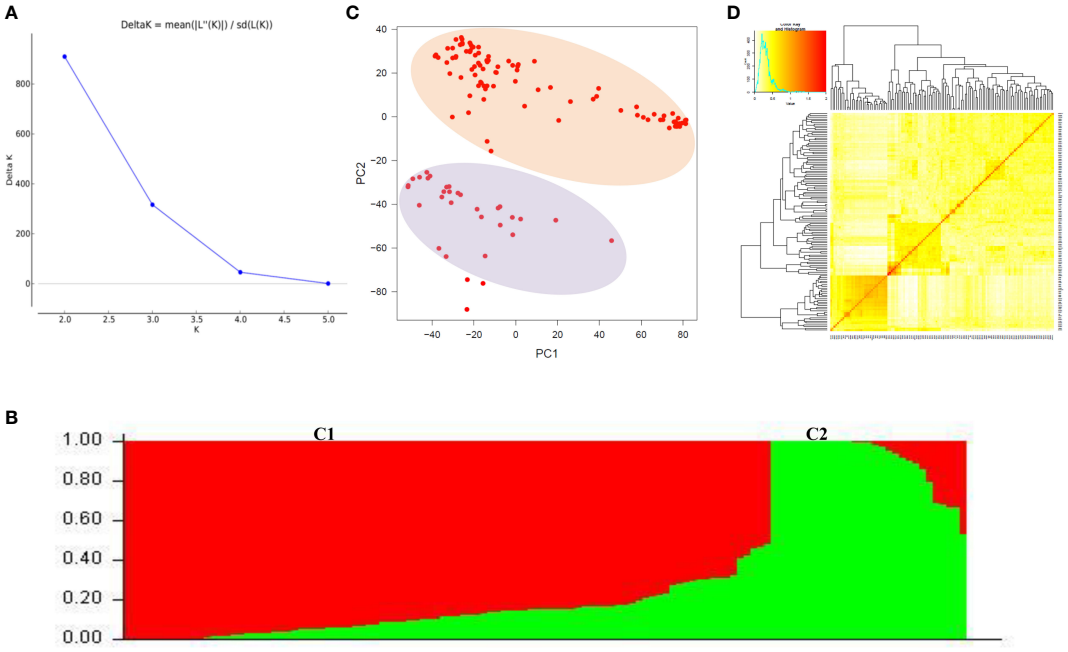


FIGURE 3
Population structure analysis of wheat association mapping panel. **(A)** Magnitude of ΔK values, rate of change from 2 to 5 in association mapping panel. **(B)** Population structure of association panel based on 125 germplasm-based SNP markers at K = 2. Different color columns represent different sub sub-populations. **(C)** Principal component analysis showing two sub sub-populations. **(D)** Heat map of kinship matrix. The heat map shows the level of relatedness among the population. The darker areas show the level of relatedness between genotypes and the dendrogram depicts clustering of sub sub-population.

GWAS model demonstrated that 28, 32, 36, and 31 SNPs were predicted for four traits (GL, GW, GLWR, and TGW respectively) by the mrMLM model while the FASTmrMLM model could detect 34, 42, 79, and 39 SNPs for GL, GW, GLWR, and TGW. FASTmrEMMA and pLARmEB model identified 13, 20, 17, and 12 and 36, 50, 69, and 45 SNPs for the four traits (GL, GW, GLWR, and TGW respectively). The ISIS EM-BASSO model detected 45, 38, 49, and 40 SNPs for GL, GW, GLWR, and TGW respectively (Figure 6).

On the basis of redundancy of SNPs in the models and locations, we combined the identified SNPs and found a total of 160 SNPs were simultaneously detected in two or more multi-locus models. These SNPs were designated as reliable *QTNs* for the respective traits. Furthermore, distribution of these 160 significant SNPs was also analyzed across the environments. Out of these 160 *QTNs*, 87 were confined to only one environment that included 13, 19, 27, and 25 QTNs for GL, GW, GLWR, and TGW respectively, and 3 *QTNs* that were associated with more than one trait (Table S6). On the other hand, a total of 73 *QTNs* were simultaneously identified in two or more environments as well as two or more models (Table 2). Among these 14, 17, 11, and 10 *QTNs* were identified for GL, GW, GLWR, and TGW traits respectively while 21 SNPs were associated with more than one trait. The physical distribution of all the 160 SNPs on chromosomes demonstrated that SNPs were present on all the chromosomes. Moreover, the highest number of SNPs were found on chr3D (7 SNPs), followed by chr2D (6 SNPs), and chr7D (6 SNPs) while chr1D, chr6B, and chr6D had only one SNPs.

Allelic effects of identified genomic regions on grain shape

To evaluate the allelic effects of *QTNs* on respective phenotype, we only analyzed those *QTNs* that were detected in more than two environments and revealed R^2 value ≥10% with at least one GWAS model. Association panel genotypes were divided into two groups according to allele types in order to test whether the mean phenotypes of the two groups were significantly different (Figure 7). Results showed that six *QTNs* had significant effect ($P \leq 0.01$) on their respective traits. Among these six *QTNs*, four QTNs (*Q.GL-GW-3D* (AX-95008504), *Q.GL-5D* (AX-94482861), *Q.GL-5D* (AX-95020206), and *Q.GL-TGW-6A* (AX-95238912)) demonstrated significant effects on GL(mm) whereas one, *QTN Q.GW-4B* (AX-94878781), had significant effect on GW and another, *Q.GLWR-2A* (AX-94736090), showed significant effect on GLWR. The QTNs with significant phenotypic effects on seed traits might contribute to their genetic variations.

Annotation of identified QTNs

All the 160 significantly associated QTNs with grain size traits that were detected in two or more models were searched for their annotation in the wheat reference genome assembly cv. Chinese Spring (IWGSCrefseq version 2.0, https://wheat.pw.usda.gov/GG3/iwgsc-2.0), available at Plant Ensemble. Of these 160 QTNs, annotation was only detected for 136 QTNs. The detailed analysis

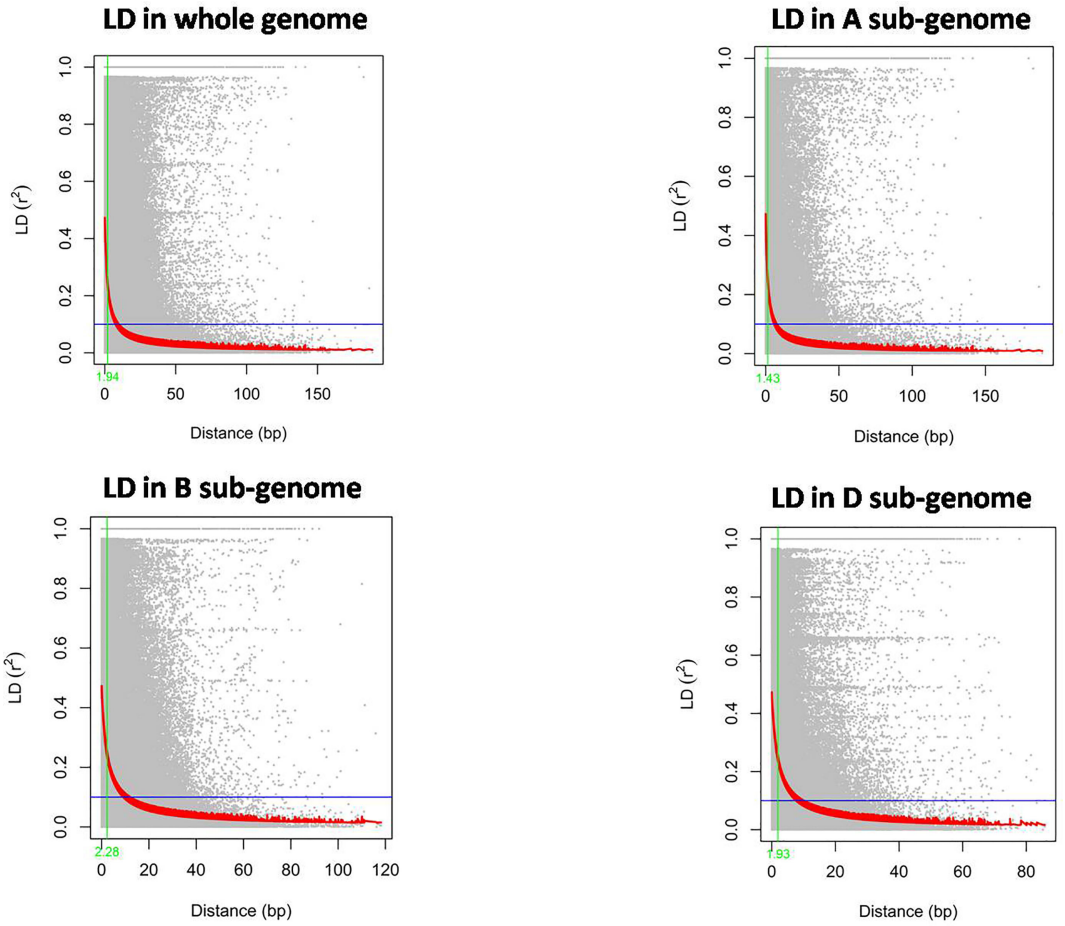


FIGURE 4
The rate of Linkage disequilibrium decay (R^2) between pairs of polymorphic markers of the whole wheat genome and its sub-genomes A, B, and D are plotted against the genetic distance (Mb).

of annotated SNPs showed that *Q.TGW-5D* (SNP-AX-95234313) was located within a gene encoding cytochrome 450 and was identified in E4 environment using mrMLM and FASTmrMLM models with LOD score ranging between 3.8 to 3.92. Another, *Q.GW-3D* (SNP-AX-94540502), for grain weight identified at the E6 environment was annotated as ABC transporter and detected by mrMLM and ISIS EM-BLASSO models. We also checked the annotation of QTNs that were identified at multiple environments and found that *Q.GL-1B* (SNP-AX-94699549), *Q.GL-GLWR-4B* (SNP-AX-94879134), *Q.GL-3D* (SNP-AX-95074739), and *Q.GW-5A* (SNP-AX-94657794) were located within genes encoding ABC transporter, WRKY transcription_factor, Glucan endo-1,3-beta-glucosidase, and Zinc_finger_protein respectively. Among these four QTLs, *Q.GL-1B* (encoding for an ABC transporter) was located at 585,331,222bp on chromosome 1B. It was identified in two different environments, E_1 and E_4, using ISIS EM-BLASSO model with LOD scores 4.17 and 5.65 and R^2 3.05% and 6.42%, respectively. *Q.GL-GLWR-4B* (encoding for a WRKY transcription factor) was located on chr4B at 63,231,309bp position and identified in four different environments: E_1 with ISIS EM-BLASSO model, E_2 with ISIS EM-BLASSO and FASTmrEMMA models, E_3 with pLARmEB model, and E_4 with ISIS EM-BLASSO model. It had an LOD score ranging between 3.52 to 8.12. This QTN was associated with GL in all locations except E_2. Further, in E_2, this *Q.GL-GLWR-4B* was also determined by three different models, namely FASTmrMLM, pLARmEB, and ISIS EM-BLASSO, but associated with GLWR trait.

Expression analysis

The transcriptome sequencing of contrasting seed size wheat genotypes, i.e. IC111905 (large-seeded) and EC575981(small-seeded), was performed at two time intervals during the seed development (i.e., 15 and 30 DPA) to quantify expression of all annotated genes within associated genomic regions.

Expression analysis demonstrated that only 123 genes were expressed in both stages (15 and 30 DPA) of small and large seeded genotypes, of which 23, 33, 27, and 28 were uniquely associated with GL, GW, GLWR, and TGW respectively. Among the identified genes, those with foldchange ≥1 and p-value<0.05 were considered as significantly differentially expressed genes. A total of 18 and 12 genes were significantly differentially regulated in large seeded

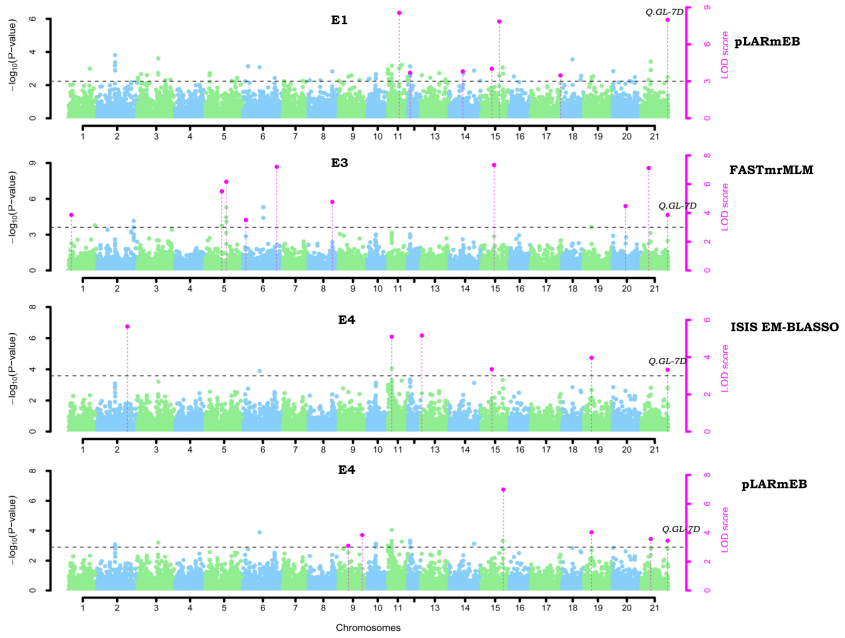


FIGURE 5
Manhattan plots of associated QTNs for grain length (GL) in wheat using multi-locus GWAS model. The x-axis shows the chromosome label and the y-axis displays - thresholds for significance (LOD score = 3) and log10 (p-value). The significant QTNs with LOD score >=3 is represented with purple dots.

cultivars at 15 and 30 DPA respectively. At 15 DPA, 11 genes were upregulated and 7 genes were downregulated, while 6 genes were significantly upregulated and downregulated in large seed cultivars at 30 DPA (Figure 8). Many genes, including *TraesCS7B02G462900* (SNP-AX-94472687; *Q.GW-7B*), *TraesCS2D02G132600* (SNP-AX-94499721; *Q.GLWR-2D*), *TraesCS1A02G187000* (SNP-AX-95120969; *Q.TGW-GW-1A*), *TraesCS2B02G260200* (SNP-AX-95129853; *Q.GLWR-GL-2B*), and *TraesCS6A02G379200* (SNP-AX-95151036; *Q.GW-6A*), were downregulated at both the time points while *TraesCS1A02G427400* (SNP-AX-94605845; *Q.TGW-1A*), *TraesCS3D02G002700* (SNP-AX-94642652; *Q.GW-3D*), *TraesCS7A02G111200* (SNP-AX-94820170; *Q.GLWR-GL-7A*), and *TraesCS3A02G180200* (SNP-AX-94960788; *Q.GL-3A*) genes were upregulated at 15 as well as 30 DPA. Result also showed that *Q.GLWR-5B* (SNP-AX-94915493) associated with GLWR was located within gene *TraesCS5B02G552400* (hypothetical protein), which was only expressed in small grain cultivars with 10 fold upregulation, which showed it has some specific role in small seed cultivars. Another gene, namely *TraesCS7D02G463100*, located within the QTN*Q.GW-7D* and identified in E_1 and E_5 location was upregulated in large grain cultivars. *TraesCS7D02G463100* is nuclear transcription factor associated with GLWR.

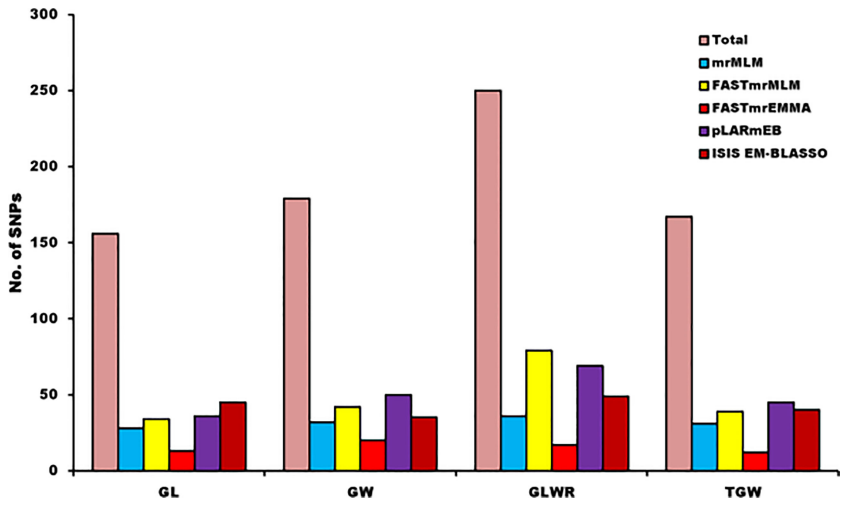


FIGURE 6
Distribution of identified and significant SNPs for each trait on the basis of detection models of multi-GWAS.

TABLE 2 List of significant QTNs detected simultaneously using two or more multi-locus GWAS methods for four wheat yield-related grain shapes across the environments.

QTN	SNP	Traits	Location	Chr	Position	LOD	R^2	Models
Q.GW-1A	AX-94592848	GW	E_1,E_2	chr1A	3.34E+08	5.69~5.88	0.92~2.1	4,2
Q.TGW-1A	AX-94932678	TGW	E_2,E_4	chr1A	5.58E+08	3.18~8.8	11.48~23.69	1,2,4,5,5
Q.GW-1A	AX-94997333	GW	E_1,E_2	chr1A	20094503	4~18.51	6.49~23.71	1,2,4,5,4
Q.TGW-GW-1A	AX-95120969	TGW,GW	E_1,E_6	chr1A	3.38E+08	3.04~5.3	5.93~6.16	5,3
Q.GLWR-1A	AX-95213485	GLWR	E_1,E_3	chr1A	25708710	3.12~26.68	3.32~33.74	1,2,4,5,4,5
Q.TGW-1B	AX-94517103	TGW	E_2,E_4	chr1B	6.41E+08	3.87~9.84	7.95~8.58	3,4
Q.GL-1B	AX-94699549	GL	E_1,E_4	chr1B	5.85E+08	4.17~5.65	3.05~6.42	5,5
Q.TGW-1B	AX-94830564	TGW	E_2,E_4	chr1B	74228850	4.11~10.66	7.94~14.14	4,1,2,5
Q.GW-GLWR-1D	AX-95098685	GW,GLWR	E_3,E_4	chr1D	4.17E+08	4.28~5.67	6.66~10.8	1,2,1,4,5
Q.GW-2A	AX-94402160	GW	E_1,E_2	chr2A	1.92E+08	3.71~10.65	0~9.68	3,4,2,3,4
Q.GLWR-2A	AX-94736090	GLWR	E_1,E_2, E_3,E_4,E_5	chr2A	7.69E+08	3.26~5.91	4~11.47	1,4,5,5,1,2,4,5,3
Q.TGW-2A	AX-94780053	TGW	E_2,E_4	chr2A	7.79E+08	3.19~7.47	3.12~14.19	1,2,4,5,5
Q.TGW-2A	AX-95012027	TGW	E_2,E_4	chr2A	71666146	6.22~6.6	4.77~7.37	4,5
Q.TGW-GW-2B	AX-94470912	TGW,GW	E_3,E_5	chr2B	7.73E+08	3.85~5.26	9.44~14.2	2,3
Q.GW-2B	AX-94519462	GW	E_3,E_4	chr2B	5.46E+08	3.87~6.47	7.7~12.23	1,2,3,1,2,3
Q.GL-GW-2B	AX-94878848	GL,GW	E_3,E_4	chr2B	5.46E+08	5.66~6.67	7.97~17.37	1,2,4,5,4,5
Q.GLWR-GL-2B	AX-95129853	GLWR,GL	E_1,E_3	chr2B	3.26E+08	4.54~8.2	3.71~28.72	2,4,1,2,4,5
Q.GL-TGW-2D	AX-94618441	GL,TGW	E_3,E_4	chr2D	6.23E+08	3.32~7.21	7.41~12.87	2,1,5
Q.GW-2D	AX-94655905	GW	E_1,E_2	chr2D	96623607	4.23~7.86	0~2.05	4,2,4
Q.TGW-2D	AX-94745278	TGW	E_2,E_4	chr2D	22451354	3.32~10.06	3.8~11.63	1,2,4,5,1,2,5
Q.GLWR-2D	AX-94774424	GLWR	E_3,E_4	chr2D	4.85E+08	4.47~11.69	3.06~6.63	2,5,4,1
Q.GLWR-2D	AX-94922377	GLWR	E_1,E_2	chr2D	2693668	3.61~4.63	0~23.04	4,1,2
Q.GL-2D	AX-95128254	GL	E_1,E_2	chr2D		3.52~5.94	2.52~9.48	1,2,5,5
Q.GW-3A	AX-95109402	GW	E_1,E_2	chr3A	25939226	4.3~5.83	0.95~3.02	4,5,4
Q.GL-GL-GL-3A	AX-95187884	GL	E_1,E_4	chr3A	10884903	3.3~4.97	0~12.07	2,1,2
Q.GLWR–3B	AX-94671460	GLWR,	E_3,E_4	chr3B	7.46E+08	3.47~5.4	0~2.63	4,5,5
Q.GLWR-3B	AX-94799334	GLWR	E_1,E_2	chr3B	8.18E+08	4.09~4.32	0~6.96	4,2
Q.GLWR-3B	AX-95150002	GLWR	E_1,E_2	chr3B		5.52~6.85	0.54~7.19	4,4,5
Q.GW-TGW-3D	AX-94401378	GW,TGW	E_2,E_4	chr3D	6.08E+08	4.24~5.96	4.76~7.65	5,2,4,1
Q.TGW-3D	AX-94406908	TGW	E_2,E_3,E_4	chr3D	2.39E+08	3.16~8.2	7.21~11.48	1,2,2,4
Q.GLWR-GW-3D	AX-94535556	GLWR,GW	E_1,E_2,E_5	chr3D		3.16~10.15	3.21~7.93	3,4,4,5,1,2,3
Q.GW-3D	AX-94642652	GW	E_1,E_2	chr3D	1253893	3.33~10.06	10.18~29.66	5,1,4,5
Q.GW-GLWR-3D	AX-94749865	GW,GLWR	E_3,E_4	chr3D	5.28E+08	4.1~6.7	4.64~21.47	1,4,5
Q.GL-GW-3D	AX-95008504	GL,GW	E_1,E_2,E_4,E_5	chr3D	1.51E+08	3.05~6.33	4.09~14.97	1,2,1,2,4,5,4,5
Q.GL-GL-3D	AX-95074739	GL	E_2,E_4	chr3D	5.77E+08	3.84~4.85	2.14~2.66	4,4
Q.GL-4A	AX-94839917	GL	E_1,E_2	chr4A	45529997	4.92~6.13	15.33~31.94	1,2,5,4,5
Q.GLWR-4A	AX-95019395	GLWR	E_1,E_2	chr4A	29453952	3.6~8.31	2.61~8.88	2,3,5,3
Q.TGW-4B	AX-94425015	TGW	E_1,E_4	chr4B	2036666	3.2~3.75	1.83~2.94	5,5

(Continued)

TABLE 2 Continued

QTN	SNP	Traits	Location	Chr	Position	LOD	R^2	Models
Q.GW-4B	AX-94433424	GW	E_1,E_2	chr4B	6.72E+08	3.78~8.49	0.62~11.76	1,2,4,4
Q.TGW-GW-GL-4B	AX-94878781	TGW,GW,GL	E_2,E_3,E_5	chr4B	6.45E+08	3.6~4.86	8.89~16.14	5,1,2,4
Q.GL -GLWR-4B	AX-94879134	GL,GLWR	E_1,E_2,E_3,E_4	chr4B	63231309	3.52~8.12	3.44~7.27	5,3,24,
Q.TGW-4B	AX-95232992	TGW	E_1,E_2	chr4B	3.89E+08	4.39~5.06	7.48~9.31	3,3
Q.GW-5A	AX-94657794	GW	E_3,E_4	chr5A	6.74E+08	3.48~4.01	3.53~3.62	4,4
Q.GL-GW-5A	AX-94909932	GL,GW	E_1,E_4	chr5A	11068517	3.17~9.9	1.54~5.26	5,1,2,4,5,5
Q.GW-5B	AX-94424550	GW	E_1,E_2	chr5B	4.76E+08	3.42~11.54	1.49~4.19	2,3,4
Q.GW-5B	AX-94531833	GW	E_1,E_2	chr5B	7.13E+08	3.5~9.78	0~6.7	2,4,5,1,2
Q.GW-GL-5B	AX-94711368	GW,GL	E_5,E_7	chr5B	7034159	3.71~3.72	2.93~4.95	4,5
Q.GLWR-5B	AX-94915493	GLWR	E_1,E_2	chr5B	7.03E+08	4.19~18.5	2.94~5.97	2,4,5,1,2,4
Q.GL-TGW-5B	AX-95189661	GL,TGW	E_1,E_3	chr5B	4.64E+08	3.61~5.05	4.61~12.21	4,2,4
Q.GL-5D	AX-94424746	GL	E_1,E_4	chr5D	3.54E+08	3.35~4.94	7.6~13.92	4,1,2,5
Q.GL-5D	AX-94482861	GL	E_1,E_2,E_4	chr5D	5.15E+08	4.91~10.32	7.8~12.41	1,2,4,4
Q.GL-5D	AX-95020206	GL	E_1,E_2,E_4	chr5D	5.02E+08	3.57~4.95	6.38~11.63	3,2,3,2,3
Q.GL-5D	AX-95078562	GL	E_2,E_3	chr5D	3.85E+08	3.92~7.33	2.37~9.56	5,1,2,3,4,5
Q.GW-5D	AX-95192563	GW	E_3,E_4	chr5D	3.82E+08	3.29~3.46	0~11.47	2,2
Q.GLWR-6A	AX-94615640	GLWR	E_3,E_4	chr6A	9253732	5.03~6.88	6.04~6.41	3,2
Q.GLWR-6A	AX-94722285	GLWR	E_3,E_4	chr6A	3.08E+08	3.47~7.59	5.03~18.87	3,2,4,5
Q.GW-6A	AX-95151036	GW	E_1,E_2	chr6A	6.00E+08	3.07~3.11	4.49~4.79	3,3
Q.GL-TGW-GL-6A	AX-95238912	GL,TGW,GL	E_3,E_4,E_5	chr6A	3.63E+08	3.46~8.23	2.6~23.8	4,4,1,2,3,4,5
Q.TGW-6A	AX-95240001	TGW	E_2,E_4	chr6A	1.61E+08	3.06~6.9	2.32~5.5	4,2,3
Q.GL–6B	AX-94405863	GL,	E_6,E_7	chr6B	5.39E+08	5.44~7.24	7.15~13.4	1,2,4,5,4,5
Q.GW-6D	AX-95143327	GW	E_3,E_4	chr6D	7088328	3.16~4.01	3.9~7.26	3,5,3,4
Q.GLWR-GL-7A	AX-94820170	GLWR,GL	E_1,E_3	chr7A	68087433	3.26~5.69	0.56~9.02	2,4,4,5
Q.GL-7A	AX-95130728	GL	E_2,E_4	chr7A	68861347	3.97~6.44	4.52~7.2	4,4,5
Q.GLWR-GL-7B	AX-94551830	GLWR,GL	E_1,E_2	chr7B	6180054	3.89~7.01	0~4.27	2,5
Q.GLWR-7B	AX-94754235	GLWR	E_1,E_2	chr7B	6.26E+08	3.81~5.17	2.72~8.3	3,4,5
Q.GLWR-GL-7B	AX-94928980	GLWR,GL	E_1,E_3	chr7B	6153426	3.62~4.26	0~5.2	2,5
Q.GL-7B	AX-95181207	GL	E_1,E_2	chr7B	6.88E+08	3.28~5.79	3.54~9.82	1,3,5,5
Q.GW-GLWR-7D	AX-94460120	GW,GLWR	E_1,E_2	chr7D		3.48~7.27	0.69~10.14	4,3,3,4,5
Q.GW-7D	AX-94674756	GW	E_1,E_5	chr7D	5.78E+08	3.44~3.81	1.63~6.26	1,4,5
Q.GL-7D	AX-94872194	GL	E_1,E_3,E_4	chr7D	6.34E+08	3.32~8	4.82~21.88	4,2,4,5
Q.GW-7D	AX-95009696	GW,GW	E_3,E_4	chr7D	4.72E+08	3.56~4.04	2.05~3.55	2,5
Q.GL-7D	AX-95014664	GL	E_1,E_2,E_4	chr7D	1.03E+08	3.22~6.76	5.76~8.2	2,4,1,2,4
Q.GW-7D	AX-95127613	GW	E_3,E_4	chr7D	1.01E+08	5.45~6.53	8.82~9.18	3,3

1, mrMLM; 2, FASTmrMLM; 3, FASTmrEMMA; 4, pLARmEB; 5, ISIS EM-BLASSO.

Discussion

Grain yield is a highly complex agronomic trait, governed by several genes and also influenced by environmental conditions (Li et al., 2022). It is essentially determined by two main components i.e., number of grains per m^2 and thousand grain weight (TGW) (Sun et al., 2009; Kumari et al., 2018; Li et al., 2019). In the breeding history, grain yield was mainly improved with increase in the grain number per m^2, which is determined by grain number per spike (Kumari et al., 2018). In the present study, we focused on the grain

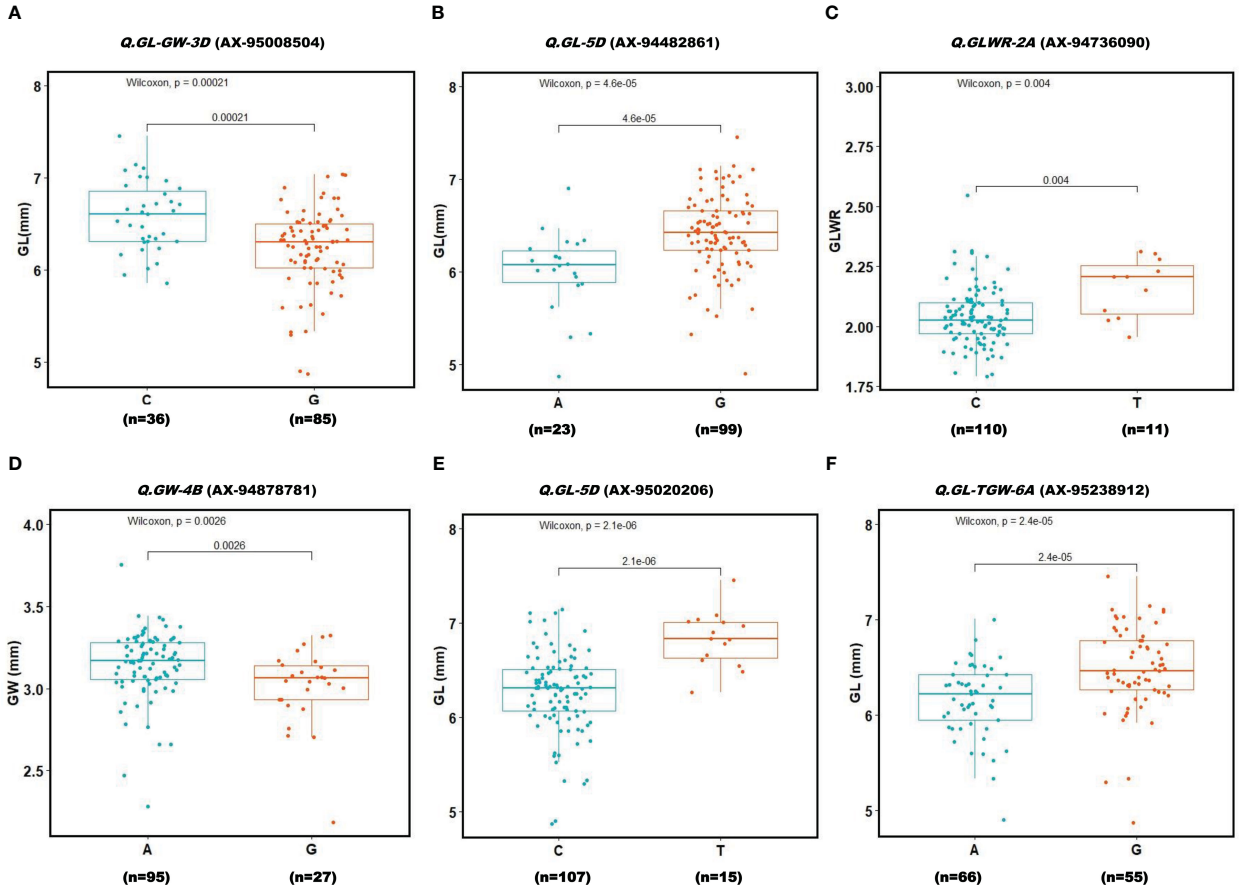


FIGURE 7
Boxplot for 6 reliable QTNs **(A–F)**. Genotypes were divided into two groups at each locus based on the allele type. A significant difference between the phenotype of these two groups was analyzed using t-test ($P \leq 0.001$). Two alleles for each QTN (Locus) are given on X-axis. Y-axis shows phenotypic values of the traits.

shape traits i.e. GL, GW, and GLWR, which determine TGW, a phenotypically stable yield contributing trait and used by the breeders for selecting high yielding varieties (Avni et al., 2018; Duan et al., 2020). The TGW and other grain size traits contribute to higher grain yield than grain number per spike. (Ji et al., 2022). Thus, it is very important to study the grain shape traits when the aim is to improve grain yield. Here, we have applied GWAS to identify genomic regions regulating variation for grain yield in a sub-set of the Indian National Genebank wheat mini core set germplasm (125 accessions). These mini core set accessions have been identified from a core set (2226 accessions), constituted from the entire wheat accessions (22416) conserved in the National Genebank of India (Phogat et al., 2020). Therefore, the mini core set accessions are a valuable genetic resource for mapping various desirable traits including grain shape traits.

The phenotyping of wheat mini core set accessions across the seven environments revealed significant variability among wheat accessions for grain parameters. High coefficients of variation for TGW under all the environments indicated broad phenotypic variation and a large improvement potential. Heritability is the proportion of genotypic variance to all observable variance in the total population. Over the environments, heritability was high for GL and moderate for GW, GLWR, and TGW. The trend of heritability is more specific to environment than traits, as we observed low/moderate heritability for E_5 and E_6. These environments fall in stress prone areas affected by less rainfall and high temperatures, which might have caused low heritability of the traits. Promising accessions for grain parameters were identified. Among them, EC578134, IC539313, IC535217, EC464070, and EC578152 were promising for GL as well as TGW. EC339611, EC578134, and IC535217 were promising for GL as well as GLWR, whereas IC335715 was promising for GW and TGW. These accessions can be used in breeding programs for trait introgression, genetics, and genomics study. The significant positive correlation of grain length and width with thousand grain weight revealed that the selection of grains with increased width and length can greatly contribute to grain weight and indirectly to grain yield. Earlier studies have reported moderate to strong correlations between TGW and size (Rasheed et al., 2018). Simmonds et al. (2016) reported that GL and GW in tetraploid and hexaploid wheat can greatly influence the TGW, as longer and broader grains have more starch accumulation and, hence, a higher weight (Simmonds et al., 2016). Previous studies have also reported positive associations among TGW, GL, and GW (Breseghello and Sorrells, 2007; Ramya et al., 2010). Principal component analysis also found GW, TGW, and GL as major contributing traits positively contributing to variations in grain shape among wheat genotypes

10.3389/fpls.2023.1148658

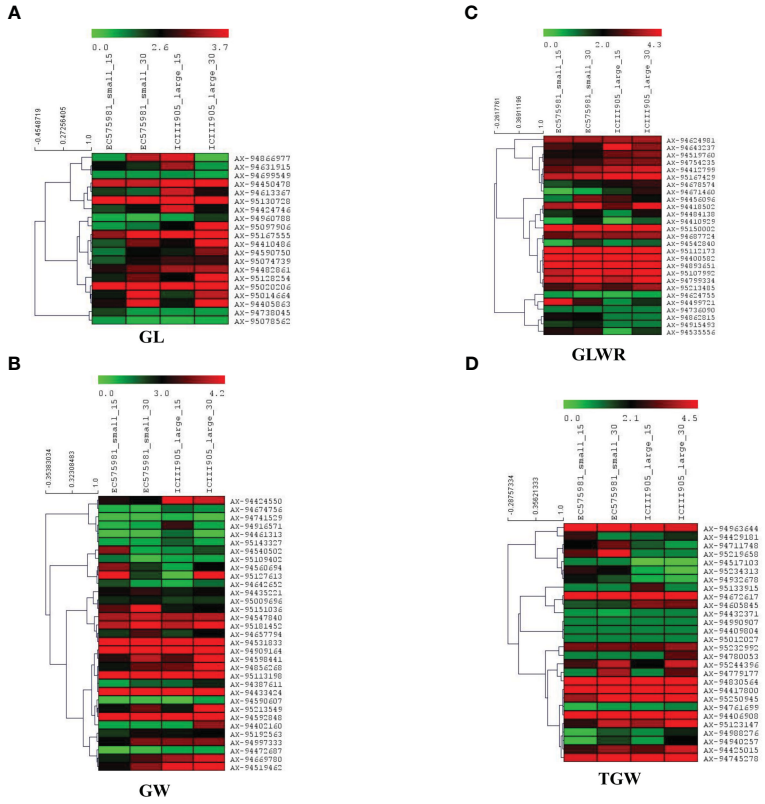


FIGURE 8
Heat maps of candidate genes identified for four-grain shape traits in small and large-size seeded wheat genotypes at the two developmental stages (15 and 30 DPAs). The figure panels show heat map for traits **(A)** GL, **(B)** GW, **(C)** GLWR, and **(D)** TGW. The genotype names are suffixed with 15 and 30, which indicate number of days after post-anthesis. Red indicates higher gene expression while green represents lower gene expression level; the gene expression levels are log2 transformed.

(Figure 1). In our study, the association between GW and GLWR is consistent and significantly negative in all environments. A negative correlation between GW and GLWR could be attributed to compensation of photosynthates to GW rather than to GL. The different correlations could be explained by the influence of the environment on the plant growth and grain development. This study shows that GL, GW, and GLWR are all expected to increase with TGW, one of the major yield components of grain yield and which can be targeted to enhance wheat yield potential. Genetic diversity and population structure in the wheat mini core subset was analyzed using 35K wheat SNP array. Both STRUCTURE and PCA analyses revealed two subpopulations in the wheat mini core set germplasm used in our study. The whole genome LD decay distance in the wheat mini core set was 1.93Mb. Further, LD decay was most rapid in A genome followed by D and B sub-genome. Many earlier studies in wheat have reported much longer LD decay distance ranging from 4Mb to 15 Mb or even more (Pang et al., 2020; Hanif et al., 2021; Li et al., 2021). This suggested the presence of a big LD block size, which has so far limited high-resolution trait mapping in wheat. One of the ways to overcome this problem is to use a very high density genic-SNP array having lakhs of SNPs derived from the coding regions for genotyping of association panels used for conducting GWAS. The high-density genotyping would facilitate construction of haplotypes maps of the associated regions that may help us in pinpointing the exact causal SNP/genes for the target traits.

The Genome-wide association study (GWAS) has been found to be a powerful tool to investigate genetic bases of complex traits in many plant species such as rice, maize, soybean, and wheat (Zegeye et al., 2014; Zhang et al., 2015; Spindel et al., 2016; Zhang et al., 2018; Chaurasia et al., 2020; Chaurasia et al., 2021). There are many statistical methods based on different algorithms that can be used to predict the true association between SNP markers and corresponding phenotypic variations in GWAS (Spindel et al., 2016). In our study, we used the ML-GWAS method for the detection of marker trait-associations for grain shape traits. Multi-locus methods are effective because of their higher statistical power which provides higher efficiency and accuracy for QTNs detection. In numerous studies, it has been found that ML-GWAS is much better than other methods (Bennett et al., 2012; Visioni et al., 2013; Spindel et al., 2016; Ma et al., 2018; Xu et al., 2018; Khan et al., 2019). Peng et al. (2018) used six ML-GWAS models to detect the genetic dissection of 20 free amino acid (AA) levels in *T. aestivum* and claimed that ML-GWAS methods are more reliable and powerful. In the current study, we used five multi-locus methods, mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, and ISIS EM-BLASSO, to perform GWAS analysis of four agronomic traits in our association panel. Among these five models, pLARmB identified the highest number of QTNs (211 SNPs), followed by FASTmrMLM (202), ISI EMBLASSO (177), mrMLM(132), and FASTmrEMMA(62).

QTNs for thousand grain weight, grain length, grain width and grain length width ratio

QTL for grain yield component traits have been extensively studied and reported on all the 21 chromosomes of wheat (Brinton et al., 2017; Cao et al., 2019; Ma et al., 2019; Ji et al., 2022). In our analysis, a total of 160 reliable QTNs were detected for four grain shape-related traits across the seven locations (Table 2; Supplementary Table 6).

For grain length, 27 QTNs were detected, which were distributed on 17 wheat chromosomes (chr1B, chr1D, chr2B, chr2D, chr3A, chr3B, chr3D, chr4A, chr4B, chr5A, chr5B, chr5D, chr6B, chr6D, chr7A, chr7B, and chr7D). Among these 27 QTNs, 10 QTNs were major ($R^2 \geq 10\%$ at least in one GWAS method), of which *Q.GL-4A* (SNP-AX-94839917), *Q.GL-7D* (SNP-AX-94872194), *Q.GL-7A* (SNP-AX-94760450) and *Q.GL-6D* (SNP-AX-94647721) were strongest because their R^2 value were ≥20%. The *Q.GL-4A* on the chr4A with highest R^2 = 31.94% may explain a significant proportion of the variation for GL in the wheat mini core germplasm. Moreover, this QTN was identified simultaneously in two environments i.e., E_1 and E_2, and with four different models. The *Q.GL-7D* is located within a gene encoding Thioredoxin M type protein with R^2 ranging between 4.82% to 21.88%. This QTN was predicted in three different environments i.e., E_1, E_3, and E_4, and using three different models, suggesting this could be a reliable QTN contributing to GL variation in wheat. In an earlier study, Thioredoxin has been shown to play an important role in preventing sprouting of developing grains in cereals (wheat and barley) by reducing the intramolecular disulfide bonds of storage proteins and other proteins in the starchy endosperm, and thereby affecting grain yield (Guo et al., 2013).

For grain width, 36 QTNs were identified that were distributed on 17 wheat chromosomes. Among these QTNs, *Q.GW-3D* (SNP-AX-94642652), *Q.GW-5B* (SNP-AX-94547840), *Q.GW-3A* (SNP-AX-94741529), *Q.GW-4D* (SNP-AX-95213549), *and Q.GW-2B* (SNP-AX-94519462) were predicted as major QTNs as the phenotypic variance explained by these QTLs was ≥10% of at least one of the ML- GWAS models. *Q.GW-3D and Q.GW-5B* were annotated as unnamed protein product and hypothetical protein respectively. Additionally, *Q.GW-2B* and *Q.GW-4D* had R^2 values ranging from 7.7 to 12.23 and 0.72 to 17.72, respectively. *Q.GW-2B* was identified at two environments E_3 and E_4, while *Q.GW-4D* was identified at E_2. Interestingly, both intragenic SNPs showed higher expression in large grain wheat cultivars than small seed cultivars. This suggested that these QTNs might have important roles in determining variation for GW in wheat.

Thirty-seven and thirty-five QTNs were predictive for GLWR and TGW traits respectively. The GLWR-associated QTNs were distributed on 17 chromosomes (chr1A, chr1D, chr2A, chr2B, chr2D, chr3A, chr3B, chr3D, chr4A, chr4B, chr5A, chr5B, chr6A, chr6B, chr6D, chr7B, and chr7D) while QTNs for TGW were spread over 16 chromosomes (chr1A, chr1B, chr1D, chr2A, chr2B, chr2D, chr3A, chr3B, chr3D, chr4B, chr4D, chr5A, chr5D, chr6A, chr7B, and chr7D). In TGW, a total of fourteen SNPs had $R^2 \geq 10$ and were considered as major genomic regions for this trait. *Q.TGW-1A* (SNP-AX-94605845) was annotated as TTL1 protein (TETRATRICOPEPTIDE-REPEAT THIOREDOXIN-LIKE 1) with R^2 = 11.78% and highly expressed in large grains. Studies have reported that *TTL1* positively regulates the stress response regulated by ABA (Guo et al., 2013). The loss of *TTL1* function causes plants to be sensitive to salt and osmotic stress during seed germination and later development (Rosado et al., 2006). So, it could be possible that the identified genomic region in our study may positively regulate the expression of *TTL1* gene and regulate seed maturation. *Q.TGW-5D* (SNP-AX-95234313) was located within a gene encoding cytochrome 450 and it was only identified at E4 environment with R^2 = 21.45. In a previous study on cytochrome family protein, CYP78A3 on chr7 has been shown to play an important role in wheat seed development by promoting integument cell proliferation (Ma et al., 2015). Thus, it could be suggested that *Q.TGW-5D* (cytochrome 450) identified in our study might also have some role in seed development. A total of 13 QTNs were associated with GLWR and were considered as strong QTNs explaining ≥10% phenotyping variance of the trait. Most of the QTNs were annotated to be either hypothetical proteins or intergenic SNPs. Three QTNs for TGW, namely *Q.GLWR-2D* (SNP-AX-94922377), *Q.GLWR-1A* (SNP-AX-95213485), and *Q.GLWR-6A* (SNP-AX-94722285), were simultaneously identified in three different environments and located on chr2D, chr1A, and chr6A respectively.

Comparison of the QTLs identified in the present and previous studies

In wheat, several candidate genes underlying grain size and weight have been identified including *TaGS* (Bernard et al., 2008), *TaGW2* (Su et al., 2011), *TaGS-D1* (Zhang et al., 2014), *TaCWI* (Jiang et al., 2015), and *Tackx4* (Chang et al., 2015). Additionally, McCartney et al. (2005) identified two major QTLs for TKW responsible for reduced plant height that were near the *Rht-B1b* and *Rht-D1b* genes that control plant height (McCartney et al., 2005; Gao et al., 2015). Another QTL, *Qtgw-cb.5A*, was identified as a key determinant of final grain weight which increased grain length by driving pericarp cell expansion (Brinton et al., 2017). We performed the comparative analysis of QTNs for grain shape identified in the present study with previously reported QTLs on the basis of their physical locations on chromosomes. Some of the previously reported grain size-associated QTLs were also predicted in our analysis. For example, *Qgl.cib-CK1-4A* associated with GL on chr4A coincided with *Q.GL-4A* (SNP-AX-94839917) for grain length trait at the same region on chr4A and identified in two environments (E_1 and E_2). Further, LOD (4.92~6.13) and R2 value (15.33~31.94) of this QTL demonstrated its importance in regulating GL trait. Goel et al. (2019) identified *qTKW.6A.1* associated with TGW on 6A at the interval 166.64-596.18 Mb (Goel et al., 2019). The QTN, *Q.TGW-6A* (SNP-AX-95240001), identified in our study appears to correspond to *qTKW.6A.1*. Interestingly, *Q.TGW-6A* was identified at two locations, E2 and E4, which showed that it is a stable genomic region for TGW. Further, we found that *Q.GL-TGW-6A* (SNP-AX-95238912) and *Q.GLWR-6A* (SNP-AX-94722285), which are located on chr6A at 362.7Mb and 307Mb, overlapped with the grain shape QTLs identified by Cao et al.

(2019) and Ji et al. (2022), respectively. Interestingly, *Q.GL-TGW-6A* was identified in three environments (E3, E4, and E5). LOD score and R^2 value of *Q.GL-TGW-6A and Q.GLWR-6A* ranged from 3.46 to 8.23 and 2.6 to 23.8 respectively. On the other hand, *Q.GLWR-6A* was present at 307Mb on chr6A with LOD score (3.47~7.59) and R2 value (5.03~18.87). Since the two QTNs on chr6A were also identified in the previous studies, these appear to represent major genomic regions for the grain shape traits.

A few other underlying genes influenced grain size and weight have been reported by Cabral et al., 2018. *TaGS-D1*, controlling GL and grain weight, is an ortholog of *OsGs3 and* located at 106.73 Mb on chr3D. Expression pattern of this *TaGS-D1* (*TraesCS7A03G0037700*) gene in our data showed relatively higher expression in large seeded genotypes as compare to small seeded genotypes. So, we examined nearby QTNs around the gene and we found two QTNs, *Q.GLWR-7*D and *Q.TGW-7D*, located in the vicinity of *TaGS-D1* and positioned at 54.9Mb on chr7DS and 100.1 Mb on chr7D respectively. The presence of these two QTNs indirectly suggested a major locus which corresponds to either *TaGS-D1* or an additional novel gene for grain shape trait on the short of chr7D. A second grain weight locus cytokinin oxidase/dehydrogenase (*TaCKX6-D1*) gene is physically located on chromosome 3D and played a key role in controlling cytokinin levels and affects grain weight in wheat (Zhang et al., 2012). *TaCKX6-D1* gene is located at 106.73 Mb on chr3D, so its expression could be influenced by nearby SNPs around the gene. On the basis of the physical location of gene, we found two significant genes, *Q.GL-GW-3D* and *Q.TGW-3D* (SNP-AX-95008504, and SNP-AX-94406908), in our analysis at 151.4 Mb and 239.3 Mb respectively. *Q.GL-GW-3D* associated with GL was identified in four environments (E_1, E_2, E_4, and E_5) with LOD value from 3.05 to 6.33 and R^2 value from 4.09 to 14.97, which showed the significance of SNP. The second, *Q.TGW-3D*, demonstrated association with TGW with LOD (3.16~8.2) and R^2 (7.21~11.48) and was identified at E2, E3, and E4 environments. Both the QTNs were annotated as hypothetical proteins and expressed in our transcriptome data. *Q.TGW-3D* was highly expressed in large seed cultivars while *Q.GL-GW-3D* also showed expression in both cultivars. In conclusion, in this study we have comprehensively phenotyped wheat mini core germplasm accessions for grain shape traits and identified promising accessions for large grain size and length which can be incorporated in breeding programs. Further, integration of phenotyping and genotyping data has enabled us to identify genomic regions/candidate genes, some of which are novel. Comparative study also showed that many QTNs identified in our study represented novel genomic regions that can be further validated for their role in determining grain size and can be potentially exploited in breeding programs to develop high-yielding varieties.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA906296.

Author contributions

JK planned and handled the field experiments, conducted formal analysis, and assisted in writing the manuscript. DL contributed to the data curation, investigation, formal analysis, and original draft writing. MY and ST planned and executed DNA extraction and SNP data generation. PJ, SS, ST, SM, and HA performed data analysis, writing, reviewing, and editing, NS, KS and KM recorded phenotyping data and reviewed and edited the manuscript, RS, MY, and GS contributed resources and participated in the writing, reviewing, and editing. AS contributed to the conceptualization, supervision, writing, reviewing, and editing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Financial support for this study was received from the Science and Engineering Research Board (project: EMR/2017/005133) and ICAR-National Innovations in Climate Resilient Agriculture (NICRA) Project (Project code 1006607)

Acknowledgments

The authors are thankful to the Director of the Indian Council of Agricultural Research (ICAR)-National Bureau of Plant Genetic Resources (NBPGR) for providing laboratory and field facilities needed to undertake this work.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1148658/full#supplementary-material

References

Allen, A. M., Winfield, M. O., Burridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L., et al. (2017). Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat *(Triticum aestivum)*. *Plant Biotechno J.* 15, 390–401. doi: 10.1111/pbi.12635

Avni, R., Oren, L., Shabtay, G., Assili, S., Pozniak, C., Hale, I., et al. (2018). Genome based meta-QTL analysis of grain weight in tetraploid wheat identifies rare alleles of GRF4 associated with larger grains. *Genes* 9, 636. doi: 10.3390/genes9120636

Bennett, D., Izanloo, A., Reynolds, M., Kuchel, H., Langridge, P., and Schnurbusch, T. (2012). Genetic dissection of grain yield and physical grain quality in bread wheat *(Triticum aestivum L.)* under water-limited environments. *Theor. Appl. Genet.* 125, 255–271. doi: 10.1007/s00122-012-1831-9

Bernard, S. M., Moller, A. L., Dionisio, G., Kichey, T., Jahn, T. P., Dubois, F., et al. (2008). Gene expression, cellular localisation and function of glutamine synthetase isozymes in wheat *(Triticum aestivum L.)*. *Plant Mol. Biol.* 67, 89–105. doi: 10.1007/s11103-008-9303-y

Breseghello, F., and Sorrells, M. E. (2007). QTL analysis of kernel size and shape in two hexaploid wheat mapping populations. *Field Crops Res.* 101, 172–179. doi: 10.1016/j.fcr.2006.11.008

Brinton, J., Simmonds, J., Minter, F., Leverington-Waite, M., Snape, J., and Uauy, C. (2017). Increased pericarp cell length underlies a major quantitative trait locus for grain weight in hexaploid wheat. *TNew Phytol.* 215, 1026–1038. doi: 10.1111/nph.14624

Cabral, A. L., Jordan, M. C., Larson, G., Somers, D. J., Humphreys, D. G., and McCartney, C. A. (2018). Relationship between QTL for grain shape, grain weight, test weight, milling yield, and plant height in the spring wheat cross RL4452/'AC domain'. *PloS One* 13, e0190681. doi: 10.1371/journal.pone.0190681

Cao, P., Liang, X., Zhao, H., Feng, B., Xu, E., Wang, L., et al. (2019). Identification of the quantitative trait loci controlling spike-related traits in hexaploid wheat *(Triticum aestivum L.)*. *Planta* 250, 1967–1981. doi: 10.1007/s00425-019-03278-0

Chang, C., Lu, J., Zhang, H. P., Ma, C. X., and Sun, G. (2015). Copy number variation of cytokinin oxidase gene Tackx4 associated with grain weight and chlorophyll content of flag leaf in common wheat. *PloS One* 10, e0145970. doi: 10.1371/journal.pone.0145970

Chaurasia, S., Singh, A. K., Kumar, A., Songachan, L. S., Yadav, M. C., Kumar, S., et al. (2021). Genome-wide association mapping reveals key genomic regions for physiological and yield-related traits under salinity stress in wheat *(Triticum aestivum L.)*. *Genomics* 113, 3198–3215. doi: 10.1016/j.ygeno.2021.07.014

Chaurasia, S., Singh, A. K., Songachan, L. S., Sharma, A. D., Bhardwaj, R., and Singh, K. (2020). Multi-locus genome-wide association studies reveal novel genomic regions associated with vegetative stage salt tolerance in bread wheat *(Triticum aestivum L.)*. *Genomics* 112, 4608–4621. doi: 10.1016/j.ygeno.2020.08.006

Duan, X., Yu, H., Ma, W., Sun, J., Zhao, Y., Yang, R., et al. (2020). A major and stable QTL controlling wheat thousand grain weight: identification, characterization, and CAPS marker development. *Mol. Breed.* 40, 68. doi: 10.1007/s11032-020-01147-3

Evanno, G., Regnaut, S., 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

Gao, F., Wen, W., Liu, J., Rasheed, A., Yin, G., Xia, X., et al. (2015). Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross zhou 8425B/Chinese spring. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01099

Goel, S., Singh, K., Singh, B., Grewal, S., Dwivedi, N., Alqarawi, A. A., et al. (2019). Analysis of genetic control and QTL mapping of essential wheat grain quality traits in a recombinant inbred population. *PloS One* 14, e0200669. doi: 10.1371/journal.pone.0200669

Guo, H., Wang, S., Xu, F., Li, Y., Ren, J., Wang, X., et al. (2013). The role of thioredoxin h in protein metabolism during wheat *(Triticum aestivum L.)* seed germination. *Plant Physiol. Biochem.* 67, 137–143. doi: 10.1016/j.plaphy.2013.03.006

Hanif, U., Alipour, H., Gul, A., Jing, L., Darvishzadeh, R., Amir, R., et al. (2021). Characterization of the genetic basis of local adaptation of wheat landraces from Iran and Pakistan using genome-wide association study. *Plant Genome* 14(3), e2009. doi: 10.1002/tpg2.20096

IWGSC, Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361 (6403), eaar7191. doi: 10.1126/science.aar7191

Ji, G., Xu, Z., Fan, X., Zhou, Q., Chen, L., Yu, Q., et al. (2022). Identification and validation of major QTL for grain size and weight in bread wheat (Triticum aestivum l.). *Crop J* 11, 564–572. doi: 10.1016/j.cj.2022.06.014

Jiang, Y., Jiang, Q., Hao, C., Hou, J., Wang, L., Zhang, H., et al. (2015). A yield-associated gene *TaCWI*, in wheat: its function, selection and evolution in global breeding revealed by haplotype analysis. *Theor. Appl. Genet.* 128, 131–143. doi: 10.1007/s00122-014-2417-5

Kato, K., Miura, H., and Sawada, S. (2000). Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. *Theor. Appl. Genet.* 101, 1114–1121. doi: 10.1007/s001220051587

Khan, S. U., Yangmiao, J., Liu, S., Zhang, K., Khan, M. H. U., Zhai, Y., et al. (2019). Genome-wide association studies in the genetic dissection of ovule number, seed number, and seed weight in Brassica napus L. *Ind. Crops Prod.* 142, 111877. doi: 10.1016/j.indcrop.2019.111877

Kumari, S., Jaiswal, V., Mishra, V. K., Paliwal, R., Balyan, H. S., and Gupta, P. K. (2018). QTL mapping for some grain traits in bread wheat *(Triticum aestivum L.)*. *Physiol. Mol. Biol. Plants* 24, 909–920. doi: 10.1007/s12298-018-0552-1

Li, T., Deng, G., Su, Y., Yang, Z., Tang, Y., Wang, J., et al. (2022). Genetic dissection of quantitative trait loci for grain size and weight by high-resolution genetic mapping in bread wheat *(Triticum aestivum L.)*. *Theor. Appl. Genet.* 135, 257–271. doi: 10.1007/s00122-021-03964-2

Li, Y., Tang, J., Liu, W., Yan, W., Sun, Y., Che, J., et al. (2021). The genetic architecture of grain yield in spring wheat based on genome-wide association study. *Front. Genet.* 12. doi: 10.3389/fgene.2021.728472

Li, N., Xu, R., and Li, Y. (2019). Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* 70, 435–463. doi: 10.1146/annurev-arplant-050718-095851

Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., et al. (2013). Draft genome of the wheat a-genome progenitor *Triticum urartu*. *Nature* 496, 87–90. doi: 10.1038/nature11997

Ma, L., Liu, M., Yan, Y., Qing, C., Zhang, X., Zhang, Y., et al. (2018). Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-wide association studies. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00561

Ma, M., Wang, Q., Li, Z., Cheng, H., Li, Z., Liu, X., et al. (2015). Expression of TaCYP78A3, a gene encoding cytochrome P450 CYP78A3 protein in wheat *(Triticum aestivum L.)*, affects seed size. *Plant J.* 83, 312–325. doi: 10.1111/tpj.12896

Ma, J., Zhang, H., Li, S., Zou, Y., Li, T., Liu, J., et al. (2019). Identification of quantitative trait loci for kernel traits in a wheat cultivar Chuannong16. *BMC Genet.* 20, 77. doi: 10.1186/s12863-019-0782-4

Malik, P., Kumar, J., Sharma, S., Sharma, R., and Sharma, S. (2021). Multi-locus genome-wide association mapping for spike-related traits in bread wheat *(Triticum aestivum L.)*. *BMC Genomics* 22, 597. doi: 10.1186/s12864-021-07834-5

McCartney, C. A., Somers, D. J., Humphreys, D. G., Lukow, O., Ames, N., Noll, J., et al. (2005). Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452x'AC domain'. *Genome* 48, 870–883. doi: 10.1139/g05-055

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. United States America* 110 (2), 453–458. doi: 10.1073/pnas.1215985110

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8 (19), 4321–4325. doi: 10.1093/nar/8.19.4321

Nehe, A., Akin, B., Sanal, T., Evlice, A. K., Unsal, R., Dincer, N., et al. (2019). Genotype x environment interaction and genetic gain for grain yield and grain quality traits in Turkish spring wheat released between 1964 and 2010. *PloS One* 14, e0219432. doi: 10.1371/journal.pone.0219432

Newell, M. A., Cook, D., Tinker, N. A., and Jannink, J. L. (2011). Population structure and linkage disequilibrium in oat *(Avena sativa L.)*: implications for genome-wide association studies. *TAG Theo.r Appl. Genet.s* 122, 623–632. doi: 10.1007/s00122-010-1474-7

Pang, Y., Liu, C., Wang, D., St Amand, P., Bernardo, A., Li, W., et al. (2020). High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. *Mol. Plant* 13 (9), 1311–1327. doi: 10.1016/j.molp.2020.07.008

Peng, Y., Liu, H., Chen, J., Shi, T., Zhang, C., Sun, D., et al. (2018). Genome-wide association studies of free amino acid levels by six multi-locus models in bread wheat. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01196

Phogat, S., Kumar, S., Kumari, J., Kumar, N., Pandey, A., Singh, T., et al. (2020). Characterization of wheat germplasm conserved in the Indian national genebank and establishment of a composite core collection. *Crop Sci.* 61, 604–620. doi: 10.1002/csc2.20285

Ramya, P., Chaubal, A., Kulkarni, K., Gupta, L., Kadoo, N., Dhaliwal, H. S., et al. (2010). QTL mapping of 1000-kernel weight, kernel length, and kernel width in bread wheat *(Triticum aestivum L.)*. *J. Appl. Genet.* 51, 421–429. doi: 10.1007/BF03208872

Rasheed, A., Xia, X., Ogbonnaya, F., Mahmood, T., Zhang, Z., Kazi., A. M., et al. (2018). Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. *BMC Plant Biol.* 14. doi: 10.1186/1471-2229-14-128

Ren, W.-L., Wen, Y.-J., Dunwell, J. M., and Zhang, Y.-M. (2018). pKWmEB: integration of kruskal–Wallis test with empirical bayes under polygenic background control for multi-locus genome-wide association study. *Heredity* 120, 208–218. doi: 10.1038/s41437-017-0007-4

Rodríguez, F., Alvarado, G., Pacheco, Á., and Burgueño, J. (2018). ACBD-R. Augmented Complete Block Design with R for Windows. Version 4.0. CIMMYT Research Data & Software Repository Network; Texcoco de Mora, Mexico

Rosado, A., Schapire, A. L., Bressan, R. A., Harfouche, A. L., Hasegawa, P. M., Valpuesta,, et al. (2006). The arabidopsis tetratricopeptide repeat-containing protein

TTL1 is required for osmotic stress responses and abscisic acid sensitivity. *Plant Physiol.* 142, 1113–1126. doi: 10.1104/pp.106.085191

Simmonds, J., Scott, P., Brinton, J., Mestre, T. C., Bush, M., Del Blanco, A., et al. (2016). A splice acceptor site mutation in TaGW2-A1 increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. *Theor. Appl. Genet.* 129, 1099–1112. doi: 10.1007/s00122-016-2686-2

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redona, E., Jannink, J. L., et al. (2016). Genome-wide prediction models that incorporate *de novo* GWAS are a powerful new tool for tropical rice improvement. *Heredity* 116, 395–408. doi: 10.1038/hdy.2015.113

Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat *(Triticum aestivum L.)*. *Theor. Appli. Genet.* 122, 211–223. doi: 10.1007/s00122-010-1437-z

Sun, C., Dong, Z., Zhao, L., Ren, Y., Zhang, N., and Chen, F. (2020). The wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. *Plant Biotechnol. J.* 18, 1354–1360. doi: 10.1111/pbi.13361

Sun, X.-Y., Wu, K., Zhao, Y., Kong, F.-M., Han, G.-Z., Jiang, H.-M., et al. (2009). QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. *Euphytica* 165, 615–624. doi: 10.1007/s10681-008-9794-2

Tamba, C. L., Ni, Y. L., and Zhang, Y. M. (2017). Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PloS Comput. Biol.* 13, e1005357. doi: 10.1371/journal.pcbi.1005357

Visioni, A., Tondelli, A., Francia, E., Pswarayi, A., Malosetti, M., Russell, J., et al. (2013). Genome-wide association mapping of frost tolerance in barley *(Hordeum vulgare L.)*. *BMC Genomics* 14, 424. doi: 10.1186/1471-2164-14-424

Wang, S. B., Feng, J. Y., Ren, W. L., Huang, B., Zhou, L., Wen, Y. J., et al. (2016). Improving power and accuracy of genome-wide association studies *via* a multi-locus mixed linear model methodology. *Sci. Rep.* 6, 19444. doi: 10.1038/srep19444

Wen, Y. J., Zhang, H., Ni, Y. L., Huang, B., Zhang, J., Feng, J. Y., et al. (2017). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* 18, 906. doi: 10.1093/bib/bbx028

Xu, Y., Xu, C., and Xu, S. (2017). Prediction and association mapping of agronomic traits in maize using multiple omic data. *Heredity* 119, 174–184. doi: 10.1038/hdy.2017.27

Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., et al. (2018). Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01311

Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opini Biotechnol.* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003

Zegeye, H., Rasheed, A., Makdis, F., Badebo, A., and Ogbonnaya, F. C. (2014). Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PloS One* 9, e105593. doi: 10.1371/journal.pone.0105593

Zhang, Y. W., Tamba, C. L., Wen, Y. J., Li, P., Ren, W. L., Ni, Y. L., et al. (2020). mrMLM v4.0.2: an r platform for multi-locus genome-wide association studies. *Genomics Proteomics Bioinf.* 18 (4), 481–487. doi: 10.1016/j.gpb.2020.06.006

Zhang, J., Feng, J. Y., Ni, Y. L., Wen, Y. J., Niu, Y., Tamba, C. L., et al. (2017). pLARmEB: integration of least angle regression with empirical bayes for multilocus genome-wide association studies. *Heredity* 118, 517–524. doi: 10.1038/hdy.2017.8

Zhang, Y., Liu, J., Xia, X., and He, Z. (2014). *TaGS-D1*, An ortholog of rice *OsGS3*, is associated with grain weight and grain length in common wheat. *Mol. Breed* 34, 1097–1107. doi: 10.1007/s11032-014-0102-7

Zhang, Y., Liu, P., Zhang, X., Zheng, Q., Chen, M., Ge, F., et al. (2018). Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00611

Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean *(Glycine max)* germplasm. *BMC Genomics* 16, 217. doi: 10.1186/s12864-015-1441-4

Zhang, L., Zhao, Y.-L., Gao, L.-F., Zhao, G.-Y., Zhou, R.-H., Zhang, B.-S., et al. (2012). *TaCKX6-D1*, The ortholog of rice *OsCKX2*, is associated with grain weight in hexaploid wheat. *New Phytol.* 195, 574–584. doi: 10.1111/j.1469-8137.2012.04194.x

Zhou, Q., Zhou, C., Zheng, W., Mason, A. S., Fan, S., Wu, C., et al. (2017). Genome-wide SNP markers based on SLAF-seq uncover breeding traces in rapeseed *(Brassica napus L.)*. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00648

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 July 2023
DOI 10.3389/fpls.2023.1107705

OPEN ACCESS

EDITED BY
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

REVIEWED BY
Shivani Nagar,
Indian Agricultural Research Institute (ICAR), India
Mohd. Anwar Khan,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India

*CORRESPONDENCE
Mohammad Jafar Tanin
jafartanin@gmail.com
Achla Sharma
achla@pau.edu

RECEIVED 25 November 2022
ACCEPTED 29 June 2023
PUBLISHED 17 July 2023

CITATION
Tanin MJ, Sharma A, Ram H, Singh S, Srivastava P, Mavi GS, Saini DK, Gudi S, Kumar P, Goyal P and Sohu VS (2023) Application of potassium nitrate and salicylic acid improves grain yield and related traits by delaying leaf senescence in *Gpc-B1* carrying advanced wheat genotypes. *Front. Plant Sci.* 14:1107705. doi: 10.3389/fpls.2023.1107705

Application of potassium nitrate and salicylic acid improves grain yield and related traits by delaying leaf senescence in *Gpc-B1* carrying advanced wheat genotypes

Mohammad Jafar Tanin*, Achla Sharma*, Hari Ram, Satinder Singh, Puja Srivastava, G. S. Mavi, Dinesh Kumar Saini, Santosh Gudi, Pradeep Kumar, Prinka Goyal and V. S. Sohu

Department of Plant Breeding and Genetics, College of Agriculture, Punjab Agricultural University, Ludhiana, Punjab, India

Grain protein content (GPC) is an important quality trait that effectively modulates end-use quality and nutritional characteristics of wheat flour-based food products. The *Gpc-B1* gene is responsible for the higher protein content in wheat grain. In addition to higher GPC, the *Gpc-B1* is also generally associated with reduced grain filling period which eventually causes the yield penalty in wheat. The main aim of the present study was to evaluate the effect of foliar application of potassium nitrate (PN) and salicylic acid (SA) on the physiological characteristics of a set of twelve genotypes, including nine isogenic wheat lines carrying the *Gpc-B1* gene and three elite wheat varieties with no *Gpc-B1* gene, grown at wheat experimental area of the Department of Plant Breeding and Genetics, PAU, Punjab, India. The PN application significantly increased the number of grains per spike (GPS) by 6.42 grains, number of days to maturity (DTM) by 1.03 days, 1000-grain weight (TGW) by 1.97 g and yield per plot (YPP) by 0.2 kg/plot. As a result of PN spray, the flag leaf chlorophyll content was significantly enhanced by 2.35 CCI at anthesis stage and by 1.96 CCI at 10 days after anthesis in all the tested genotypes. Furthermore, the PN application also significantly increased the flag leaf nitrogen content by an average of 0.52% at booting stage and by 0.35% at both anthesis and 10 days after anthesis in all the evaluated genotypes. In addition, the yellow peduncle colour at 30 days after anthesis was also increased by 19.08% while the straw nitrogen content was improved by 0.17% in all the genotypes. The preliminary experiment conducted using SA demonstrated a significant increase in DTM and other yield component traits. The DTM increased by an average of 2.31 days, GPS enhanced by approximately 3.17 grains, TGW improved by 1.13g, and YPP increased by 0.21 kg/plot. The foliar application of PN and SA had no significant effect on GPC itself. The findings of the present study suggests that applications of PN and SA can effectively mitigate the yield penalty associated with *Gpc-B1* gene by extending grain filling period in the wheat.

KEYWORDS

salicylic acid, potassium nitrate, *Gpc-B1* gene, isogenic lines, wheat

Introduction

Wheat (*Triticum aestivum* L.) is the second most widely cultivated crop in the world that is consumed on a daily basis in many parts of the world. Additionally, during the 2019–2020 growing season, total global wheat production and productivity reached records of 764 million metric tonnes and 3.52 metric tonnes per hectare, respectively (https://worldpopulationreview.com/country-rankings/wheat-production-by-country). The protein content of harvested wheat grains is one of the most important quality traits as it significantly modulates the quality of the final food product(s). The world population is growing at an alarming rate and providing enough nutritionally enriched food for this rapidly expanding population will be a critical challenge for breeders. The recommended daily protein amount required for adults (more than 18 years old) is 0.8 g per kg or human weight (Wolfe et al., 2017). Acceptable protein-enriched food intake is a prerequisite factor for meeting human metabolic requirements and ensuring standard health criteria, especially in light of the current climate change crisis (Tanin et al., 2022b).

Grain protein content (GPC) is one of the highly valuable grain quality characteristics that determine the processing quality, nutritional capacity, and grain market properties. Breeders have made several attempts to improve the GPC in wheat using conventional breeding methods, but the desired results have not been achieved. This could be due to: (1) the substantial effect of the environment on GPC; (2) the negative relationship of GPC with grain yield; and (3) the complex quantitative genetic control of GPC and its low heritability. QTLs for GPC have been identified and mapped on nearly all chromosomes of wheat (Shewry, 2009; Saini et al., 2020; Saini et al., 2022b). *Gpc-B1* is the most critical QTL identified for GPC so far (Uauy et al., 2006; Saini et al., 2022b; Singh K. et al., 2022, http://www.wheatqtldb.net/). This QTL was discovered in a wild accession of tetraploid wheat, *Triticumturgidum* var. *dicoccoides* (Uauy et al., 2006). Later, the same accession (i.e., FA15-3) was used to generate a complete set of chromosome substitution lines with the background of modern durum wheat (Joppa and Cantrell, 1990). The *Gpc-B1* gene was then mapped on chromosome arm 6BS, which explained 66% of the phenotypic variation of the GPC and significantly improved GPC by an average of 14 g kg^{-1} in hexaploid wheat and by an average of 3.3 g kg^{-1} in tetraploid wheat under different environmental conditions (Uauy et al., 2006). The map-based cloning of *Gpc-B1* revealed that encodes a NAC transcription factor (*NAM-B1*).

The *NAM-B1*/*Gpc-B1* gene has been found to be associated with early senescence, improved nitrogen remobilization from flag leaf to the developing grains, and pleiotropic effects on protein accumulation in wheat grains, yield, and key yield-related traits, including, 1000-grain weight (TGW), number of spikes, number of spikelets per spike, and number of grains per spike (Uauy et al., 2006; Tabbita et al., 2017). The introgression of the functional allele of the *Gpc-B1* gene into the background of advanced wheat genotypes has resulted in the release of numerous cultivars in different countries (see Tabbita et al., 2017). However, GPC enhancement in genotypes with the *Gpc-B1* gene in their genetic background has been associated with a reduction in grain size and eventually the grain yield. In other words, the increased GPC could be attributed to more efficient nutrient transport from flag leaf to developing grains. The association of the *Gpc-B1* gene with accelerated senescence could explain both mechanisms. Even in the optimum grain filling environment, the TGW could be reduced due to a shorter period of grain filling, but earlier senescence may result in more effective nutrient transportation due to earlier initiation of the translocation process (Kade et al., 2005; Tanin et al., 2018).

The higher chlorophyll degradation in flag leaf, the lower spike water content, and the increased percentage of yellow peduncle have also been found to be associated with *Gpc-B1* (Uauy et al., 2006). Plants with delayed senescence are known as “stay-green”, and cosmetic types result from impeded chlorophyll catabolism (Thomas and Ougham, 2014; Kamal et al., 2019). In multiple crops, functional stay-green phenotypes have been found to be associated with increased or prolonged photosynthetic activities, conferring tolerance to drought, heat, and low nitrogen stresses (Thomas and Ougham, 2014). Significant positive correlation between green canopy and grain filling duration (Luche et al., 2016; Pinto et al., 2016) endorse the potential breeding versatility of stay-green traits, as a prolonged grain filling period that may boost grain size and final grain yield.

Potassium nitrate (PN, KNO_3) is an inorganic salt that regulates different biochemical and physiological processes which influence plant growth and metabolism, including protein accumulation in grains, seed germination and emergence, stomatal regulation, phloem transport, cation-anion balance, photosynthesis, energy transfer, nutrient balance, and carbohydrate metabolism (Hasanuzzaman et al., 2018). The foliar application of PN showed the highest plant height, tiller number, spike length, and yield in an earlier study (Devi et al., 2017). In this study, uptake of nitrogen, phosphorus, and potassium was observed at the highest level with the PN spray at a rate of 3%. In another study, foliar spray of 0.5% PN at booting and anthesis stages increased the yield and yield component traits of wheat genotype HD2985 under late sown conditions (Singh and Singh, 2020). Similarly, salicylic acid (SA, $C_7H_6O_3$) is a phenolic compound that regulates plant growth and development as well as their responses to biotic and abiotic stresses (Khan et al., 2015). The foliar application of SA has been shown to increase grain filling period, TGW, GPS, grain yield, and some other key yield-related traits in wheat genotypes (Karim and Khursheed, 2011; Ziasmin et al., 2017; Singh et al., 2019; Jatana et al., 2021; Stanislawska-Glubiak and Korzeniowska, 2021; Abdi et al., 2022). The foliar application of 0.75 mM SA at 30 days after sowing (DAS) and seed treatment with 0.75 mM SA produced the best results under late sown and timely sown conditions, thereby, significantly increasing grain filling period, grain yield, proline content, and soluble sugar content in leaves, plant height, and tiller number in two wheat varieties (viz., NW5054 and NW2036) when compared to the control (Singh et al., 2019).

In addition to high GPC, *Gpc-B1 is* linked to the yellow rust resistance gene *Yr36*, and excellent gene-based markers are available for monitoring combination of these genes (Uauy et al., 2005). Lastly, pyramiding and introgression of major QTLs/genes for various traits of economic importance via marker-assisted

selection (MAS) have been shown to have the positive results in wheat (e.g., Sharma et al., 2021; Gill et al., 2022; Gupta et al., 2022). We at Punjab Agricultural University (PAU), Ludhiana, India have successfully used MAS for GPC enrichment in our wheat breeding programs, pyramiding the *Gpc-B1* gene and the stripe rust resistance gene *Yr15* in the background of a popular wheat cultivar "PBW550" using a unique combination of multi-disciplinary and multi-pronged approaches (e.g., Tanin et al., 2022b). This germplasm has higher GPC levels, which is important for reducing mass malnutrition and hunger. Consumers who are most affected by rising food prices will be the primary beneficiaries, ultimately improving food security and sustainable production.

The grain yield and GPC have been reported to have a negative correlation, which makes it difficult to increase GPC without sacrificing yield, developing high GPC wheat genotypes unlikely to be commercially successful due to yield penalty. This situation necessitates a detailed examination of how the wheat genotypes achieve high GPC, as well as a simple method of using economically feasible intervention(s) to compensate the yield losses associated with high GPC. To achieve the above-mentioned objective, the present study was planned to investigate the effects of foliar applications of PN and SA on senescence, grain filling period, yield, and yield components, with a particular emphasis on minimizing the yield penalty associated with the *Gpc-B1* gene in order to combine higher GPC with no yield penalty in an individual wheat genotype, combining both higher grain yield and increased GPC.

Materials and methods

A breeding program was previously initiated in the Department of Plant Breeding and Genetics, Punjab Agricultural University (PAU), to generate a set of near isogenic lines (NILs) with a special focus on transferring the high GPC loci *Gpc-B1* from a hard red spring wheat cultivar 'GLUPRO' into the genetic background of an elite wheat variety (PBW550). The wheat cultivar PBW550, widely known for its desirable yield and quality characteristics, was developed in PAU and released to be grown under irrigated conditions in the Northern Western Plains Zone, India. The current study employed nine NILs containing the *Gpc-B1* gene and three commercial wheat varieties lacking the *Gpc-B1* gene (Table 1). The successful introgression of high GPC loci into the backgrounds of all nine NILs was confirmed using a *Gpc-B1*-specific KASP marker (data not provided) (Tanin et al., 2022b).

Field management and agronomic practices

To study the effect of PN and SA on GPC, senescence and key yield related traits in wheat, two independent field experiments were conducted during the main season 2020-21 in the wheat experimental area of the Department of Plant Breeding and Genetics, PAU, Punjab, India. The experimental area is located at an altitude of 247 meters above the mean sea level at a latitude of N and a longitude of 75˚48´E and, therefore, experiences a sub-tropical and semi-arid climate with cold (November to January) and mild winters (February and March), and also hot and dry summer (April to July). Both PN (@2%) and SA (@75 ppm) were sprayed twice (first spray at booting stage and second spray at anthesis stage) on the wheat genotypes grown in different field experiments. Concentrations of PN and SA tested and developmental stages targeted were selected based on a thorough survey of the earlier published studies and preliminary experiments (data not shown) conducted earlier by us. A total of 2 mg of PN powder was dissolved in 100 ml of distilled water to obtain a 2% potassium nitrate solution. Similarly, 50 mg of standard SA was dissolved in 95% alcohol in a volumetric flask to dilution formula

TABLE 1 Wheat genotypes (nine NILs, and three elite varieties) evaluated in the current investigation.

Serial No.	Genotypes	*Gpc-B1*	Pedigree
1	BWL6964	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
2	BWL7503	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
3	BWL7504	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
4	BWL7506	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
5	BWL7507	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
6	BWL7508	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
7	BWL7509	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
8	BWL7510	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
9	BWL7511	+	PBW550//*Yr15*/6*Avocet/3/2*PBW550/4/GLUPRO/3*PBW568//3*PBW550
10	PBW761	–	PBW550//*Yr15*/6*Avocet/3/2*PBW550
11	PBW725	–	PBW621//GLUPRO/3*PBW568/3/PBW621
12	HD3086	–	DBW14/HD 2733//HUW 468

+ confers the presence of Gpc-B1 gene, while – confers the absence of Gpc-B1 gene.

from the ratio between the concentration and the volume (Saputra et al., 2022). Both PN and SA were sprayed on leaves at an equal rate of 100 cc per m^2 to every genotype/plot (540 cc per plot/genotype) using a backpack CO_2-based pressure sprayer. For the control treatments, PN and SA were replaced with distilled water.

In both the experiments, the environmental conditions (including rainfall, temperature, and humidity) differed significantly (Supplementary Table 1). For each chemical, the experiment was conducted in three replications using a randomized block design (RBD). Each genotype was grown at the standard rate of 100 kg seed per hectare recommended by PAU, in a separate plot (plot size of 5.4 m^2, six rows of 4.5 m long with a row-to-row distance of 20 cm). In case of the PN and SA experiments, the trials were sown on November 22 and November 26, 2020, respectively. During the growing season, the standard field management practices were applied according to PAU recommendations (online available at https://www.pau.edu/content/ccil/pf/pp_rabi.pdf) for raising the crops. For better crop growth, a PAU-recommended package of NPK fertilizers (N = 50 kg/acre, P = 25 kg P_2O_5/acre, and K = 12 kg K_2O/acre) was applied. Furthermore, manual weeding and chemical weed control were implemented before and during the growing season inside and across the surrounding border areas of the fields. In the experimental fields, standard methods to control insect-pests and diseases were also used when required.

Data recorded on different traits

Our study mainly aimed to investigate the effects of PN application on a range of agronomic and physiological traits. Additionally, a preliminary experiment was conducted to assess the effects of SA application on some selected agronomic traits. In the experiments conducted for the evaluation of the effect of foliar applications of PN and SA, data on number of days to flowering (DTF), and number of days to maturity (DTM) was recorded when 75% of the plants in a plot were in that stage. Furthermore, a set of five plants was randomly selected in each plot for recording the data on following traits- plant height (PH), number of spikelets per plant (SPS), number of grains per spike (GPS), 1000-grain weight (TGW), and grain protein content (GPC) as described in one of our earlier studies (Tanin et al., 2022b). The yield per plot (YPP) was recorded by weighing the total wheat grains harvested from each plot. In addition to the traits listed above, the data on following physiological traits was also recorded during the experimental trial conducted to evaluate the effect of PN application-

(i) Chlorophyll content in flag leaf

A total of 10 plants were tagged in each plot. Consequently, the chlorophyll content was measured on the flag leaf of each individual plant using a hand-held chlorophyll concentration meter (MC-100, Utah 84321, USA). The data for each genotype was recorded at booting stage (Chl. BS), anthesis stage (Chl. AS), 10, 20, 25, and 30 days after anthesis (Chl. 10DAA, Chl. 20DAA, Chl. 25DAA, and Chl. 30DAA). In each replication, the value of each genotype consisted of the average of 10 flag leaf samples (3 readings per flag leaf). The experimental error was decreased as a result of the large number of samples that were collected for the determination of the chlorophyll content. The chlorophyll content reading was recorded as chlorophyll content index (CCI).

(ii) Senescence in peduncle

To measure the senescence in wheat peduncles, 10 peduncles were randomly tagged from each plot, and the percentage of peduncles completely turned yellow at 30 DAA (YP 30DAA) was determined by counting the number of yellow peduncles from these randomly selected peduncles. The readings for the yellow peduncles were recorded in percentage (%).

(iii) Flag leaf nitrogen content

For measuring the leaf nitrogen content, five wheat plants were randomly tagged in each plot, and subsequently flag leaf samples from each tagged plant were collected (five samples per plot), packed in envelops, and dried using an oven. The sampling dates for flag leaf nitrogen content analysis were similar to those for chlorophyll content measurement viz., flag leaf nitrogen content analysis at booting stage (N BS), anthesis stage (N AS), 10, 20, 25, and 30 days after anthesis (N 10DAA, N 20DAA, N 25DAA, and N 30DAA, respectively). Later on, the flag leaf samples were subjected to nitrogen content analysis following the Kjeldahl's distillation method as described by (Piper, 2019). About 0.5 g of dried grinded leaf samples were mixed with a pinch of digestion mixture of potassium sulphate (K_2SO_4), copper sulphate ($CuSO_4$), selenium powder, and mercury oxide for this purpose, and then digested in 10 ml of concentrated sulphuric acid (H_2SO_4). All of the chemical mixture was placed in the digestion tube and left overnight before being heated in Kjeldahl's digestion unit from 75^0C to 375^0C for 2.5 hour/each sample during the next day to digest the colourless content. The digestion mixture was then used for distillation by filling a volumetric flask to 50 ml with distilled water. Following that, the final titration was performed, and the nitrogen content (per leaf sample) was calculated using the following equation given by Islam et al. (2003):

$$\mathrm{N}(\%) = \frac{14 \times Normality\ of\ Acid \times Titration\ Volume - Blank\ Value}{Sample\ Weight \times 10}$$

(iv) Measurements of grain nitrogen content and straw nitrogen content

Grain nitrogen content (GNC; %) and straw nitrogen content (SNC; %) were determined following the Kjeldahl's distillation method as discussed above for the estimation of nitrogen content in flag leaves. In this method, half a gram of completely dried and grinded grain and straw samples were mixed with a small amount of digestion mixture and later digested in 10 ml of H_2SO_4 for 150 minutes per sample. The digestion tubes containing the chemical mixture were kept overnight at room temperature and then heated in the digestion unit for 150 minutes at a temperature of up to 375°C to digest the content. The mixture was later subjected to a distillation process following the final titration, and the nitrogen

content (per grain sample and straw sample) was calculated using the above-mentioned equation.

Statistical analysis

The experiments were based on randomized block designs. The pooled analysis of variance (ANOVA) was performed for the data collected from each experiment, using the agricolae package of RStudio software version 4.0.3. The least significant difference (LSD) test was utilized in the context of the ANOVA when the F-ratio suggested the rejection of the null hypothesis H_0, that is, when the difference between the treatments and genotypes means were significant. This test was useful in identifying treatments with statistically different means.

Results

Experiments conducted for the evaluation of the effect of foliar application of SA

Based on the data recorded in experimental trial conducted to evaluate the effect of SA, a combined ANOVA was carried out to determine the effects of the genotypes and the treatments that were tested in this study. The combined ANOVA involving all the eight traits is provided in Supplementary Table 2. The variation due to genotype (G) was highly significant for all the eight traits at a 0.001% level of significance, while the variation due to SA treatment (T) was found to be significant at different levels of significance, for instance, DTM (0.001%), GPS (0.001%), TGW (0.01%), and YPP (0.001%). Overall, in this experiment, a significant high level of variation was observed among the wheat genotypes in terms of GPC and yield-related traits.

Effect of SA on yield and related traits in advanced wheat genotypes

The data provided in Table 2 represents the significant level of variations due to the SA treatment among the genotypes. Foliar application of SA significantly increased the DTM by an average of 2.31 days as compared to the control. Among the genotypes, HD3086 recorded the highest DTM (129.33 days), whereas the lowest DTM was observed in BWL7511 (122 days). The genotype, PBW725, recorded the highest increase in DTM (3.33 days) with the SA spray over the control. The increase in grain filling period due to SA application was higher in the genotypes without *Gpc-B1* as compared with the genotypes with *Gpc-B1* gene. The GPS was significantly increased by an average of 3.17 grains due to foliar spray of SA in all the genotypes under study. PBW725 had the most GPS (63 grains), which was significantly higher than all other genotypes tested during the experiment. BWL7507 had the lowest GPS (45 grains), which was significantly lower than all other genotypes. The greatest increase in GPS was observed in PBW725 (6 grains) with SA application when compared to the control. When

TABLE 2 Effects of SA treatment on yield component compared to the control.

	GPS		DTM		TGW		YPP	
Genotypes	SA	Control	SA	Control	SA	Control	SA	Control
BWL6964	48.67 ± 0.58	47.33 ± 0.58	122.67 ± 0.58	120.33 ± 0.58	28.05 ± 0.05	27.5 ± 0.50	2.19 ± 0.03	2.16 ± 0
BWL7503	51 ± 1	47.67 ± 0.58	124.33 ± 0.58	122.33 ± 0.58	29.65 ± 1.55	29.43 ± 1.59	2.6 ± 0.01	2.47 ± 0.01
BWL7504	51 ± 1	49 ± 0	123.33 ± 0.58	121.33 ± 0.58	30 ± 1	29.33 ± 1.53	3.03 ± 0.06	2.84 ± 0.08
BWL7506	60 ± 1	55 ± 0	123.67 ± 0.58	121.33 ± 0.58	33.7 ± 3.10	31.93 ± 3.27	3.51 ± 0.03	3.24 ± 0
BWL7507	45 ± 1	44.67 ± 0.58	123 ± 0	121.33 ± 0.58	32.35 ± 1.65	31.57 ± 1.25	2.03 ± 0.03	2.01 ± 0.02
BWL7508	59 ± 1	54.67 ± 0.58	125 ± 0	122.33 ± 0.58	32.85 ± 0.35	31.25 ± 0.72	3.41 ± 0.01	3.2 ± 0.03
BWL7509	48 ± 1	45.67 ± 0.58	124 ± 0	122.33 ± 0.58	33.5 ± 0.40	33.17 ± 0.67	2.26 ± 0.02	2.24 ± 0.03
BWL7510	54.67 ± 0.58	52 ± 0	123.33 ± 0.58	121.33 ± 0.58	32.25 ± 0.95	31.93 ± 1.01	2.79 ± 0.03	2.72 ± 0.03
BWL7511	62 ± 1	58.33 ± 0.58	122 ± 1	120.33 ± 0.58	33.3 ± 1.30	31.7 ± 0.85	4.08 ± 0.03	3.78 ± 0
PBW761	56 ± 1	54.33 ± 0.58	124.33 ± 0.58	121.33 ± 0.58	39 ± 1	37.5 ± 1.32	3.12 ± 0.09	2.79 ± 0.08
PBW725	63 ± 1	57 ± 0	129 ± 1	125.67 ± 0.58	40 ± 0	37.67 ± 0.58	4.42 ± 0.02	3.8 ± 0.03
HD3086	62 ± 1	56.67 ± 0.58	129.33 ± 0.58	126.33 ± 0.58	36.45 ± 1.05	34.67 ± 0.58	4.29 ± 0.02	3.85 ± 0.01
Means	55.03 ± 6.22	51.86 ± 4.78	124.5 ± 2.32	122.19 ± 1.91	33.43 ± 3.58	32.30 ± 3.08	3.14 ± 0.82	2.93 ± 0.65
LSD 5%								
G	0.82		0.68		1.57		0.04	
T	0.34		0.28		0.64		0.01	

DTM, number of days to maturity; GPS, number of grains per spike; TGW, 1000-grain weight; YPP, yield per plot; SA, salicylic acid; T, treatment, and G, genotype.

compared to the control, the lowest GPS improvement was observed in BWL6964 (1.34 grains) with SA spray.

Furthermore, in the current experiment, all of the tested genotypes conferred a highly significant improvement in GPS with SA spray over their respective control(s). The SA spray on the leaves increased the TGW and YPP significantly, with an average of 1.13 g and 0.21 kg per plot, respectively. PBW725 had the highest TGW (40 g) and yield (4.42 kg per plot), which was significantly higher than all other genotypes for both the traits. PBW725 had the highest levels of enhancement for TGW (2.23 g) and yield (0.62 kg per plot). Similarly, the smallest increases in TGW and YPP were found in BWL7510 (0.32 g), BWL7507 (0.02 kg per plot), and BWL7509 (0.02 kg per plot). Overall, in the light of the results obtained, the foliar application of SA was reported to be extremely successful in delaying the senescence and extending the grain filling period to reduce the yield penalty associated with the *Gpc-B1* gene, and furthermore, the improvement in yield and yield components due to SA spray was higher in the genotypes without *Gpc-B1* as compared to genotypes with the *Gpc-B1* gene. This could be due to the comparatively longer grain filling period with SA application observed in the genotypes which did not carry the *Gpc-B1* gene (Figure 1). Furthermore, in the current study, SA foliar spray had no significant effect on DTF, PH, SPS, or GPC.

Experiment conducted for the evaluation of the effect of foliar application of PN

Based on the data recorded in experimental trial conducted at Ludhiana, a pooled ANOVA was carried out to determine the effects of the genotypes and the treatments that were tested in this study. The pooled ANOVA concerning all the twenty-three traits (which is recorded on different dates) is provided in Supplementary Table 3. The variation due to genotype (G) was highly significant for all the traits either at a 0.001% or 0.01% level of significance, while in the case of treatment (T), the variation was highly significant for DTM, GPS, TGW, YPP, Chl. AS, Chl. 10DAA, N BS, N AS, N 10DAA, YP 30DAA, and SNC at 0.001% level of significance. Furthermore, the G x T interaction was also observed as significant for DTF, PH, SPS, GPS, Yield, Chl. BS, Chl. 25DAA, N BS, N AS, N 10DAA, YP 30DAA, and SNC. In the present study, a significant level of differences was observed for all the physiological, and yield-related traits.

Effect of PN on yield and related traits in advanced wheat genotypes

Table 3 represents the differences in the values of yield and its components achieved after foliar application of PN. The foliar spray of PN significantly increased the GPS by an average of 6.42 grains in all tested genotypes over their corresponding control. Among the genotypes, BWL7502 recorded the highest number of GPS (61.3 grains), which was significantly higher than all other genotypes, whereas, lowest number of GPS (48.3 grains) was recorded by HD3086. The genotype BWL7502 showed the highest improvement in GPS (14.5 grains) with PN spray as compared to its respective control. Similarly, all the genotypes showed significantly higher GPS with PN spray as compared to the control. The DTM was significantly increased with PN application by an average of 1.03 days over the control in all the genotypes considered in the present study. PBW725 showed the highest DTM (133.67 days), which was significantly higher than all other tested genotypes. The BWL6964 and BWL7502 recorded the lowest DTM (125 days), which was significantly lower than all other genotypes. Among the genotypes, BWL6964 exhibited the highest improvement in DTM (2 days) due to foliar application of PN over the control.

Similarly, the TGW was significantly improved with the foliar application of PN by an average enhancement of 1.97 g in all the genotypes. The genotype PBW725 conferred the highest TGW

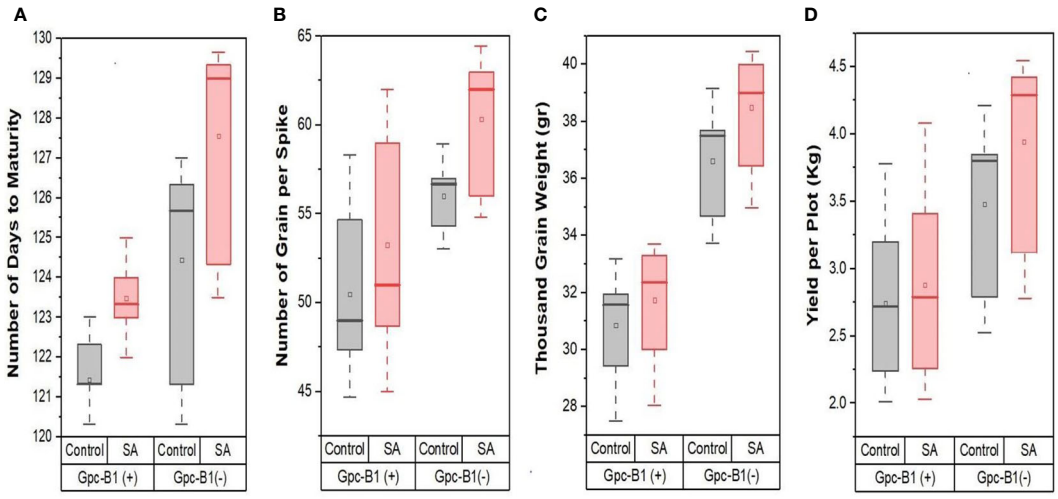


FIGURE 1
The effects of SA treatment on key agronomic traits such as **(A)** number of days to maturity, **(B)** number of grain per spike, **(C)** thousand grain weight, and **(D)** yield per plot of wheat genotypes with and without the *Gpc-B1* gene.

TABLE 3 Comparison of the effects of PN treatment on yield components compared to control.

	GPS		DTM		TGW		YPP	
Genotypes	PN	Control	PN	Control	PN	Control	PN	Control
BWL6964	57.4 ± 0.4	49.6 ± 1.91	125 ± 1	123 ± 1	32.68 ± 0.79	30.56 ± 0.5	3.4 ± 0.05	3.35 ± 0.08
BWL7502	61.3 ± 1.3	46.8 ± 2	125 ± 1	123.5 ± 0.5	32.67 ± 1.89	31.56 ± 0.33	3.57 ± 0.14	3.5 ± 0.1
BWL7504	54 ± 2.56	50.8 ± 2.17	126 ± 0	124.5 ± 0.5	29.94 ± 0.25	28.42 ± 0.5	3.10 ± 0.06	3 ± 0.09
BWL7506	50.33 ± 2.66	45 ± 2.23	125.33 ± 0.58	124.5 ± 0.5	36.36 ± 0.58	34.84 ± 0.5	3.34 ± 0.15	3.1 ± 0.05
BWL7507	52 ± 2.6	43.6 ± 1.83	126 ± 0	125 ± 0	34.88 ± 1.3	33.02 ± 0.48	3.12 ± 0.07	3 ± 0.08
BWL7508	53.3 ± 1.47	50.4 ± 1.91	125.67 ± 0.58	124.5 ± 0.5	32.42 ± 0.63	30.87 ± 0.5	3.16 ± 0.08	3 ± 0.09
BWL7509	54.9 ± 2.52	50 ± 2.17	125.33 ± 0.58	124.5 ± 0.5	32.07 ± 0.9	30.28 ± 0.5	3 ± 0.33	2.05 ± 0.09
BWL7510	52.3 ± 0.5	46 ± 2.23	125.33 ± 0.58	124.5 ± 0.5	33.25 ± 0.9	31.17 ± 0.17	3.74 ± 0.04	3.55 ± 0.08
BWL7511	53.53 ± 2.25	48 ± 1.8	125.33 ± 0.58	124.5 ± 0.5	33.98 ± 1.75	31.54 ± 0.5	3.14 ± 0.11	3 ± 0.05
PBWL761	54.2 ± 2.51	50 ± 2.23	125.67 ± 0.58	125.5 ± 0.5	35.68 ± 1.3	33.8 ± 0.5	3.46 ± 0.06	3.4 ± 0.02
PBW725	54.3 ± 3.7	45.4 ± 1.87	133.67 ± 0.58	132.5 ± 0.5	39.44 ± 1.51	34.31 ± 0.44	4.13 ± 0.05	4.05 ± 0.1
HD3086	48.3 ± 2.5	43.2 ± 1.91	131 ± 1	130.5 ± 0.5	35.09 ± 0.46	34.49 ± 0.5	2.19 ± 0.09	2 ± 0.08
Means	53.82 ± 3.29	47.4 ± 2.75	126.61 ± 2.75	125.58 ± 2.87	34.04 ± 2.46	32.07 ± 2	3.28 ± 0.47	3.08 ± 0.59
LSD 5%								
G	2.46		0.69					
T	1		0.28		0.82		0.14	
					0.34		0.06	

DTM, number of days to maturity; GPS, number of grain per spike; TGW, 1000-grain weight; YPP, yield per plot; PN, potassium nitrate; T, treatment, and G, genotype.

(39.44 g), which was significantly higher than all other genotypes evaluated in this study. The lowest TGW was observed in BWL7504 (29.94 g), which was significantly lower than other genotypes. PBW725 also showed the highest enhancement in TGW (5.13 g) with PN application as compared to the control. The PN application on leaves significantly increased the total YPP by an average of 0.2 kg per plot in all the genotypes. The highest YPP (4.13 kg per plot) was recorded by PBW725, which was significantly higher than all other studied genotypes. The lowest YPP (2.19 kg per plot) was observed in HD3086, which was significantly lower as compared to other genotypes. The highest YPP enhancement (0.95 kg per plot) was recorded by BWL7509 with the PN spray over the control. Overall, the enhancement in yield and yield components due to PN foliar application was greater in the genotypes without *Gpc-B1* over the genotypes with the *Gpc-B1* gene (Figure 2). This could be due to the comparatively longer grain filling period with SA application observed in the genotypes which did not carry the *Gpc-B1* gene. Compared to the control, the foliar spray of PN had no significant effect on DTF, PH, SPS, and GPC traits.

Effect of PN on flag leaf chlorophyll content in wheat NILs and cultivars

The Table 4 represents the differences in flag leaf chlorophyll contents achieved after foliar spray of PN. As clear from the Table 4, the foliar spray of PN significantly increased the flag leaf chlorophyll content by an average of 2.35 CCI at AS and 1.96 CCI at 10DAA as compared to the control. Among the genotypes, BWL7510 with 31.93 CCI at AS and BWL7507 with 28.4 CCI at 10DAA showed the highest flag leaf chlorophyll contents, that were significantly higher than all other genotypes evaluated in the current study. The HD3086 recorded the lowest flag leaf chlorophyll content with 25.65 CCI at AS and 23.5 CCI at 10DAA) as compared to other genotypes. The genotypes, BWL7502 with 5.55 CCI at AS, and HD3086 with 4.30 CCI at 10DAA were observed to have the highest improvements in flag leaf chlorophyll contents with the PN application over the control. Furthermore, the foliar application of PN had no significant effects on the flag leaf chlorophyll contents at the BS, 20DAA, 25DAA, and 30DAA, compared to the control. Furthermore, chlorophyll content enhancement due to PN application in *Gpc-B1* positive wheat genotypes was much higher than in *Gpc-B1* negative wheat genotypes (Figures 3A, B).

Effect of PN on flag leaf nitrogen content, straw nitrogen content and peduncle senescence in wheat NILs and cultivars

The Table 4 contains the data revealing the differences for nitrogen contents in flag leaf and straw, and senescence in peduncles among the genotypes achieved after treating with PN. The flag leaf nitrogen content was significantly increased with the

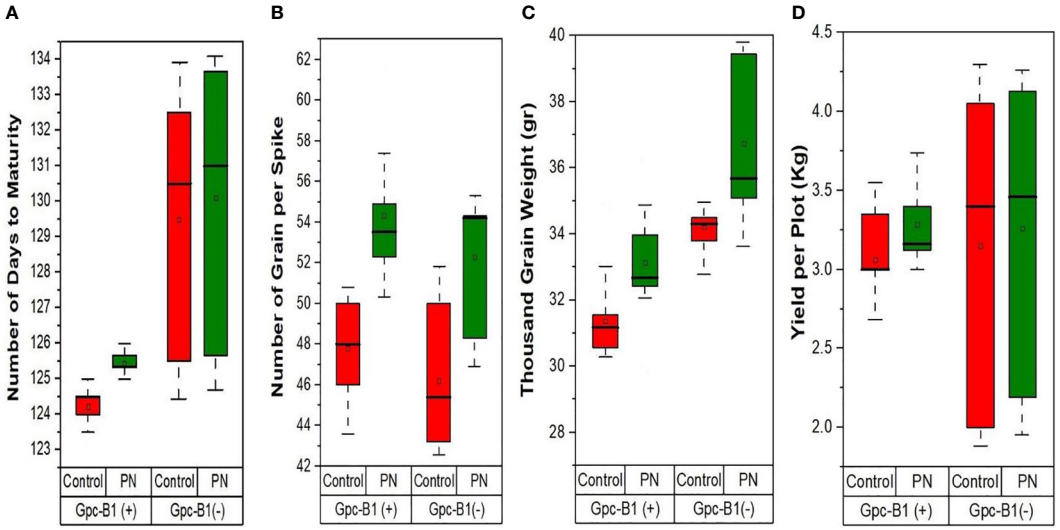


FIGURE 2
The effects of PN treatment on key agronomic traits such as **(A)** number of days to maturity, **(B)** number of grain per spike, **(C)** thousand grain weight, and **(D)** yield per plot of wheat genotypes with and without the *Gpc-B1* gene.

foliar spray of PN by an average of 0.52% at BS, and 0.35% at both AS and 10DAA in the evaluated genotypes. The highest flag leaf nitrogen contents were observed in BWL7507 (3.9% at BS), PBW761 (4.3% at AS), and BWL7506 (3.45% at 10DAA), which were significantly higher than all other genotypes in each developmental stage. The lowest flag leaf nitrogen content was recorded by BWL7509 with 2.54% at BS, followed by HD3086 with 3.12% at AS, and BWL7510 with 2.43% at 10DAA which were significantly lower as compared to other genotypes at each developmental stage. The highest flag leaf nitrogen enhancements with PN was observed in BWL7509 with 1.38% at BS, followed by BWL6964 with 1.16% at AS, and BWL7511 with 0.94% at 10DAA after the PN spray as compared to the control. As a result, the flag leaf nitrogen content was higher in genotypes with no *Gpc-B1* gene compared to the genotypes carrying the *Gpc-B1* gene at BS and AS (Figures 3C, D).

The foliar application of PN significantly increased the SNC by an average of 0.17% over the control. BWL7504 was observed to have the highest SNC with 0.79%, which was significantly higher than all other genotypes. The lowest SNC was recorded by BWL7506 with 0.29%, which was statistically similar to BWL7507, BWL7509, BWL7510, BWL7511, and PBW761 but significantly lower than the remaining genotypes. The genotype BWL6964 recorded the highest enhancement in SNC (0.46%) with PN over the control. The YP 30DAA was significantly increased by an average of 19.08% due to foliar application of PN in all the genotypes as compared to the control. The highest YP 30DAA was observed in BWL7504 (91%) which was statistically similar to BWL6964, BWL7502, BWL7506, and HD3086 but significantly higher than all other genotypes. The lowest YP 30DAA was recorded by PBW761 (18.33%) which was significantly lower than all other genotypes. The highest enhancement of YP 30DAA with PN was recorded by BWL7506 (61.67%) as compared to the control. Overall, the rate of senescence in peduncle was much higher in *Gpc-B1* negative genotypes than in *Gpc-B1* positive wheat genotypes (Figure 3E). However, the foliar spray of PN had no significant effects on N 20DAA, N 25DAA, N 30DAA, and GNC traits.

Discussion

In wheat, the GPC is considered the most important factor affecting nutritional quality. Grain protein, therefore, has significant potential to increase economic value of wheat grains (Gudi et al., 2022a). *Gpc-B1* is the major regulator of GPC in wheat, which is responsible for more than 60% of phenotypic variation for the protein content in grains (Uauy et al., 2006; Saini et al., 2020). It is commonly known that higher GPC is negatively associated with grain yield in wheat, thus hampering the breeding efforts aimed at simultaneous improvement of GPC and yield in wheat genotypes (Tanin et al., 2022b). In addition to significant effects on GPC, the *Gpc-B1* is associated with the earlier initiation of senescence, shorter grain filling period, higher percentage of yellow peduncle, and more efficient nitrogen remobilization in wheat (Uauy et al., 2006). The SA and PN are known to play significant roles in delaying the leaf senescence and improving tolerance against biotic and abiotic stresses (El-Tayeb, 2005; Ibrahim et al., 2014; Jatana et al., 2021; Gudi et al., 2022b; Tanin et al., 2022a).

Application of SA on leaves significantly affects the wheat physiology, including seed germination, normal growth, yield production, nutrient uptake and remobilization, photosynthesis efficiency, stomatal density and transpiration rate, root development, and biomass enhancement (Azmat et al., 2020; Dias et al., 2021; Wang et al., 2021);. Under stress conditions, SA regulates crucial plant physiological processes, including photosynthesis, proline metabolism, nitrogen metabolism,

10.3389/fpls.2023.1107705

TABLE 4 Comparison of the effects of PN treatment on physiological and other key traits.

	Chl. AS		Chl. 10DAA		N BS		N AS	
Genotypes	PN	Control	PN	Control	PN	Control	PN	Control
BWL6964	30.2 ± 0.6	28.4 ± 1.95	27.45 ± 0.65	24.2 ± 1.65	2.95 ± 0.29	2.08 ± 0.03	3.93 ± 0	2.77 ± 0.08
BWL7502	29.55 ± 0.05	24 ± 2	25.4 ± 0.20	23.3 ± 2	2.77 ± 0	2.54 ± 0.10	3.35 ± 0.35	2.77 ± 0.1
BWL7504	31.4 ± 0.1	29.2 ± 2.08	27.40 ± 0.53	26.6 ± 1.4	2.95 ± 0.06	2.77 ± 0.08	3.35 ± 0.12	3.01 ± 0.09
BWL7506	31.7 ± 0.46	29.8 ± 1.95	28.37 ± 0.12	26.6 ± 2	3.27 ± 0.14	3.01 ± 0.08	3.85 ± 0.05	3.7 ± 0.07
BWL7507	30.53 ± 0.21	28 ± 2.23	28.4 ± 0.95	26.8 ± 1.2	3.90 ± 0.05	3.7 ± 0.05	3.88 ± 0.08	3.7 ± 0.07
BWL7508	28.7 ± 0.56	26.5 ± 1.5	26.87 ± 0.57	24.1 ± 1.9	3.35 ± 0.13	3.01 ± 0.09	3.86 ± 0.10	3.7 ± 0.05
BWL7509	26.2 ± 2	24.9 ± 1.1	23.2 ± 3.9	21.6 ± 1.4	2.54 ± 0.46	1.16 ± 0.04	3.36 ± 0.35	3.01 ± 0.07
BWL7510	31.93 ± 0.67	29.4 ± 1.6	24.85 ± 0.35	22.8 ± 1.2	2.72 ± 0.17	2.31 ± 0.09	3.47 ± 0.21	3.24 ± 0.07
BWL7511	26.95 ± 0.15	26.5 ± 1.5	25.1 ± 0.3	24.7 ± 1.3	2.6 ± 0.40	1.39 ± 0.06	3.57 ± 0.37	3.24 ± 0.06
PBWL761	28.2 ± 1.18	26.7 ± 1.7	24.97 ± 0.59	24.3 ± 1.7	3.60 ± 0.10	3.24 ± 0.06	4.30 ± 0.12	4.16 ± 0.04
PBW725	29.8 ± 5.3	24.6 ± 1.5	26.33 ± 0.5	24.2 ± 1.8	2.99 ± 0.01	2.54 ± 0.07	3.87 ± 0.03	3.7 ± 0.05
HD3086	25.65 ± 0.45	24.6 ± 2.14	23.5 ± 0.4	19.2 ± 1.8	2.89 ± 0.12	2.54 ± 0.06	3.12 ± 0.12	2.77 ± 0.08
Means	29.23 ± 2.13	26.88 ± 2.06	25.99 ± 1.76	24.03 ± 2.19	3.04 ± 0.41	2.52 ± 0.73	3.66 ± 0.34	3.31 ± 0.47
LSD 5%								
G	1.93		1.55		0.18		0.17	
T	0.79		0.63		0.07		0.07	
	N 10DAA		SNC		YP 30DAA			
Genotypes	PN	Control	PN	Control	PN	Control		
BWL6964	2.70 ± 0.08	2.54 ± 0.06	0.69 ± 0.23	0.23 ± 0.01	90 ± 10	40 ± 20		
BWL7502	2.89 ± 0.35	2.08 ± 0.10	0.46 ± 0	0.23 ± 0.05	86.67 ± 11.55	80 ± 10		
BWL7504	2.77 ± 0.23	2.54 ± 0.06	0.79 ± 0.06	0.69 ± 0.01	91 ± 3.61	86.67 ± 5.77		
BWL7506	3.45 ± 0.08	3.24 ± 0.03	0.29 ± 0.03	0.23 ± 0.02	85 ± 15	23.33 ± 15.28		
BWL7507	2.94 ± 0.09	2.77 ± 0.06	0.35 ± 0.12	0.23 ± 0.01	60 ± 10	30 ± 20		
BWL7508	2.91 ± 0.06	2.77 ± 0.06	0.58 ± 0.12	0.23 ± 0.01	45 ± 15	30 ± 20		
BWL7509	2.78 ± 0.46	2.31 ± 0.07	0.32 ± 0.06	0.23 ± 0.02	75 ± 25	70 ± 10		
BWL7510	2.43 ± 0.11	2.31 ± 0.06	0.35 ± 0.12	0.23 ± 0.01	25 ± 15	23.33 ± 15.28		
BWL7511	2.79 ± 0.01	1.85 ± 0.05	0.35 ± 0.12	0.23 ± 0.01	40 ± 30	30 ± 20		
PBW761	2.54 ± 0.69	2.31 ± 0.06	0.35 ± 0.12	0.23 ± 0.02	18.33 ± 7.64	10 ± 10		
PBW725	3 ± 0.23	2.54 ± 0.06	0.37 ± 0.09	0.23 ± 0.01	45 ± 35	10 ± 10		
HD3086	2.52 ± 0.03	2.31 ± 0.02	0.36 ± 0.02	0.23 ± 0.02	88 ± 2.65	80 ± 10		
Means	2.81 ± 0.27	2.46 ± 0.36	0.44 ± 0.16	0.27 ± 0.13	61.86 ± 26.4	42.78 ± 28.35		
LSD 5%								
G	0.24		0.09		19.19			
T	0.1		0.04		7.83			

Chl. AS, chlorophyll content at anthesis stage; Chl. 10DAA, chlorophyll content at 10 days after anthesis; N BS, nitrogen content at booting stage; N AS, nitrogen content at anthesis stage; N 10DAA, nitrogen content at 10 days after anthesis; YP 30DAA, yellow peduncle at 30 days after anthesis; SNC, straw nitrogen content, PN, potassium nitrate; T, treatment, and G, genotype.

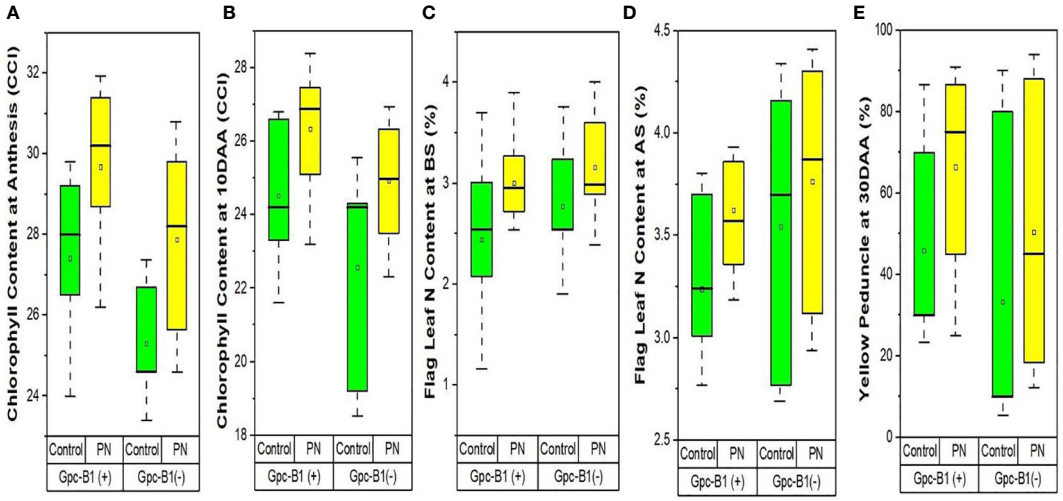


FIGURE 3
The effects of PN treatment on traits such as **(A)** chlorophyll content at anthesis, **(B)** chlorophyll content at 10DAA, **(C)** flag leaf N content at booting stage, **(D)** flag leaf N content at anthesis stage, and **(E)** yellow peduncle at 30DAA of wheat genotypes with and without the *Gpc-B1* gene.

glycine-betaine production, antioxidant defence system (generation of reactive oxygen species), and plant-water relations, and thus protects plants from abiotic stresses (reviewed in Khan et al., 2015). Furthermore, the SA has been shown to play an important role in the regulation of leaf senescence via its effects on lipid metabolism (Xiao et al., 2010), autophagy (Yoshimoto et al., 2009), and the generation of reactive oxygen species (Guo et al., 2017). Apart from influencing senescence progression, SA influences senescence initiation by simultaneously inducing both positive and negative senescence regulators (Besseau et al., 2012). Grain yield enhancement was observed in wheat due to foliar spray of SA (Safar-Noori et al., 2018). Similarly, the PN regulates different biochemical and physiological processes that influence plant growth and metabolism, including seed germination and emergence, stomatal regulation, phloem transport, cation-anion balance, photosynthesis, energy transfer and contributes to the survival of plants exposed to various abiotic stresses (Wang et al., 2013). In this study, nine NILs carrying the *Gpc-B1* gene and three elite wheat varieties with no *Gpc-B1* gene were used to test the effects of foliar application of SA and PN on different yield-related traits, nitrogen concentrations in different tissues and some key physiological parameters. Previously, different concentrations of SA and PN were used to delay senescence and extend the grain filling period in cereal crops, including wheat, in order to increase grain yield (Singh et al., 2019; Singh and Singh, 2020; Jatana et al., 2021; Singh et al., 2021; Abdi et al., 2022).

The foliar spray of SA significantly increased the number of DTM in the current study, extending the grain filling period in all the evaluated genotypes that resulted into the increased grain yield. These results were consistent with studies earlier conducted in wheat (Singh et al., 2019; Jatana et al., 2021). It is observed that the SA application supports and further improves the photosynthetic mechanism of the wheat crop (Zheng et al., 2013). In addition, Fan et al. (2022) reported that foliar spray of SA significantly abated the negative effect of high temperature on flag leaf area in wheat and therefore the potential of flag leaf for detecting the light energy remain stable, thus the senescence is delayed in wheat flag leaves. Further, the carotenoids and chlorophyll are the important pigments produced during photosynthesis in plants, and chlorophyll is the major regulator of light energy-based physiological processes. The exogenous application of SA preserves the normal production of chlorophyll content in flag leaves under heat stress and further improves the photosynthetic efficiency and chlorophyll production in wheat flag leaves as a result of involvement of SA in preserving the structure of the chloroplast membrane and maintaining the optimum balance of the antioxidant mechanism in plant chloroplasts (Fan et al., 2022).

Furthermore, the foliar application of SA significantly enhanced the Fv/Fm of flag leaves in wheat, and therefore, due to this improvement, the photosynthetic potential and efficiency of light energy utilization were significantly increased, paving the way for better storage of the photosynthetic outcomes and delaying senescence, thus keeping the photosynthetic stability normal (Aldesuquy et al., 2018; Fan et al., 2022). Gonzalez-Bayon et al. (2019) observed that reduction of SA level in an *Arabidopsis thaliana* accession through introduction of a bacterial gene (i.e., *NahG*) resulted in plant size enhancement after 3 weeks of growth. The *TL1 BINDING TRANSCRIPTION FACTOR1* is an important gene regulating the SA-induced balance among plant's defense and growth pathways (Smith, 2019), and its expression levels were significantly suppressed in *Arabidopsis* F1s (derived from C24 x Ler) and C24 *NahG* individuals, which may be the reason for growth enhancement. Similarly, the down-regulation of several genes (e.g., *WRKY53*, *ORESARA1*, and *NAC-CONTAINING PROTEIN 29*) associated with SA-induced chlorophyll degradation (senescence) in these individuals could be the regulator of the delayed senescence, thereby causing a longer period of photosynthesis in leaves.

In wheat, the *NAC-S* gene has been reported to confer positive interaction with GPC and delayed senescence with no yield penalty (Sultana et al., 2021). Furthermore, SA spray on wheat leaves also significantly improved the yield components, including GPS and TGW, in all the twelve genotypes tested in the present study, which is in agreement to the previous studies conducted in wheat (Karim and Khursheed, 2011; Dias et al., 2021). The enhancement in yield components with SA application on leaves could be attributed to changes in wheat physiological mechanisms such as influencing the antioxidant defense system (Hasanuzzaman et al., 2019), ion transportation (Simaei et al., 2012), flowering time (Wada and Takeno, 2010), and photosynthesis efficiency (Fariduddin et al., 2003). Further, earlier studies on maize (Ahmad et al., 2018), rice (Jini and Joseph, 2017), barley (Pirasteh-Anosheh et al., 2022), and *Brassica* (Muhal and Solanki, 2016) supports our findings regarding improvements in yield component traits as a result of SA application. The grain yield enhancement with SA spray may be due to higher chlorophyll content, TGW, GPS, and longer grain filling period. The wheat genotype "Prodi" also showed higher grain yield with SA application as compared to the control (Ziasmin et al., 2017). In a more recent study, SA (1.5 mM; once every two days) and NaCl (100 mM, 200 mM; once a week) were sprayed on leaves starting from tillering stage until the complete spike formation stage, resulting a significant increase in shoot weight, root weight, GPC, yield and its components in two wheat varieties (viz., PAN3476, and SST806) (Abdi et al., 2022). In this study, the SA in combination with 100 mM NaCl showed greater enhancement than the SA in combination with 200 mM NaCl.

The foliar application of PN significantly increased the yield and its components, flag leaf chlorophyll content at anthesis and 10DAA, flag leaf nitrogen content at booting stage, anthesis and 10DAA, SNC, and yellow peduncle colour at 30DAA. The number of DTM was significantly increased by an average of more than one day with PN spray over the control. A similar increase in DTM with PN spray was also observed in earlier studies in wheat (Singh et al., 2013; Devi et al., 2017; Agrawal et al., 2021). Therefore, senescence is delayed and the grain filling duration is extended, thus improving key yield-related traits. In wheat, yield is a highly complex trait regulated by a number of genes and is determined by two major attributes, viz., TGW and the GPS (Yadav et al., 2021; Saini et al., 2022a). The grain yield improvement due to foliar application of PN could be due to a longer grain filling period, higher TGW, and GPS as compared to the control in wheat, the GPS is measured by integrating both the number of spikelets per spike and the number of grains per spikelet. Wheat spikelets carry more than a single grain, unlike the spikelets of other cereal crops such as rice and barely (Zhou et al., 2021; Singh R. et al., 2022). Thus, wheat spikelet is one of the most important grain yield attributes.

The capacity for GPC in wheat is significantly affected by the spike architecture. Wheat is known to have an unbranched inflorescence. Furthermore, spikelets are placed along the primary axis of inflorescence (Koppolu and Schnurbusch, 2019). Since a number of florets are available on rachilla to establish the spikelet, the GPS is finally determined by the number of spikelets having the higher number of fertile florets (Li et al., 2021). The increase in the GPS with PN application may also be due to the longer length of spike and the higher number of grains per spikelet. A similar increase in GPS with the PN application was reported in wheat by Singh and Singh (2020). Additionally, higher grain yield along with the favourable physiological, phenological, and biochemical characteristics were observed with foliar spray of thio-urea at a rate of 500 ppm under drought and irrigated conditions in two wheat genotypes viz., K1006, and K307 (Yadav et al., 2019). Similarly, a number of studies have found that PN spray can significantly improve yield and its components in different crops, including rice (Ali et al., 2021), sorghum (Ávila et al., 2022), maize (Ahmad et al., 2022), and coriander (Elhindi et al., 2016).

Flag leaf chlorophyll content is known for its role as one of the most essential determinants in wheat growth and development (Zhang et al., 2016). The higher rate of chlorophyll degradation and accelerated senescence are correlated with the higher GPC and yield penalty in wheat genotypes (Kade et al., 2005). Foliar application of PN at booting and anthesis stages significantly increased the flag leaf chlorophyll content at anthesis and 10DAA. In addition, at both the stages, the chlorophyll content improvement with PN spray was comparatively higher in wheat genotypes with no *Gpc-B1* gene as compared to the genotypes carrying the *Gpc-B1* gene (Figures 3A, B). It is assumed that the lower improvement in flag leaf chlorophyll content with PN application on the leaves is associated with the *Gpc-B1* gene. Further, the improvement in chlorophyll content in genotypes with no *Gpc-B1* gene was comparatively higher during 10DAA as compared to the anthesis stage. However, the expression of *Gpc-B1* in leaves is initiated before anthesis (Uauy et al., 2006). Because the *Gpc-B1* gene starts to express after anthesis, genotypes carrying it may experience faster chlorophyll degradation at 10DAA (Uauy et al., 2006).

During the pre-anthesis period, the nitrogen is provisionally stored in flag leaf, and later on, after anthesis is initiated, it is remobilized from flag leaf to developing grains (Distelfeld et al., 2014). The efficient remobilization of nitrogen from flag leaf into grains is the key factor which significantly determines the protein accumulation in wheat grains. In this study, the application of PN significantly increased the flag leaf nitrogen content at booting, anthesis, and 10DAA. The lower nitrogen content and higher chlorophyll content in flag leaf at anthesis in genotypes carrying the *Gpc-B1* as compared to the genotypes with no *Gpc-B1* (Figures 3C, D) provides evidence that *Gpc-B1* gene becomes active in leaves before the grain formation. A similar result previously reported by Kade et al. (2005) support the findings of the present study. In addition, the initiation of *Gpc-B1* expression in leaves after grain formation further suggests a significant role of *Gpc-B1* on nitrogen accumulation in grain at source (leaves) over the sink (grain). The higher rate of senescence enhancement in peduncles could be a combined result of PN spray and high GPC gene in genotypes carrying the *Gpc-B1* gene compared to the genotypes with no *Gpc-B1* gene (Figure 3E). In addition, we observed that the pleiotropic effect of the *Gpc-B1* gene on various traits, such as peduncle senescence rate, grain size, and yield, was inconsistent following PN application.

To improve wheat performance in the field, physiological interventions (use of SA and PN) and genetics still need to be

combined. In the present study, to mitigate the yield penalty associated with the *Gpc-B1* gene in our germplasm, we sprayed SA and PN on leaves to improve the wheat potential (genetics) aimed to produce a higher grain yield with no negative effect on GPC. The results indicated that the foliar application of SA and PN was an extremely efficient approach to extend the grain filling period and enhance the yield components (TGW and GPS) in wheat lines introgressed with the Gp*c-B1* gene. The comparative result indicates that foliar application of SA was more effective to mitigate the yield penalty associated with the *Gpc-B1* gene as compared with the result obtained with the PN spray. The DTM and YPP were significantly improved by 2.31 days and 0.21 kg/plot, respectively, due to the foliar application of SA, while the PN spray increased the DTM and YPP by 1.03 days and 0.2 kg/plot, respectively. Furthermore, both the SA and PN did not show any significant influence on GPC. In addition, a rapid and more efficient photosynthesis was also observed due to the application of PN which subsequently resulted in higher chlorophyll production in flag leaves. Similarly, physiological intervention and genetics could be more effective to improve the yield potential of wheat when further supported by agronomic intervention (e.g., appropriate date of sowing) (Jatana et al., 2021).

Conclusions

The results of the present investigation provided evidence that the yield reduction due to the presence of the *Gpc-B1* gene could be compensated through the foliar application of SA and PN in wheat. As a result of the spraying of growth regulators on leaves, wheat growth and development were enhanced. The dose and time (stages) of spraying of growth regulators (PN: 2% at booting and anthesis stages; SA: 75 ppm at booting and anthesis stages) used in present experiment is strongly recommended to mitigate the early senescence caused by the *Gpc-B1* gene. This yield improvement in tested wheat genotypes was caused by growth regulator-mediated longer grain filling period. Furthermore, we showed that the application of SA and PN favorably modulated grain yield and other important traits. Future research might explore how these two chemicals interact with the expression and activity of the gene *Gpc-B1*. The preliminary experiment with SA focused on a limited number of agronomic traits; therefore, future experiments can be designed to investigate the effects of SA on additional key agronomic and physiological traits. More research is needed to determine the exact role of SA and PN sprays in the performance of wheat genotypes with the *Gpc-B1* gene in field experiments.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding authors.

Author contributions

AS, HR, VSS, PS, GSM and SS conceptualized the study, MJT, SG, PK, and PG performed the phenotyping and recorded the original data. MJT analyzed the data. MJT and AS wrote the original draft. DKS helped MJT during this research and further extended his support for improvement of the original draft. All authors contributed to the article and approved the submitted version.

Acknowledgments

The first author is extremely thankful to the USAID-supported Grain Research and Innovation (GRAIN) project for providing the GRAIN graduate degree scholarship to MJT for his doctoral studies.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1107705/full#supplementary-material

References

Abdi, N., Van Biljon, A., Steyn, C., and Labuschagne, M. T. (2022). Salicylic acid improves growth and physiological attributes and salt tolerance differentially in two bread wheat cultivars. *Plants* 11, 14. doi: 10.3390/plants11141853

Agrawal, S., Yadav, V. K., Verma, A. K., and Mishra, R. C. (2021). Studies on foliar application of potassium nitrate on morpho-physiological and yield potential in different wheat (*Triticum aestivum l.*) cultivars under two condition conditions. *Int. J. Agric. Sci.* 12 (2), 41–45. doi: 10.53390/ijas.v12i2.1

Ahmad, H., Khan, I., Liaqat, W., Jan, M. F., and Ahmadzai, M. D. (2018). Effect of salicylic acid on yield and yield components of maize under reduced irrigation. *Int. J. Environ. Sci. Nat. Res.* 9 (3), 1–5. doi: 10.19080/ARTOAJ.2018.15.555947

Ahmad, S., Wang, G. Y., Muhammad, I., Zeeshan, M., and Zhou, X. B. (2022). Melatonin and KNO_3 application improves growth, physiological and biochemical characteristics of maize seedlings under waterlogging stress conditions. *Biology* 11, 1. doi: 10.3390/biology11010099

Aldesuquy, H. S., Ibraheem, F. L., and Ghanem, H. E. (2018). Exogenously supplied salicylic acid and trehalose protect growth vigor, chlorophylls and thylakoid membranes of wheat flag leaf from drought-induced damage. *J. Agric. For. Meteo. Res.* 1 (1), 13–20.

Ali, L. G., Nulit, R., Ibrahim, M. H., and Yien, C. Y. S. (2021). Efficacy of KNO3, SiO2 and SA priming for improving emergence, seedling growth and antioxidant enzymes of rice (*Oryza sativa*), under drought. *Sci. Rep.* 11 (3864), 1–11. doi: 10.1038/s41598-021-83434-3

Ávila, R. G., Magalhães, P. C., Silva, E. M. D., Alvarenga, A. A. D., Reis, C. O. D., Custódio, A. M., et al. (2022). Foliar application of potassium nitrate induces tolerance to water deficit in pre-flowering sorghum plants. *Acta Scientiarum. Agron.* 44, 1–12. doi: 10.4025/actasciagron.v44i1.53069

Azmat, A., Yasmin, H., Hassan, M. N., Nosheen, A., Naz, R., Sajjad, M., et al. (2020). Co-Application of bio-fertilizer and salicylic acid improves growth, photosynthetic pigments and stress tolerance in wheat under drought stress. *PeerJ* 8, e9960. doi: 10.7717/peerj.9960

Besseau, S., Li, J., and Palva, E. T. (2012). *WRKY54* and *WRKY70* co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. *J. Expt. Bot.* 63 (7), 2667–2679. doi: 10.1093/jxb/err450

Devi, K. N., Devi, K. M., Lhungdim, J., Athokpam, H., Chanu, Y. B., Singh, K. L., et al. (2017). Foliar application of potassium under water deficit conditions improved the growth and yield of wheat (*Triticum aestivum l.*). *Bioscan.* 12 (3), 1777–1781.

Dias, R. C., Melo, C. A. D., Tropaldi, L., Silva, P. V., Silva, D. V., and Reis, M. R. (2021). Effect of salicylic acid in two stages of application on wheat crop. *Rev. Bras. Ciências Agrárias.* 16 (3), 1–7. doi: 10.5039/agraria.v16i3a8374

Distelfeld, A., Avni, R., and Fischer, A. M. (2014). Senescence, nutrient remobilization, and yield in wheat and barley. *J. Exp. Bot.* 65 (14), 3783–3798. doi: 10.1093/jxb/ert477

Elhindi, K. M., El-Hendawy, S., Abdel-Salam, E., Schmidhalter, U., Rahman, S., and Hassan, A. (2016). Foliar application of potassium nitrate effects the growth and photosynthesis in coriander (*Coriander sativum l.*) plants under salinity. *Prog. Nutt.* 18 (1), 63–73.

El-Tayeb, M. A. (2005). Response of barley grain to the interactive effect of salinity and salicylic acid. *Plant Physio. Biochemi.* 36 (10), 767–772. doi: 10.1007/s10725-005-4928-1

Fan, Y., Lv, Z., Li, Y., Qin, B., Song, Q., Ma, L., et al. (2022). Salicylic acid reduces wheat yield loss caused by high temperature stress by enhancing the photosynthetic performance of the flag leaves. *Agronomy* 12, 6. doi: 10.3390/agronomy12061386

Fariduddin, Q., Hayat, S., and Ahmad, A. (2003). Salicylic acid influences net photosynthetic rate, carboxylation efficiency, nitrate reductase activity, and seed yield in *Brassica juncea l. Photosynthetica* 41 (2), 281–284. doi: 10.1023/B:PHOT.0000011962.05991.6c

Gill, H. S., Halder, J., Zhang, J., Rana, A., Kleinjan, J., Amand, P. S., et al. (2022). Whole-genome analysis of hard winter wheat germplasm identifies genomic regions associated with spike and kernel traits. *Theor. Appl. Genet.* 135 (9), 2953–2967. doi: 10.1007/s00122-022-04160-6

Gonzalez-Bayon, R., Shen, Y., Groszmann, M., Zhu, A., Wang, A., Allu, A. D., et al. (2019). Senescence and defense pathways contribute to heterosis. *Plant Physio.* 180 (1), 240–252. doi: 10.1104/pp.18.01205

Gudi, S., Kumar, P., Singh, S., Tanin, M. J., and Sharma, A. (2022b). Strategies for accelerating genetic gains in crop plants: special focus on speed breeding. *Phys. Mol. Bio. Plants.*, 1921–1938. doi: 10.1007/s12298-022-01247-8

Gudi, S., Saini, D. K., Singh, G., Halladakeri, P., Kumar, P., Shamshad, M., et al. (2022a). Unravelling consensus genomic regions associated with quality traits in wheat using meta-analysis of quantitative trait loci. *Planta* 255 (6), 1–19. doi: 10.1007/s00425-022-03904-4

Guo, P., Li, Z., Huang, P., Li, B., Fang, S., and Chu, J. (2017). A tripartite amplification loop involving the transcription factor WRKY75, salicylic acid, and reactive oxygen species accelerates leaf senescence. *Plant Cell.* 29 (11), 2854–2870. doi: 10.1105/tpc.17.00438

Gupta, P. K., Balyan, H. S., Chhuneja, P., Jaiswal, J. P., Tamhankar, S., Mishra, V. K., et al. (2022). Pyramiding of genes for grain protein content, grain quality, and rust resistance in eleven Indian bread wheat cultivars: a multi-institutional effort. *Mol. Breedi.* 42 (4), 1–16. doi: 10.1007/s11032-022-01277-w

Hasanuzzaman, M., Bhuyan, M. H. M., Nahar, K., Hossain, M. D., Mahmud, J. A., Hossen, M., et al. (2018). Potassium: a vital regulator of plant responses and tolerance to abiotic stresses. *Agronomy* 8 (3), 1–29. doi: 10.3390/agronomy8030031

Hasanuzzaman, M., Matin, M. A., Fardus, J., Hasanuzzaman, M., Hossain, M. S., and Parvin, K. (2019). Foliar application of salicylic acid improves growth and yield attributes by upregulating the antioxidant defense system in *Brassica campestris* plants grown in lead-amended soils. *Acta Agrobota.* 72 (2), 1–16. doi: 10.5586/aa.1765

Ibrahim, O. M., Bakry, B. A., Thalooth, A. T., and El-Karamany, M. F. (2014). Influence of nitrogen fertilizer and foliar application of salicylic acid on wheat. *Agric. Sci.* 5 (13), 1316–1325. doi: 10.4236/as.2014.513140

Islam, M. R., Ishida, M., Ando, S., Nishida, T., and Yamada, T. (2003). A method for determination of nitrogen in ruminant feedstuffs and products. *Asian-Aust. J. @ Ani. Sci.* 16 (10), 1438–1442. doi: 10.5713/ajas.2003.1438

Jatana, B. S., Ram, H., Gupta, N., and Kaur, H. (2021). Wheat response to foliar application of salicylic acid at different sowing dates. *J. Crop Improv.* 36 (3), 369–388. doi: 10.1080/15427528.2021.1971131

Jini, D., and Joseph, B. (2017). Physiological mechanism of salicylic acid for alleviation of salt stress in rice. *Rice Sci.* 24 (2), 97–108. doi: 10.1016/j.rsci.2016.07.007

Joppa, L. R., and Cantrell, R. G. (1990). Chromosomal locations of genes for grain protein content of wild tetraploid wheat. *Crop Sci.* 30 (5), 1059–1064. doi: 10.2135/cropsci1990.0011183X003000050021x

Kade, M. A., Barneix, J., Olmos, S., and Dubcovsky, J. (2005). Nitrogen uptake and remobilization in tetraploid langdon durum wheat and a recombinant substitution line with the high grain protein gene *Gpc-B1*. *Plant Breed.* 124 (4), 343–349. doi: 10.1111/j.1439-0523.2005.01110.x

Kamal, N. M., Gorafi, Y. S. A., Abdelrahman, M., Abdellatef, E., and Tsujimoto, H. (2019). Stay-green trait: a prospective approach for yield potential, and drought and heat stress adaptation in globally important cereals. *Int. J. Mol. Sci.* 20, 23. doi: 10.3390/ijms20235837

Karim, F. M., and Khursheed, M. Q. (2011). Effect of foliar application of Salicylic acid on growth, yield components and chemical constituents of Wheat (*Triticum aestivum L. var. Cham 6*). *5th Scientific Conference of College of Agriculture-Tikrit University from 26-27 April 2011.* 672–682.

Khan, M. I. R., Fatma, M., Per, T. S., Anjum, N. A., and Khan, N. A. (2015). Salicylic acid -induced abiotic stress tolerance and underlying mechanisms in plants. *Fron. Plant Sci.* 6. doi: 10.3389/fpls.2015.00462

Koppolu, R., and Schnurbusch, T. (2019). Developmental pathways for shaping spike inflorescence architecture in barley and wheat. *J. Integ. Plant Bio.* 61 (3), 278–295. doi: 10.1111/jipb.12771

Li, T., Deng, G., Tang, Y., Su, Y., Wang, J., Cheng, J., et al. (2021). Identification and validation of a novel locus controlling spikelet number in bread wheat (*Triticum aestivum l.*). *Fro. Plant Sci.* 12. doi: 10.3389/fpls.2021.611106

Luche, H. D. S., Silva, J. A. G. D., Nornberg, R., Hawerroth, M. C., Silveira, S. F. D. S., Caetano, V. D. R., et al. (2016). Stay-green character and its contribution in Brazilian wheats. *Ciênc. Rural.* 47 (1), 1–7. doi: 10.1590/0103-8478cr20160583

Muhal, S., and Solanki, N. S. (2016). Effect of seeding dates and salicylic acid foliar spray on growth, yield, phenology and agrometeorological indices of brassica species. *J. @ Oilseed Brassica.* 1 (1), 183–190.

Pinto, R. S., Lopes, M. S., Collins, N. C., and Reynolds, M. P. (2016). Modelling and genetic dissection of staygreen under heat stress. *Theo. App. Gen.* 129 (11), 2055–2074. doi: 10.1007/s00122-016-2757-4

Piper, C. S. (2019). *Soil and plant analysis* (India: Scientific Publishers).

Pirasteh-Anosheh, H., Ranjbar, G., Hasanuzzaman, M., Khanna, K., Bhardwaj, R., and Ahmad, P. (2022). Salicylic acid -mediated regulation of morpho-physiological and yield attributes of wheat and barley plants in deferring salinity stress. *J. Plant Growth Reg.* 41 (3), 1291–1303. doi: 10.1007/s00344-021-10358-7

Safar-Noori, M., Assaha, D. V. M., and Saneoka, H. (2018). Effect of salicylic acid and potassium application on yield and grain nutritional quality of wheat under drought stress condition. *Cereal Res. Commun.* 46 (3), 558–568. doi: 10.1556/0806.46.2018.026

Saini, D. K., Devi, P., and Kaushik, P. (2020). Advances in genomic interventions for wheat biofortification: a review. *Agronomy* 10 (62), 1–28. doi: 10.3390/agronomy10010062

Saini, P., Sheikh, I., Saini, D. K., Mir, R. R., Dhaliwal, H. S., and Tyagi, V. (2022b). Consensus genomic regions associated with grain protein content in hexaploid and tetraploid wheat. *Front. Genet.* 13. doi: 10.3389/fgene.2022.1021180

Saini, D. K., Srivastava, P., Pal, N., and Gupta, P. K. (2022a). Meta-QTLs, ortho-meta-QTLs and candidate genes for grain yield and associated traits in wheat (Triticum aestivum l.). *Theor. Appl. Genet.* 135 (3), 1049–1081. doi: 10.1007/s00122-021-04018-3

Saputra, R., Hamzah, A., Puspita, F., Irfandri.,, Nasrul, B., and Naibaho, T. E. (2022). "Study of salicylic acid concentration in suppressing the development of ganoderma spp. causes of stem rot disease *in vitro* and its effect on oil palm seedling growth," in *IOP Conference Series: Earth and Environmental Science*, Vol. 978, 1–6.

Sharma, A., Srivastava, P., Mavi, G. S., Kaur, S., Kaur, J., Bala, R., et al. (2021). Resurrection of wheat cultivar PBW343 using marker-assisted gene pyramiding for rust resistance. *Fron. Plant Sci.* 12 (570408), 1–19. doi: 10.3389/fpls.2021.570408

Shewry, P. R. (2009). Wheat. *J. Exp. Bot.* 60 (6), 1537–1553. doi: 10.1093/jxb/erp058

Simaei, M., Khavari-Nejad, R. A., and Bernard, F. (2012). Exogenous application of salicylic acid and nitric oxide on the ionic contents and enzymatic activities in NaCl-stressed soybean plants. *Am. J. Plant Sci.* 3, 1495–1503. doi: 10.4236/ajps.2012.310180

Singh, R., Kaur, N., Praba, U. P., Kaur, G., Tanin, M. J., Kumar, P., et al. (2022). A prospective review on selectable marker-free genome engineered rice: past, present and future scientific realm. *Front. Gen.* 13 (882836), 1–15. doi: 10.3389/fgene.2022.882836

Singh, K., Saini, D. K., Saripalli, G., Batra, R., Gautam, T., Singh, R., et al. (2022). WheatQTLdb V2.0: a supplement to the database for wheat QTL. *Mol. Breed.* 42, 56. doi: 10.1007/s11032-022-01329-1

Singh, K. M., and Singh, H. K. (2020). Effect of foliar application of potassium nitrate on late sown wheat (*Triticum aestivum l.*) in mitigating terminal heat stress. *J. Pharmacogn. Phytochem.* 9 (6), 492–495.

Singh, S., Singh, H., and Choudhary, R. N. (2013). Heat stress management in late sown wheat (*Triticum aestivum l.*) under climate change scenario. *Annals. Agric. Bio. Res.* 18 (2), 135–138.

Singh, P. K., Singh, S., Singh, M. P., Verma, R. K., Singh, V. B., and Pandey, S. K. (2021). Beneficial effects of foliar application of potassium nitrate in late-sown wheat in village gaurahi of district sonbhadra. *Pharma Innov. J.* 10 (4), 95–97.

Singh, P., Yadav, V. K., Khan, A. H., and Kumar, A. (2019). Effect of salicylic acid on growth, physio-chemical changes and yield of wheat (*Triticum aestivum l.)* under date of sowing at anthesis. *J. Pharmacog. Phytochem.* 8 (4), 721–726.

Smith, L. M. (2019). Salicylic acid, senescence, and heterosis. *Plant physio.* 180 (1), 3–4. doi: 10.1104/pp.19.00260

Stanislawska-Glubiak, E., and Korzeniowska, J. (2021). Effect of salicylic acid foliar application on two wheat cultivars grown under zinc stress. *Agronomy* 12 (1), 1–11. doi: 10.3390/agronomy12010060

Sultana, N., Islam, S., Juhasz, A., and Ma, W. (2021). Wheat leaf senescence and its regulatory gene network. *Crop J.* 9 (4), 703–717. doi: 10.1016/j.cj.2021.01.004

Tabbita, F., Pearce, S., and Barneix, A. J. (2017). Breeding for increased grain protein and micronutrient content in wheat: ten years of the *Gpc-B1* gene. *J. Cereal Sci.* 73, 183–191. doi: 10.1016/j.jcs.2017.01.003

Tanin, M. J., Kumar, A., and Gupta, S. K. (2018). Phenotypic stability for yield and some quality traits in *Brassica juncea l. Int. J. Micorbiol. App. Sci.* 7 (2), 479–485. doi: 10.20546/ijcmas.2018.702.060

Tanin, M. J., Saini, D. K., Sandhu, K. S., Pal, N., Gudi, S., Chaudhary, J., et al. (2022a). Consensus genomic regions associated with multiple abiotic stress tolerance in wheat and implications for wheat breeding. *Sci. Rep.* 12 (1), 1–17. doi: 10.1038/s41598-022-18149-0

Tanin, M. J., Sharma, A., Saini, D. K., Singh, S., Kashyap, L., Srivastava, P., et al. (2022b). Ascertaining yield and grain protein content stability in wheat genotypes having the *Gpc-B1* gene using univariate, multivariate, and correlation analysis. *Fron. Gen.* 13 (1001904), 1–21. doi: 10.3389/fgene.2022.1001904

Thomas, H., and Ougham, H. (2014). The stay-green trait. *J. Exp. Botany.* 65 (14), 3889–3900. doi: 10.1093/jxb/eru037

Uauy, C., Brevis, J. C., Chen, X., Khan, I., Jackson, L., Chicaiza, O., et al. (2005). High-temperature adult-plant (HTAP) stripe rust resistance gene *Yr36* from triticum turgidum ssp. dicoccoides is closely linked to the grain protein content locus *Gpc-B1*. *Theor. Appl. Genet.* 112 (1), 97–105. doi: 10.1007/s00122-005-0109-x

Uauy, C., Brevis, J. C., and Dubcovsky, J. (2006). The high grain protein content gene *Gpc-B1* accelerates senescence and has pleiotropic effects on protein content in wheat. *J. Exp. Bot.* 57 (11), 2785–2794. doi: 10.1093/jxb/erl047

Wada, K. C., and Takeno, K. (2010). Stress-induced flowering. *Plant Signal. Behav.* 5 (8), 944–947. doi: 10.4161/psb.5.8.11826

Wang, L., Feng, Z., and Schjoerring, J. K. (2013). Effects of elevated atmospheric CO_2 on physiology and yield of wheat (*Triticum aestivum l.*): a meta-analytic test of current hypotheses. *Agric. Ecosyst. Environ.* 178, 57–63. doi: 10.1016/j.agee.2013.06.013

Wang, W., Wang, X., Huang, M., Cai, J., Zhou, Q., Dai, T., et al. (2021). Alleviation of field low-temperature stress in winter wheat by exogenous application of salicylic acid. *J. Plant Growth Reg.* 40, 811–823. doi: 10.1007/s00344-020-10144-x

Wolfe, R. R., Cifelli, A. M., Kostas, G., and Kim, I. Y. (2017). Optimizing protein intake in adults: interpretation and application of the recommended dietary allowance compared with the acceptable macronutrient distribution range. *Adv. Nutr.* 8 (2), 266–275. doi: 10.3945/an.116.013821

Xiao, S., Gao, W., Chen, Q. F., Chan, S. W., Zheng, S. X., and Ma, J. (2010). Overexpression of arabidopsis acyl-CoA binding protein ACBP3 promotes starvation-induced and age-dependent leaf senescence. *Plant Cell.* 22 (5), 1463–1482. doi: 10.1105/tpc.110.075333

Yadav, R. K., Kumar, M., Saini, P. K., Singh, A. K., Singh, A. K., Kumar, B., et al. (2021). Effect of foliar application of different nutrients on growth and yield of wheat (Triticum aestivum l.) under sodic soil. *Pharma Inn. J.* 10 (1), 589–594. doi: 10.22271/tpi.2021.v10.i1h.5594

Yadav, V. K., Singh, A. K., Singh, P., Kumar, R., Srivastav, S. K., Yadav, V., et al. (2019). Effect on foliar application of potassium nitrate on physiological, biochemical and yield potential in different wheat (*Triticum aestivum l.)* cultivars under drought and irrigated conditions. *J. Pharmacogn. Phytochemi.* 8 (2), 254–257.

Yoshimoto, K., Jikumaru, Y., Kamiya, Y., Kusano, M., Consonni, C., and Panstruga, R. (2009). Autophagy negatively regulates cell death by controlling NPR1-dependent salicylic acid signalling during senescence and the innate immune response in arabidopsis. *Plant Cell.* 21 (19), 2914–2927. doi: 10.1105/tpc.109.068635

Zhang, J., Li, H., Xu, B., Li, J., and Huang, B. (2016). Exogenous melatonin suppresses dark-induced leaf senescence by activating the superoxide dismutase-catalase antioxidant pathway and down-regulating chlorophyll degradation in excised leaves of perennial ryegrass (*Loliumperenne l.*). *Front. Plant Sci.* 7 (1500), 1–15. doi: 10.3389/fpls.2016.01500

Zheng, J. J., Zhao, H. J., Hu, W. W., Zhao, X. J., and Zhao, Y. D. (2013). Effect of heat and high irradiation stress on Deg1 protease and D1 protein in wheat chloroplasts and the regulating role of salicylic acid. *Acta Ecol. Sin.* 33 (9), 2930–2935.

Zhou, H., Riche, A. B., Hawkesford, M. J., Whalley, W. R., Atkinson, B. S., Sturrock, C. J., et al. (2021). Determination of wheat spike and spikelet architecture and grain traits using X-ray computed tomography imaging. *Plant Methods* 17 (1), 1–9. doi: 10.1186/s13007-021-00726-5

Ziasmin,, Islam, M. M., Ahmed, K. U., Islam, S., and Parvin, S. (2017). Response of wheat to salicylic acid. *Int. J. Sci. Res. Publ.* 7 (2), 738–742.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 August 2023
DOI 10.3389/fpls.2023.1226297

OPEN ACCESS

EDITED BY
Shouvik Das,
Regional Centre for Biotechnology (RCB), India

REVIEWED BY
Shakeel Ahmad,
Ministry of Environment, Water and Agriculture, Saudi Arabia
Yohei Koide,
Hokkaido University, Japan

*CORRESPONDENCE
Joong Hyoun Chin
jhchin@sejong.ac.kr

RECEIVED 21 May 2023
ACCEPTED 31 July 2023
PUBLISHED 18 August 2023

CITATION
Navea IP, Maung PP, Yang S, Han J-H, Jing W, Shin N-H, Zhang W and Chin JH (2023) A meta-QTL analysis highlights genomic hotspots associated with phosphorus use efficiency in rice (*Oryza sativa* L.).
Front. Plant Sci. 14:1226297.
doi: 10.3389/fpls.2023.1226297

A meta-QTL analysis highlights genomic hotspots associated with phosphorus use efficiency in rice (*Oryza sativa* L.)

Ian Paul Navea[1,2], Phyu Phyu Maung[1,2], Shiyi Yang[3], Jae-Hyuk Han[1,4], Wen Jing[3], Na-Hyun Shin[1,2], Wenhua Zhang[3] and Joong Hyoun Chin[1,2]*

[1]Food Crops Molecular Breeding Laboratory, Department of Integrative Biological Sciences and Industry, Sejong University, Seoul, Republic of Korea, [2]Convergence Research Center for Natural Products, Sejong University, Seoul, Republic of Korea, [3]College of Life Sciences, National Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Nanjing Agricultural University, Nanjing, China, [4]The International Rice Research Institute-Korea Office, National Institute of Crop Science, Rural Development Administration, Iseo-myeon, Republic of Korea

Phosphorus use efficiency (PUE) is a complex trait, governed by many minor quantitative trait loci (QTLs) with small effects. Advances in molecular marker technology have led to the identification of QTLs underlying PUE. However, their practical use in breeding programs remains challenging due to the unstable effects in different genetic backgrounds and environments, interaction with soil status, and linkage drag. Here, we compiled PUE QTL information from 16 independent studies. A total of 192 QTLs were subjected to meta-QTL (MQTL) analysis and were projected into a high-density SNP consensus map. A total of 60 MQTLs, with significantly reduced number of initial QTLs and confidence intervals (CI), were identified across the rice genome. Candidate gene (CG) mining was carried out for the 38 MQTLs supported by multiple QTLs from at least two independent studies. Genes related to amino and organic acid transport and auxin response were found to be abundant in the MQTLs linked to PUE. CGs were cross validated using a root transcriptome database (RiceXPro) and haplotype analysis. This led to the identification of the eight CGs (*OsARF8, OsSPX-MFS3, OsRING141, OsMIOX, HsfC2b, OsFER2, OsWRKY64*, and *OsYUCCA11*) modulating PUE. Potential donors for superior PUE CG haplotypes were identified through haplotype analysis. The distribution of superior haplotypes varied among subspecies being mostly found in *indica* but were largely scarce in *japonica*. Our study offers an insight on the complex genetic networks that modulate PUE in rice. The MQTLs, CGs, and superior CG haplotypes identified in our study are useful in the combination of beneficial alleles for PUE in rice.

KEYWORDS

rice, phosphorus use efficiency, Meta-QTL analysis, Quantitative Trait Loci, candidate genes, superior haplotypes

1 Introduction

Phosphorus (P) is one of the most important macronutrients in plants and is required in large quantities. Inorganic phosphate (Pi) is a crucial component of phospholipids and plays a significant role in deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) synthesis and other nucleotide-containing molecules. Moreover, P present in adenosine triphosphate (ATP), adenosine diphosphate (ADP), and nicotinamide adenine dinucleotide phosphate (NADPH), provides plant cells with energy required for metabolic and catabolic cellular processes (Heuer et al., 2017). In rice (*Oryza sativa*), P deficiency substantially decreases overall productivity by reducing plant height, tiller number, panicle length (Irfan et al., 2020) and suppresses the root growth (Liu, 2021). Delayed maturity, high sterility, and poor grain quality (Ismail et al., 2007) have also been reported.

Approximately 5.7 billion hectares of land lack plant-available P due to its immobility in soil (Dobermann and Fairhurst, 2000). P tends to get fixed in soils with extreme levels of pH due to the complexation of P by aluminum (Al) or iron (Fe) in acidic conditions and by calcium (Ca) in alkaline soils (Haefele et al., 2014). Unlike N, which has an unlimited source due to its abundance in the atmosphere thanks to the Haber-Bosch process (Lynch, 2011), P fertilizer sources are finite, mainly consisting of phosphate rocks that are estimated to last for only 300~400 years (Shimizu et al., 2004; Van Kauwenbergh, 2010).

Plants have evolved to cope with P-starvation by undergoing physiological changes in their root morphology, such as the promotion of lateral root and root hair growth, and the inhibition of primary root development. Additionally, plants foster symbiosis between their roots and arbuscular mycorrhizal fungi to scavenge Pi from the soil (Péret et al., 2011). The uptake of P in plants is determined by three primary factors: (1) root morphology; (2) P uptake efficiency; and (3) internal P-use efficiency. Among these mechanisms, genetic variations in root morphology are the primary causal factor of P uptake, whereas P uptake efficiency and internal P-use contribute less (Ismail et al., 2007). Consequently, enhancing root morphology through breeding efforts may lead to the development of rice varieties with high phosphorus use efficiency (PUE).

Development of P-efficient rice varieties can be achieved through improved uptake of phosphate from soil (P-acquisition efficiency) (Mori et al., 2016) and improved biomass and/or yield per unit P taken up [internal P-utilization efficiency/P-use efficiency] (PUE) (Rose and Wissuwa, 2012). Developing P-efficient rice genotypes has gained significant attention to breeders. The major PUE quantitative trait locus (QTL), *Pup1*, has been extensively utilized due to its large additive effect. The QTL was mapped from a backcross population derived from a cross between the P-deficiency intolerant Nipponbare and P-deficiency tolerant Kasalath and is located on the long arm of chromosome 12. The use of *Pup1* in molecular marker-assisted backcrossing (MABC) has proven successful (Chin et al., 2011). The *Pup1* QTL harbors the Kasalath-derived *OsPSTOL1* gene, encoding a protein kinase that enhances early-stage root morphology in rice (Gamuyao et al., 2012). Similarly, a minor QTL on the long arm of chromosome 6 (Shimizu et al., 2004) where three of the known rice Pi-responsive regulatory genes *OsERF3*, *OsTHS1* (Wasaki et al., 2003), and *OsPTF1* (Yi et al., 2005) collocate, has been mapped but is yet to be implemented in large-scale breeding programs. In addition, numerous promising genes that are involved in PUE have been identified through overexpression and knockout studies in rice and are now undergoing advanced testing in cereal crops (Heuer et al., 2017). These genes include *AVP1* (Yang et al., 2014), *PHO1* (Khan et al., 2014), *OsPHT1;6* (Ai et al., 2009), and *SPX-MFS* (Wang et al., 2015). Moreover, the next-generation sequencing (NGS) has facilitated the development of high-density rice linkage maps, which have been utilized to identify QTLs associated with various rice traits, including PUE. Several studies have used SNP-based linkage maps to identify PUE QTLs (Koide et al., 2013; Ogawa et al., 2014; Navea et al., 2017; Fu et al., 2019; Ranaivo et al., 2022; Wang K et al., 2014).

Despite the progress in ascertaining QTLs and genes underlying PUE, their practical utility in breeding programs remain elusive due to their unstable effects across different genetic backgrounds and environments (Arcade et al., 2004; Daware et al., 2017; Navea et al., 2022) as well as their complex relationship with soil status (Heuer et al., 2017) and linkage drag (Kumar et al., 2020). It is therefore necessary to identify genomic regions conferring robust and stable effects across a wide range of genetic background and environments, as well as QTLs with small CI to increase the efficiency and precision of genomics-assisted breeding (Collard and Mackill, 2008).

Meta-QTL analysis is a powerful tool that can help achieve breeding precision by compiling QTLs identified from various mapping populations used in independent studies. It can also provide target genomic regions with considerably small CI (Goffinet and Gerber, 2000; Khahani et al., 2021). Previous studies successfully identified the meta-QTL's (MQTL) underlying yield (Khahani et al., 2021; Aloryi et al., 2022), nitrogen-use efficiency (Sandhu et al., 2021), salinity tolerance (Islam et al., 2019), grain zinc content (Joshi et al., 2023), drought (Selamat and Nadarajah, 2021), and grain traits (Selamat and Nadarajah, 2021) in rice. However, to date, MQTLs associated with PUE remain to be explored in rice. The present study aims to fill this gap by identifying promising MQTLs through an extensive literature search on previous PUE QTLs mapped in various independent studies and identify potential rice donors for pyramiding of the beneficial alleles of the candidate genes (CGs) identified within the MQTLs.

2 Materials and methods

2.1 Compilation of PUE QTLs from various studies

An exhaustive literature search on rice QTLs linked to PUE from studies published between 1998 and 2022 was performed using the publicly available QTL databases, such as PubMed (https://pubmed.ncbi.nlm.nih.gov), Google scholar (https://scholar.google.com), and Gramene QTL database (https://

archive.gramene.org). QTLs and traits associated with PUE were defined as traits contributing to biomass and yield (Rose and Wissuwa, 2012) under low P condition, such as shoot dry weight (SDW); absolute value of root traits (RT); root-shoot ratio (RSR); total dry biomass (BM); seed P content (SPC); relative response to P (RRP); internal P translocation (IPT); and yield component (YLD). QTLs identified under low P input were chosen for the meta-QTL (MQTL) anlaysis. Additionally, only QTL mapping studies carried out under field conditions were selected. Sixteen independent studies were used to perform the MQTL analysis. A total of 192 PUE QTLs mapped from 15 non-overlapping bi-parental populations were utilized in the analysis. Information on the type of parental lines, population types, size of the mapping population, type of molecular markers, number of QTLs, and trait types were recorded. Individual QTLs were assessed for QTL names, type of mapping population, trait class, year of study, linkage group, logarithm of odds (LOD) scores, phenotypic variance explained (PVE), peak position (cM), and CI. These attributes were used to generate the "QTL file". The peak positions were inferred from the midpoint between two flanking markers whenever information on the peak position was lacking. An LOD threshold of 3.0 was set when LOD scores were not supplied. Missing CI values were estimated using the following formulas proposed in a previous study (Darvasi and Seller, 1997; Weller and Soller, 2004):

$$(\mathrm{I})\mathrm{CI} = \frac{163}{\mathrm{N} \times \mathrm{PVE}} \quad (\mathrm{II})\mathrm{CI} = \frac{530}{\mathrm{N} \times \mathrm{PVE}}$$

Where N = population size and PVE = phenotypic variance explained. Equation I was used to estimate CI in recombinant inbred lines (RILs) whereas equation II was used for populations derived from F_2, backcross inbred lines (BILs), and chromosome substitution lines (CSSLs). Missing PVE values were calculated using the following formula (Nagelkerke, 1991):

$$\mathrm{PVE} = 1 - 10\left(-\frac{2\mathrm{LOD}}{\mathrm{N}}\right)$$

2.2 Construction of consensus map and QTL projection

A high-density consensus map was constructed by integrating SNP markers from the 6K Infinium SNP array (Thomson et al., 2017) and the flanking markers of 192 QTLs. The physical locations of the flanking markers were determined by aligning their sequences to the Nipponbare reference genome (IRGSP v. 1.0) using the Basic Local Alignment Tool (https://www.ncbi.nlm.nih.gov).Then, the closest markers from the Infinium SNP array were used to project QTLs on the consensus map. Markers were arranged according to their physical positions. The consensus map was created by converting the physical position into centimorgan units (cM) using the conversion factor of 1 cM = 250 kb (Raghavan et al., 2017). An individual "map file" for each chromosome was generated, containing information on linkage group, marker name, and the genetic position in centimorgan (cM). The QTL and map files were used as input files to project the consolidated QTLs on the consensus map.

2.3 Meta-QTL analysis

After QTL projection, MQTL analysis was performed using Biomercator v4.2.3 software (Arcade et al., 2004; Sosnowski et al., 2012). Two different approaches were applied based on the number of QTLs on each chromosome, namely the Goffinet and Gerber method when the number of QTLs was ≤ 10, and the Veyrieras approach when the number of QTLs was > 10 (Goffinet and Gerber, 2000; Veyrieras et al., 2007). The algorithms and statistical procedures for both methods were previously described in detail (Sosnowski et al., 2012). The "true" number of MQTLs per chromosome was defined from the model with the lowest Akaike information criterion (AIC) value, which was selected as it contained the least amount of information loss (Akaike, 1987). The PVE value for each MQTL were estimated using that of its "initial QTL" members. The best model, along with its corresponding AIC values, MQTL peak positions, 95% CI, and physical positions, are presented in (Table 1). Additionally, only MQTLs with an average PVE value of ≥ 5% and supported by QTLs identified in at least two independent studies were selected for further analysis. The QTLs identified under P-deficient or P-non-supplied conditions were categorized into the following traits: 1) shoot dry weight (SDW); 2) absolute values of root traits (RT); 3) root-shoot ratio (RSR); 4) total dry biomass (BM); 5) seed P content (SPC); 6) relative response to P (RRP); 7) internal P translocation (IPT); and 8) yield components (YLD).

2.4 Identification of CGs underlying PUE in rice

After identifying MQTLs with an average PVE value of ≥ 5% and support from at least two independent studies using the model with the lowest AIC value, we aimed to mine CGs associated with PUE in rice. To this end, we utilized the physical position of markers flanking the MQTLs as query terms in the Rice Annotation Project Database (IRGSP v1.0 and MSU 7) to batch download functionally annotated gene models within the MQTLs. All genes within the MQTLs were initially considered as CGs, which were further filtered based on gene ontology (GO) terms and/or keywords related to PUE such as P homeostasis, phosphate, inorganic phosphate (Pi) transporter, crown root, root hairs, phosphorus translocation, phosphorus uptake, abiotic stress, and/or secondary traits associated with PUE, as described in a review article by Heuer et al. (2017).

To identify P-responsive genes, we conducted *in silico* analysis on the CGs identified in the previous step. We used a microarray dataset (RXP_5002, available at https://ricexpro.dna.affrc.go.jp) containing root gene expression data of 7-day-old seedlings grown under P-deficient and control conditions. Details of the plant growth conditions and treatments can be found on the RiceXPro website (https://ricexpro.dna.affrc.go.jp/RXP_5002/details-of-methods.html). Briefly, 7-day-old Nipponbare seedlings were exposed to P-non-supplied and control (P-supplied) conditions. Root samples were collected at 6- and 24-hour (h) post-treatment for RNA extraction. The RNA was labeled with Cy3

TABLE 1 QTL studies used in the QTL meta-analysis for PUE in rice.

Parents (Subspecies or species)	Population type[a]	Population size	Type of marker[b]	Number of marker	Number of QTLs identified	Traits[c]	LOD scores/ PVE[d] (%)	Country of experiment	Year	Reference
Zhonghui9308 x Xieyou9308 (*indica x indica*)	CSSL	75	SSR + InDel	120	7	SDW, RT, SDW, BM.	2.00~3.32 10.82~18.46	China	2018	Anis et al., 2018
ZYQ8 x JX17 (*indica x japonica*)	DH	127	RFLP	444	6	RRP, RT, RSR	2.03~6.79 8.8~25.2	China	2000	Ming et al., 2000
GH128 x W6827 (*indica x japonica*)	F_2	262	SNP	25,117	21	BM, RRP, YLD, SPC, IPT	2.53~9.50 1.56~8.19	China	2019	Fu et al., 2019
IR20 x IR55178 (*indica x indica*)	RIL	284	AFLP + RFLP	178	6	RRP	3.40~9.10 13~21	China	2001	Hu et al., 2001
Wazuhophek x Samba Mahsuri (*indica x indica*)	RIL	330	SSR	78	15	SDW, YLD, RT, BM, RSR, SPC	5.29~5.84 2.25~21.84	India	2021	Kale et al., 2021
Nerica10 x Hitomebore (*indica x japonica*)	$F_{2:3}$	91	SNP + SSR	128	15	SDW, YLD, BM, IPT	2.60~4.90 12~23.7	Japan	2013	Koide et al., 2013
Mizukagami x OHA15 (*japonica x indica*)	CSSLs + F_2	35 + 176	SSR	9 (Chr. 6 only)	3	YLD, RT, SDW	2.20~12.70 5.5~27.5	Japan	2022	Kokaji et al., 2022
Dasanbyeo x TR22183 (*indica x japonica*)	RIL	172	SNP	236	30	YLD	3.15~12.76 8.9~22.52	Philippines/ Korea	2017	Navea et al., 2017
IR20 x IR55178 (*indica x indica*)	RIL	285	AFLP	217	10	SDW, BM, RRP	2.42~16.98 6.8~19.5	Philippines	1998	Ni et al., 1998
Curinga x IRGC105491 (*O. sativa x O. rufipogon*)	CSSL	48	SNP	238	3	RT	3.0 1.25	Colombia	2014	Ogawa et al., 2014
IRAT109 x Yuefa (*indica x japonica*)	DH	116	RFLP +SSR	165	17	YLD	3.02~5.13 2.65~20.78	China	2008	Ping et al., 2008
DJ123 x Nerica4 (*aus x indica*)	BIL	201	SNP	1578	10	RT, SDW, BM	4.46~7.65 1.5~19.2	Japan and Madagascar	2022	Ranaivo et al., 2022
Gimbozu x Kasalath (*japonica x indica*)	$F_{2:3}$	82	SSR	97	11	SDW, BM, RSR, RRP	3.50~6.46 9.1~24.6	Japan	2004	Shimizu et al., 2004
Zhenshan 97 x Minghui 63 (*indica x indica*)	RIL	113	SNP	1,619	36	BM, YLD, PUP, SPC	3.0 1.4~15.8	China	2014	Wang W et al., 2019

(Continued)

TABLE 1 Continued

Parents (Subspecies or species)	Population type[a]	Population size	Type of marker[b]	Number of marker	Number of QTLs identified	Traits[c]	LOD scores/ PVE[d] (%)	Country of experiment	Year	Reference
Nipponbare x Kasalath (*japonica x indica*)	BIL	98	RFLP	245	13	PUP, SDW, YLD	2.82~10.74 5.8~30	Japan	1998	Wissuwa et al., 1998
Shuhui527 x Yetuozai (*japonica x indica*)	BIL	60	SSR	96	20	RT, RSR, YLD	3.78~6.81 3.6~15.9	China	2015	Zhang et al., 2014

a CSSL, chromosome substitution lines; **DH**, doubled haploid lines; **RIL**, recombinant inbred lines; **$F_{2,3}$**, F_2-derived F_3 lines; **BIL**, backcross inbred lines.
b SSR, single sequence repeats; **RFLP**, restriction fragment length polymorphism; **AFLP**, amplified fragment length polymorphism; **SNP**, single nucleotide polymorphism.
[c] **SDW**, shoot dry weight; **RT**, absolute value of root traits; **RSR,** root-shoot ratio; **BM**, total dry biomass; **SPC**, seed P content; **RRP**, relative response to P; **IPT**, internal P translocation; **YLD**, yield component.
[d] **LOD**,logarithm of odds; **PVE**, phenotypic variation explained.

and subjected to hybridization using the Agilent one-color microarray analysis system. The resulting gene expression profile was presented in terms of raw signal intensity. We carried out gene ontology (GO) enrichment analysis using the Shiny GO 0.76 database (Ge et al., 2020) to determine the most enriched biological pathways in the P-responsive genes with at least 1.5-fold change in expression (P-non-supplied/control condition). We used all the unfiltered genes within the MQTLs as the background against the P-responsive genes in the GO enrichment analysis. In the next step, we further filtered the CGs to those that were responsive to P at both 6- and 24-h post-treatment, with a consistent direction of regulation at both time points, and at least a 2-fold change in expression levels at 24-h post-treatment (P-non-supplied relative to control).

2.5 Statistical analysis

Statistical analysis was performed using the two-tailed t-test to determine the significance of differences in the root gene expression levels at different P application at $P<0.10$, $P<0.05$, and $P<0.01$ using Minitab release v.14.

2.6 Identification of potential donors for the pyramiding of beneficial CG alleles

Haplotype analysis carried out for the PUE CGs with significant responses to P treatment was performed using the SNP Seek database's built-in tool (Mansueto et al., 2016) in the 3K Rice Genome Project (RGP) database. The database included all rice sub-populations with Nipponbare as the reference genome. PUE CGs with non-synonymous SNPs, at least a 2-fold change in gene expression under P-deficient vs. control conditions, and with at least two haplotypes were utilized in the analysis. The number of haplotypes were determined using the default parameters with the Calinski criterion for determining the optimal number of groups (Caliński and Harabasz, 1974). Haplotypes with fewer than three genotypes were excluded. Genotypes with more than 20% missing SNPs or heterozygous loci were filtered out from the data set. Haplotypes identical to those of the beneficial allele donors (whenever available in the 3K RGP dataset) in the QTL studies used in the MQTL analysis were regarded as superior haplotypes. The abundance of superior haplotypes across the subpopulations was calculated, and rice accessions with superior haplotypes in *indica*, *japonica*, and *aus* backgrounds from the 3K RGP were identified. These lines were considered as potential donors for pyramiding superior CG haplotypes for improving PUE in rice.

3 Results

3.1 Features of the QTL studies used in MQTL analysis

The key features of the PUE QTL studies used in the PUE MQTL analysis are presented in (Table 1). A total of 192 QTLs from 16 independent studies published between 1998 and 2022 were used to identify PUE MQTLs. These QTLs were mapped from 15 non-overlapping bi-parental populations, including CSSLs (chromosome segment substitution lines) (3), DH (doubled haploids) (2), F_2 (2), RILs (recombinant inbred lines) (5), $F_{2:3}$ (F_2-derived F_3) (2), and BILs (backcross inbred lines) (3). The markers employed in mapping PUE QTLs included SNP (single nucleotide polymorphism) (6); AFLP (amplified fragment length polymorphism) (2); SSR (simple sequence repeat) (6); RFLP (restriction fragment length polymorphism) (4); and InDel (insertion/deletion polymorphism) (1). The numbers of the initial QTLs per trait group were as follows: SDW (n=18); RT (n=25); RSR (n=4); BM (n=31); SPC (n=10); RRP (n=24); PUE (n=36); YLD (n=44). The distribution of the initial QTLs varied widely across chromosomes (Figure 1). There were largest numbers of QTLs located on chromosomes 2 (n=32) and 6 (n=33). On the other hand, only seven and five QTLs were located on chromosomes 3 and 9, respectively. The PVE values and LOD scores had ranges of 1.25% ~27.9% and 2.0~16.98, respectively. More than half (51%) of the PVE values were between 3 and 6% (Figure 2A), whereas most of the LOD scores were less than five (Figure 2B).

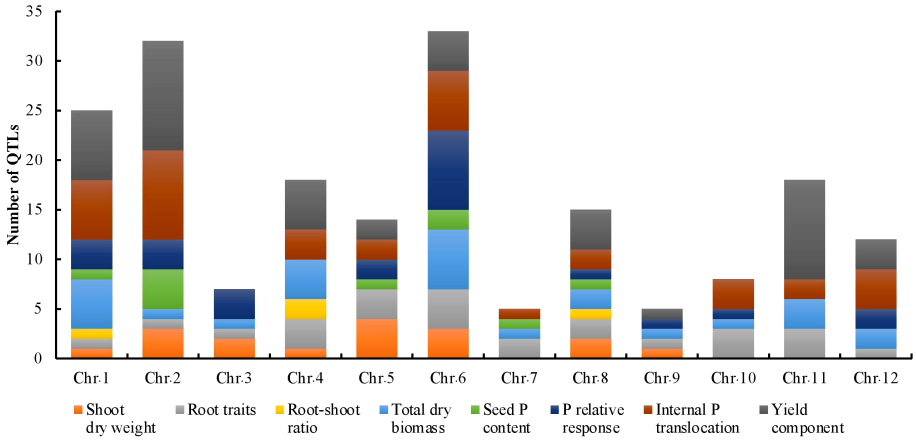


FIGURE 1
Phenotypic trait classes and chromosome-wise distribution of QTLs utilized in the MQTL analysis for PUE in rice.

3.2 The rice consensus map and QTL projection

The consensus map included a total of 5,694 markers, comprising both SNPs from the 6K Infinium SNP array and the flanking markers of the QTLs used in the MQTL analysis. They were well-distributed across the twelve chromosomes. The SNP density per 500 kb window is presented in (Figure S1). In addition, the positions of known PUE genes, such as *OsPSTOL1*, *OsPTF1*, *OsERF3*, and *OsTHS1*, were incorporated onto the consensus map. The cumulative length of the consensus map was 1,637 cM. The average distance between the adjacent markers was 0.29 cM. Chromosomes 1 (171.9 cM) and 6 (160 cM) were the longest, while chromosomes 9 (108.5 cM) and 10 (92.2 cM) were the shortest. The number of markers varied from 332 to 653 per chromosome. Chromosomes 1 and 2 had the highest counts of 653 and 574 markers, respectively, while chromosomes 9 (n=354) and 10 (n=332) were the least saturated. We projected the PUE QTLs onto the consensus map using both the physical position of the QTLs and the SNP markers (Figure 3). The number of QTLs ranged from 5 to 33. Chromosomes 6 (n=33), 2 (n=32), and 1 (n=25) were the most saturated regions, while chromosomes 9 (n=5), 7 (n=5), and 3 (n=7) had the fewest.

3.3 MQTL analysis

MQTL analysis of the 192 initial PUE QTLs resulted in the identification of 60 MQTLs (Table 2; Figure 3). The distribution of MQTLs across chromosomes was uneven. The highest number of MQTLs were detected on chromosomes 1 (n=9) and 5 (n=8), while chromosomes 3 (n=2), 7 (n=2), and 9 (n=2) had the least. Reductions in MQTL CI were observed, with fold reductions ranging from 2.8 to 11.4 with an overall average reduction of 9.7 cM (Figure 4). The CI for MQTLs located on chromosomes 3, 5, 7, 9, and 10 were at least five times smaller than their corresponding initial QTLs, while those on chromosomes 1, 2, 6, and 11 were at least twice as small. Chromosomes 4 and 12 exhibited the least reduction in CI, with fold change values of 1.7 and 1.02, respectively. The MQTLs had PVE values from 2.3 to 20.3.

Subsequently, we selected MQTLs that had an average PVE value of at least five and were supported by QTLs from at least two independent studies (Table 3). This resulted in 38 MQTLs, with the number of QTL members ranging from 2 to 11. MQTL1.1, MQTL8.1, and MQTL8.2 had the most QTLs (n=11), while MQTL3.1, MQTL4.1, and MQTL10.1 had only two initial QTLs. The MQTLs had CI ranging from 0.2 to 9 cM, with an average of 3.5 cM per MQTL. The smallest CI (0.2~2 cM) were observed on

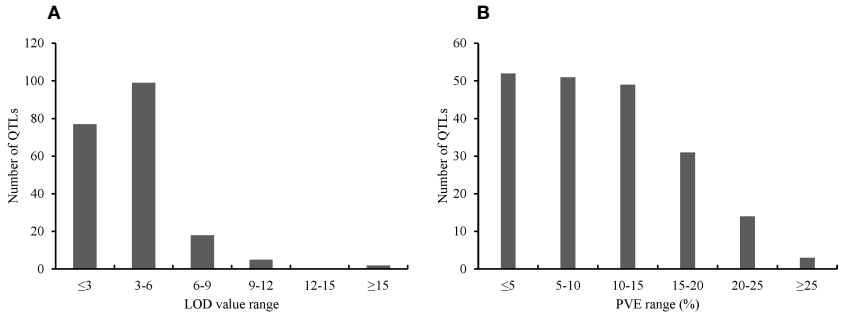


FIGURE 2
Features of the PUE QTLs used in the MQTL analysis. **(A)** Distribution of LOD scores **(B)** Distribution of PVE values of the initial PUE.

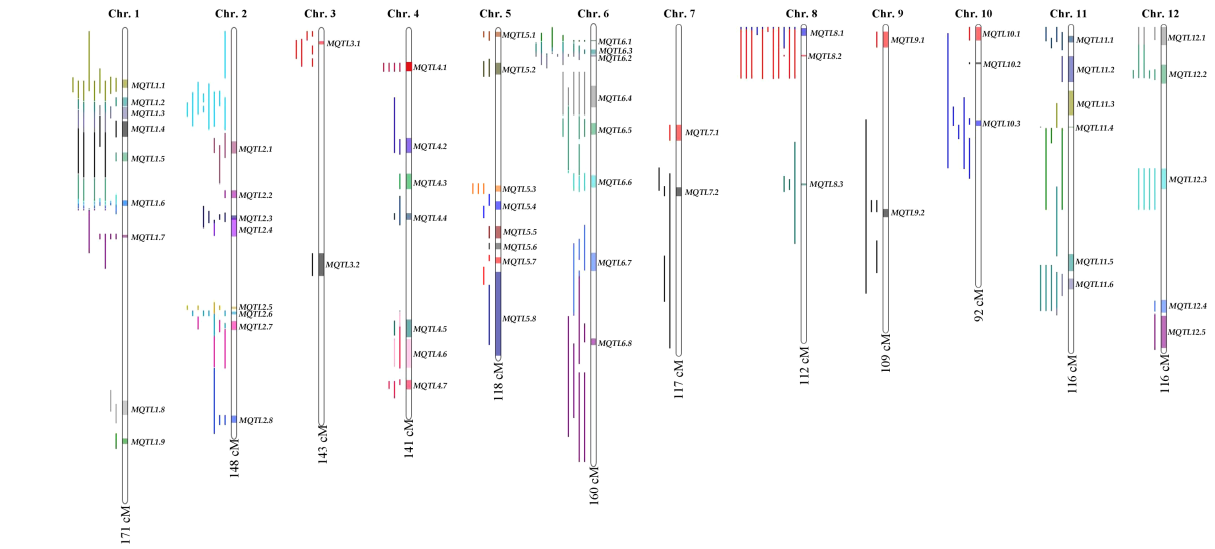


FIGURE 3
PUE MQTLs identified across 12 rice chromosomes.

MQTL1.7, MQTL2.1, MQTL3.1, MQTL4.5, MQTL5.1, MQTL6.1, MQTL6.2, MQTL6.3, MQTL8.2, MQTL8.3, MQTL10.2, MQTL10.3, and MQTL11.4, while the largest CI (7~9 cM) were observed on MQTL6.7, MQTL6.4, MQTL11.3, MQTL12.1, and MQTL12.2. The initial QTL members per MQTL varied from two to six per MQTL. MQTL1.5, MQTL1.6, MQTL1.7, and MQTL6.4 had the greatest number of initial QTL at n=6. Interestingly, MQTL2.1 had only one initial QTL (RRP). MQTL1.5, MQTL1.6, and MQTL1.7 had the colocalizing QTLs underlying BM, SCP, RSR, PUE, YLD, and RRP. Initial QTLs modulating BM, RRP, PUE, SPC, and YLD were all collocated on MQTL6.1, MQTL6.2, and MQTL6.3. YLD and BM were the most abundant traits in the MQTLs, present in 66% and 61% of the total MQTLs, respectively. In contrast, RSR was the least abundant trait, present in only 23% of the MQTLs.

TABLE 2 Number of initial QTLs and MQTLs identified on rice chromosomes.

Chromosome	No. of the Initial QTLs[a]	No. of MQTLs
1	25	9
2	32	8
3	7	2
4	18	7
5	14	8
6	33	5
7	5	2
8	15	3
9	5	2
10	8	3
11	18	6
12	12	5
Total	**192**	**60**

[a] **Initial QTLs:** QTLs used for MQTL analysis.

3.4 Identification of CGs underlying PUE in rice and functional analysis

The MQTLs harbored a total of 4,370 non-redundant genes. The number of genes per MQTL varied widely, from seven to 355, with MQTL11.4 (n=355) and MQTL12.1 (n=303) having the highest number of genes, while MQTL6.1 and MQTL8.2 having the fewest genes involved (seven and nine genes, respectively) (Table 3). After filtering for PUE-related terms, we identified 273 CGs (Table S1), of which 238 had root expression data available in the RiceXPro database (Table S2). Further analysis of CGs revealed that 209 genes had consistent direction of regulation under P non-supplied condition (184 upregulated, 25 downregulated) across the two time-points (6h and 24h) post-treatment (Table S3). The number of CGs had a strong positive correlation (R^2 = 64%) with CI (Figure S3), with exceptions on some MQTLs. In some MQTLs with small CI values, high number of PUE CGs were investigated, say, MQTL2.7 (CI=4.57 cM, n CGs=24), MQTL4.3 (CI=2.7 cM, n CGs=31), MQTL11.4 (CI=0.33 cM, n CGs=16). In contrast, MQTL4.1 (CI=6.6 cM, n CGs=4), MQTL10.1 (CI=4.9 cM, n CGs=1), MQTL11.3 (CI=9 cM, n CGs=7) had a relatively larger CI but fewer CGs.

To understand the biological pathways involved, we conducted gene ontology (GO) enrichment analysis using the 103 PUE CGs. Amino acid transmembrane transport, organic transport, and response to auxin were the most enriched biological pathways with five, five, and six gene members, respectively. Transcription, DNA templated, nucleic acid-templated transcription, and regulation of nucleobase-containing compound metabolic process

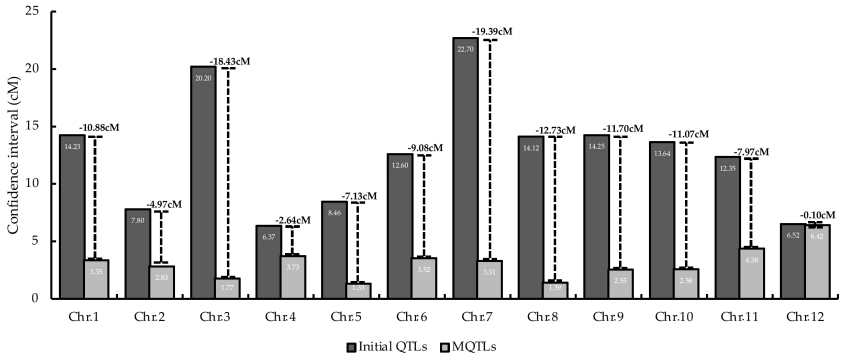


FIGURE 4
Comparison of average CI between initial the QTLs and MQTLs on the 12 rice chromosomes. Values on top of the lines represent the reduction in the average CI.

had the most gene members (20, 21, and 21, respectively) (Figure 5). To identify genes that are responsive to P treatment, we set a threshold of a 1.5-fold change in expression (P non-supplied condition vs. control), which narrowed down the list of CGs to 103 (Table 4), of which 86 were upregulated and 17 were downregulated (Figure 6).

3.5 Haplotype analysis and identification of potential donors

We performed haplotype analysis on the selected PUE CGs and identified potential donors for pyramiding the beneficial PUE CG alleles. We found 25 genes with at least a 2-fold change in expression levels at 24h post-treatment (Table S6). Among these, the eight genes (*OsARF8, OsSPX MFS3, OsRING141, OsMIOX, HsfC2b, OsFER2, OsWRKY64*, and *OsYUCCA11*) (Figure 7) had at least two haplotypes in the 3K RGP (Table 5) and were therefore used for further analysis. The number of synonymous SNPs ranged from three to nine (Table S7). *OsARF8*, *OsSPX-MFS3*, and *HSfC2b* had five SNPs, while *OsYUCCA11* had the most with nine SNPs. *OsRING141*, *OsMIOX*, *OsFER2, OsWRKY64* had four SNPs. The number of haplotypes ranged from two to six (Table S8). Four CGs had four haplotypes each (*OsARF8*, *OsMIOX*, *HSfC2b*, *OsFER2*, and *OsWRKY64*). Two CGs had six haplotypes each (*OsSPX-MF3* and *OsYUCCA11*). *OsRING141* had only two haplotypes.

We inferred the superior haplotypes from the beneficial allele donors in the original QTLs used in MQTL analysis. Kasalath and IR20, which were donors of the beneficial alleles in MQTLs from which the eight PUE CGs were located, had haplotype information in the 3K RGP. Thus, we used them to infer superior haplotypes. The Kasalath-types were regarded as the superior haplotype for *OsARF8* (Haplotype 2), *OsRING141* (Haplotype 4), *OsSPX-MFS3* (Haplotype 4), *and HSfC2b* (Haplotype 2). IR20-types were selected for *OsMIOX (Haplotype 1) OsFER2 (Haplotype 2), OsWRKY64 (Haplotype 2)*, and *OsYUCCA11 (Haplotype 2)* (Table S8). The abundance of the superior haplotypes was evaluated in the 3K RGP (Table S9; Figure 8). Superior haplotypes were found most abundant in the *indica* variety group. The frequency of superior alleles in *japonica* was considerably lower than that of the other subpopulations; in fact, superior haplotype for *OsSPX-MFS3* was completely absent. We identified rice accessions from the 3K RGP that cover the beneficial haplotypes for the eight CGs (Table 6). Four accessions were identified as the potential donors for the *indica* breeding programs, whereas two and three accessions were selected for the PUE breeding programs for *aus* and *japonica*, respectively. The population development using only Kasalath and IR20 can be suggested to pyramid the superior haplotypes of the eight PUE CGs.

4 Discussion

The development of rice varieties requiring less agricultural inputs is essential in the face of a changing climate and the rising cost of agricultural inputs such as fertilizers. Unlike N fertilizers, P fertilizer sources are estimated to last for only another four to five generations (Shimizu et al., 2004; Van Kauwenbergh, 2010). This is further complicated by the lack of plant-available P, despite fertilizer application due to the inherently low PUE in rice (Rose et al., 2011). It is therefore vital to utilize genetic variation in rice to address these issues. PUE-related QTLs and genes have been previously studied and mapped. However, their practical breeding utility has been hampered by their unstable effect across environments, genetic backgrounds, QTL to QTL or gene to gene interactions, and linkage drag caused by the large introgression size of QTLs that have not undergone fine-mapping. One example is the instability of the Kasalath-derived major PUE QTL *Pup1*, which confers tolerance to low P application as well as to mild drought. *Pup1* improves yield under low P in several *indica* varieties such as IR64, IR74, Situ Bagendit, Batur, and Dodokan under tropical conditions (Chin et al., 2011). However, *Pup1* introgression into Dasanbyeo, a Tongil-type *indica*, neither improved phosphorus uptake nor yield under low P application in temperate regions (Navea et al., 2022). In addition, a study has revealed the inability of *Pup1* to confer tolerance to low P in early shoot growth when introgressed simultaneously with the submergence tolerance major QTL *Sub1* into IR64 (Shin et al., 2021). The inconsistencies in QTL

TABLE 3 Summary of rice PUE MQTLs supported by multiple QTLs from at least two independent studies and with at least 5% average PVE values.

MQTL	Chromosome	AIC value (Model) [a]	No. of QTLs involved	Initial QTL average PVE (%)	Traits Involved [b]	Position (cM)	95% CI[c]. (cM)	Start (cM)	Stop (cM)	Left Marker	Right Marker	Physical Position (Mb)	Number of annotated genes underlying MQTL
MQTL1.1	1	230.18 (10)	11	12.3	YLD, BM, RRP, SPC, RSR, BM	19.51	3.10	17.96	21.06	SNP144359	SNP163177	4.49 – 5.17	85
MQTL1.2	1		7	13.0	BM, YLD, SPC, RSR	26.14	3.39	24.45	27.84	SNP193627	id1005232	6.13 – 6.9	104
MQTL1.3	1		7	12.5	PUP, YLD, BM, SPC, RSR	30.25	4.55	27.98	32.53	SNP222467	SNP245712	7.12 – 7.92	127
MQTL1.4	1		7	12.3	YLD, BM, SPC, RSR, IPT, BM	36.07	5.79	33.18	38.97	SNP255699	SNP298163	8.30 – 9.88	217
MQTL1.5	1		6	8.9	YLD, BM, SPC, RSR, IPT, RRP	46.33	3.29	44.69	47.98	SNP337593	SNP368391	11.16 – 11.94	82
MQTL1.6	1		7	8.3	BM, SPC, RSR, IPT, YLD, RRP	63.29	2.12	62.23	64.35	SNP513565	SNP528131	15.57 – 16.03	32
MQTL1.7	1		5	9.3	BM, SPC, RSR, IPT, YLD, RRP	75.62	1.20	75.02	76.22	SNP634609	SNP642207	18.89 – 19.09	22
MQTL2.1	2	278.41 (3)	3	9.5	RRP	44.31	1.09	43.77	44.86	id2005182	SNP1717430	11.00 – 11.23	24
MQTL2.7	2		7	11.7	SPC, BM, IPT, RT	105.31	4.57	103.03	107.60	id2011296	SNP2252638	25.77 – 26.88	159
MQTL3.1	3	35.98 (4)	2	9.3	BM, RT	0.88	1.77	0.02	1.79	SNP2491701	SNP2492869	0.40 – 0.47	8
MQTL4.1	4	124.61 (4)	2	9.1	RT, BM	31.98	6.60	28.68	35.28	SNP3933751	SNP3993698	7.17 – 8.84	118
MQTL4.3	4		3	8.5	RT, YLD	73.19	2.47	71.96	74.43	id4009413	id4010200	28.56 – 30.27	272
MQTL4.4	4		3	10.0	SDW, YLD	99.15	5.63	96.34	101.97	SNP4511040	SNP4543652	24.10 – 25.44	246
MQTL4.5	4		3	8.0	BM, IPT, YLD	127.50	0.20	127.40	127.60	SNP4715080	id4010985	31.78 – 32.04	46
MQTL5.1	5	144.34 (8)	3	7.0	SPC, RT	1.33	1.92	0.37	2.29	id5000043	SNP4821710	0.10 – 0.63	96
MQTL6.1	6	305.93 (5)	7	10.7	BM, RRP, IPT, SPC, YLD	5.63	0.50	5.38	5.88	id6000911	SNP5865517	1.38 – 1.42	9
MQTL6.2	6		9	6.5		9.62	1.71	8.77	10.48	SNP5883472	SNP5895767	2.13 – 2.61	108

(Continued)

10.3389/fpls.2023.1226297

TABLE 3 Continued

MQTL	Chromosome	AIC value (Model) [a]	No. of QTLs involved	Initial QTL average PVE (%)	Traits Involved [b]	Position (cM)	95% CI[c]. (cM)	Start (cM)	Stop (cM)	Left Marker	Right Marker	Physical Position (Mb)	Number of annotated genes underlying MQTL
					BM, RRP, IPT, SPC, YLD								
MQTL6.3	6		8	8.9	BM, RRP, IPT, SPC, YLD	10.98	0.82	10.57	11.39	SNP5895767	SNP5903052	2.61 – 2.87	47
MQTL6.4	6		6	20.4	SDW, RRP, IPT, SPC, YLD, RT	26.18	8.12	22.12	30.24	SNP5980679	SNP6051078	5.71 – 7.58	278
MQTL6.5	6		5	13.0	RT, YLD, BM	37.87	4.42	35.66	40.08	id6005608	SNP6147112	8.73 – 10.2	138
MQTL6.6	6		5	13.5	BM, RT	57.20	4.68	54.86	59.54	SNP6294233	SNP6351040	13.73 – 14.94	62
MQTL6.7	6		5	13.3	RRP, BM	86.61	6.91	83.16	90.07	SNP6613416	SNP6686237	20.83 – 22.52	179
MQTL6.8	6		7	14.2	RRP, BM, RT	116.04	2.61	114.74	117.35	SNP6889734	SNP6906770	28.69 – 29.33	95
MQTL7.2	7	34.17 (3)	3	9.3	PUP, SPC	58.54	3.31	56.89	60.20	SNP7468473	id7002392	14.11 – 15.01	57
MQTL8.1	8		11	8.5	YLD, SPC, BM, RSR, RT	1.87	2.80	0.47	3.27	id8001299	SNP8007342	0.12 – 0.82	126
MQTL8.2	8	95.18 (3)	11	15.1	YLD, SPC, BM, RSR, RT	10.43	0.68	10.09	10.77	id8002025	wd8001250	2.52 – 2.69	7
MQTL8.3	8		4	9.0	IPT, RRP	56.46	0.70	56.11	56.81	SNP8545780	SNP8550504	14.05 – 14.18	10
MQTL9.2	9	37.81 (4)	4	13.0	YLD, RRP	54.08	2.55	52.81	55.36	SNP9600918	SNP9623212	13.15 – 13.79	53
MQTL10.1	10		2	9.4	IPT, RT	2.45	4.90	0.00	4.90	SNP9898598	SNP9941068	0.15 – 1.01	70
MQTL10.2	10	48.96 (4)	3	11.5	IPT, RRP	13.07	0.78	12.68	13.46	id10000881	SNP10063204	3.02 – 3.4	21
MQTL10.3	10		7	9.2	RRP, IPT, SDW, RT	34.54	2.05	33.51	35.56	SNP10309364	id10002487	8.40 – 8.83	19
MQTL11.1	11		5	6.3	RT, YLD, IPT	4.52	2.48	3.28	5.76	SNP10840785	SNP10859595	0.81 – 1.42	94
MQTL11.3	11	148.94 (6)	4	14.0	IPT, RT, YLD	27.36	9.00	22.86	31.86	SNP10987441	SNP11075456	5.70 – 7.86	239
MQTL11.4	11		6	14.0	YLD, RT, RRP	36.00	0.33	35.84	36.17	id11005058	id11007108	15.45 – 19.35	355

(Continued)

10.3389/fpls.2023.1226297

10.3389/fpls.2023.1226297

TABLE 3 Continued

MQTL	Chromosome	AIC value (Model) [a]	No. of QTLs involved	Initial QTL average PVE (%)	Traits Involved [b]	Position (cM)	95% CI [c]. (cM)	Start (cM)	Stop (cM)	Left Marker	Right Marker	Physical Position (Mb)	Number of annotated genes underlying MQTL
MQTL11.5	11		4	9.8	YLD, IPT, BM, RT	84.47	6.17	81.39	87.56	SNP11608239	SNP11688144	17.69 – 19.18	202
MQTL11.6	11		6	14.4	YLD, IPT, BM, RT	92.03	3.90	90.08	93.98	SNP11730518	SNP11760343	21.49 – 21.84	77
MQTL12.1	12	95.86 (5)	4	6.0	YLD, RRP	2.33	8.45	0.00	6.56	SNP12006654	SNP12045914	0.12 – 1.64	303
MQTL12.2	12		6	11.0	BM, YLD, RRP	16.99	6.90	13.54	20.44	SNP12088643	id12002348	3.14 – 5.13	250

[a] For the model number in the parentheses of AIC value, see the materials and methods.
[b] **SDW**, shoot dry weight; **RT**, absolute value of root traits; **RSR**, root-shoot ratio; **BM**, total dry biomass; **SPC**, seed P content; **RRP**, relative response to P; **IPT**, internal P translocation; **YLD**, yield component.
[c] **CI**, Confidence interval.

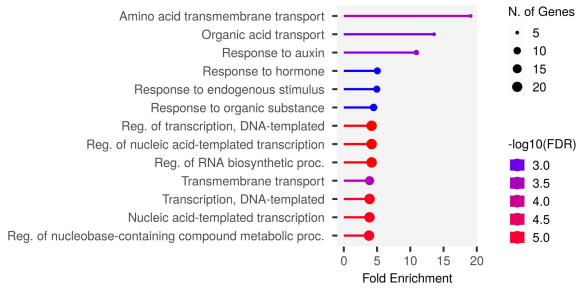


FIGURE 5
The most enriched biological pathways in the PUE CGs.

effects necessitate the identification of fine-mapped genomic regions related to PUE that are effective across diverse genetic backgrounds and environments.

We conducted an MQTL analysis on 192 PUE QTLs identified in 16 independent studies, in 15 non-overlapping bi-parental mapping populations. MQTL analysis offers the benefit of identifying reliable and robust genomic regions with reduced numbers of QTLs and narrower introgression sizes (Goffinet and Gerber, 2000; Kumari et al., 2021; Aloryi et al., 2022; Anilkumar et al., 2022; Joshi et al., 2023). To the best of our knowledge, our study is the first attempt to identify MQTLs underlying PUE in rice. Initially, we identified 60 PUE MQTLs and further selected 38 MQTLs that were supported by at least two independent studies and had PVE values of at least 5%. The majority of QTLs underlying these MQTLs had small PVE values ranging from 3% to 6%, implying that PUE in rice is controlled by many minor loci. This observation is consistent with results from several PUE QTL mapping studies (Navea et al., 2017; Wang et al., 2014). The distribution of the initial QTLs largely varied throughout the whole chromosome (Figure 1). However, it is not largely attributable to the SNP density (Figure S2). This is in contrast to the results from other MQTL studies in rice (Khahani et al., 2021; Selamat and Nadarajah, 2021; Aloryi et al., 2022; Joshi et al., 2023), wherein the uneven distribution of the initial QTLs mostly depended on the number of markers used in constructing the consensus map. Chromosome 6 was the chromosome with the largest number of initial PUE QTLs, as reported in the various previous PUE studies in rice (Ismail et al., 2007; Wissuwa et al., 2015; Heuer et al., 2017).

Linkage drag, due to the large CI of the QTLs utilized in breeding programs, is one of the hindrances in achieving successful in marker-assisted breeding (MAB) and genomic selection (GS) programs (Collard and Mackill, 2008; Kumar et al., 2020). Here, the average CI of the PUE MQTLs was reduced compared to their initial QTL members (Figure 4), except in case of chromosomes 4 and 12. MQTL analysis was not able to narrow down the CI on chromosomes 4 and 12 partly due to the small number of the initial QTL members as well as the already small CI of the QTLs. With the exceptions of the MQTLs detected on chromosomes 4 and 12, the average CI was reduced from 14.04 cM to 2.70 cM, suggesting the power of MQTL analysis in identifying precise genomic segments modulating a trait of

10.3389/fpls.2023.1226297

TABLE 4 P-responsive candidate genes underlying PUE MQTLs supported by multiple QTLs from at least two independent studies and with at least 5% average PVE values.

MQTL	MSU ID	RAP ID	Symbol	Description
MQTL1.1	*LOC_Os01g09550*	*Os01g0191300*	*ONAC003*	Similar to NAC-type transcription factor;NAC transcription factor, Drought resistance.
	LOC_Os01g09700	*Os01g0192900*	*OsACS5*	1-aminocyclopropane-1-carboxylic acid synthase, Submergence response.
	LOC_Os01g09850	*Os01g0195000*	*OsIDD2*	Zinc finger and indeterminate domain (IDD) family transcription factor, Regulation of secondary cell wall formation.
MQTL1.2	*LOC_Os01g11520*	*Os01g0213400*	*OsRFPH2-8*	Zinc finger, RING/FYVE/PHD-type domain containing protein.
	LOC_Os01g12260	*Os01g0222500*	*OsVHA-E3*	Similar to Vacuolar ATP synthase subunit E (EC 3.6.3.14) (V-ATPase E subunit) (Vacuolar proton pump E subunit).
	LOC_Os01g12400	*Os01g0223700*	*SDRLK-68*	S-Domain receptor like kinase-68, Partial S-domain containing protein, Response to drought in tolerant rice genotypes
	LOC_Os01g12440	*Os01g0224100*	*OsERF#053*	Similar to DNA binding protein-like protein; Similar to DNA binding protein-like protein.
MQTL1.3	*LOC_Os01g13030*	*Os01g0231000*	*OsIAA3*	Similar to Auxin-responsive protein (Aux/IAA) (Fragment). Similar to Auxin-responsive protein (Aux/IAA) (Fragment).
	LOC_Os01g13770	*Os01g0239200*	*OsTPT1*	Triose phosphate/phosphate translocator
MQTL1.4	*LOC_Os01g14870*	*Os01g0252200*	*OsC3H3*	Zinc finger, CCCH-type domain containing protein.
	LOC_Os01g15300	*Os01g0256800*	*OsC3H4*	Similar to zinc finger helicase family protein. Hypothetical conserved gene.
	LOC_Os01g16260	*Os01g0268100*	*OsZIFL1*	Similar to Major facilitator superfamily antiporter.
	LOC_Os01g16940	*Os01g0276500*	–	Similar to Histidine biosynthesis bifunctional protein hisIE, chloroplast precursor; Phosphoribosyl-ATP pyrophosphatase (EC 3.6.1.31) (PRA-PH)].
	LOC_Os01g17240	*Os01g0279700*	*OsPT21*	Phosphate transporter 4;1;Major facilitator superfamily protein.
MQTL1.5	*LOC_Os01g20930*	*Os01g0311400*	*OsRING108*	Zinc finger, *RING/FYVE/PHD*-type domain containing protein.
	LOC_Os01g20940	*Os01g0311500*	*OsPHS1a*	Similar to *PHS1* (PROPYZAMIDE-HYPERSENSITIVE 1); phosphoprotein phosphatase/protein tyrosine/serine/threonine phosphatase.
	LOC_Os01g21120	*Os01g0313300*	*OsERF#068*	Similar to *EREBP-3* protein (Fragment).
MQTL2.7	*LOC_Os02g41800*	*Os02g0628600*	*OsARF8*	Similar to Auxin response factor 8.
	LOC_Os02g42690	*Os02g0639800*	*OsRDCP2*	Hypothetical gene. Zinc finger, *RING/FYVE/PHD*-type domain containing protein.
	LOC_Os02g42990	*Os02g0643800*	*OsSAUR11*	Auxin-responsive protein, Negative regulation of deep sowing tolerance, Mesocotyl elongation.
	LOC_Os02g43170	*Os02g0646200*	*OsBBX6*	Zinc finger, B-box domain containing protein.
	LOC_Os02g43620	*Os02g0652800*	*OsGlpT1*	Major facilitator superfamily *MFS*_1 protein.
	LOC_Os02g43790	*Os02g0654700*	*OsERF#091*	*AP2/ERF* family protein, Abiotic stress response.
	LOC_Os02g43820	*Os02g0655200*	*OsERF#095*	Similar to Ap25. ERF family protein, Transcriptional regulator, Salt stress tolerance.
	LOC_Os02g43940	*Os02g0656600*	*OsERF#032*	Similar to Dehydration responsive element binding protein 2B (DREB2B protein).
	LOC_Os02g44090	*Os02g0658200*	–	Zinc finger, PHD-type domain containing protein.
	LOC_Os02g44120	*Os02g0659100*	*OsDLN62*	Zinc finger, C2H2-type domain containing protein.
MQTL4.4	*LOC_Os04g41350*	*Os04g0490900*	*OsAAP11G*	Similar to *OSIGBa0130B08.4* protein.
	LOC_Os04g42090	*Os04g0498600*	*SamDC*	S-adenosylmethionine decarboxylase, Polyamine biosynthesis, Salt and drought stresses, Abiotic stress.
	LOC_Os04g42570	*Os04g0504500*	*OsPLT4*	Similar to protein *BABY BOOM 1*.
MQTL4.3	*LOC_Os04g48050*	*Os04g0568900*	*OsRINGzf1*	RING zinc finger protein, E3 ubiquitin ligase, Regulation of drought resistance.
	LOC_Os04g48170	*Os04g0570000*	*OsCYP87A3*	Cytochrome P450 87A3, Auxin signaling in the regulation of coleoptile growth.
	LOC_Os04g48410	*Os04g0573200*	*SOD*	Copper chaperone for superoxide dismutase, Target of miR398b, Resistance to rice blast disease; Similar to OSIGBa0147H17.7 protein.
	LOC_Os04g49000	*Os04g0579200*	*OsRING328*	Zinc finger, *RING/FYVE/PHD*-type domain containing protein.

(Continued)

10.3389/fpls.2023.1226297

TABLE 4 Continued

MQTL	MSU ID	RAP ID	Symbol	Description
	LOC_Os04g49410	*Os04g0583500*	*OsEXPA10*	Expansin, Al-inducible expansin, Root cell elongation;Similar to Expansin-A10.
	LOC_Os04g49510	*Os04g0584600*	*OsCDPK13*	Similar to Calcium dependent protein kinase. Group I calcium-dependent protein kinase, Cold and salt/drought tolerance.
	LOC_Os04g49620	*Os04g0585700*	*OsFLZ12*	Protein of unknown function DUF581 family protein.
	LOC_Os04g49650	*Os04g0585900*	*OsFLZ13*	Protein of unknown function DUF581 family protein. Protein of unknown function DUF581 family protein.
	LOC_Os04g49660	*Os04g0586000*	*OsFLZ14*	Protein of unknown function DUF581 family protein.
	LOC_Os04g49670	*Os04g0586100*	*OsFLZ15*	Protein of unknown function DUF581 family protein.
	LOC_Os04g49680	*Os04g0586200*	*OsFLZ16*	Similar to H0307D04.13 protein.
	LOC_Os04g51890	*Os04g0608300*	*OsSAUR20*	Auxin responsive SAUR protein domain containing protein.
	LOC_Os04g52900	*Os04g0620000*	*OsABCC1*	C-type ATP-binding cassette (ABC) transporter, Arsenic (As) detoxification, Reduction of As in grains.
MQTL4.5	*LOC_Os04g53612*	*Os04g0628000*	*OsISC40*	Protein of unknown function DUF794, plant family protein.
MQTL5.1	*LOC_Os05g01610*	*Os05g0106700*	*OsPRAF2*	Similar to PRAF1; Ran GTPase binding/chromatin binding/zinc ion binding.
	LOC_Os05g01990	*Os05g0110500*	*OsRH17*	Similar to DEAD-box ATP-dependent RNA helicase 17. DEAD-box RNA helicase protein, Stress responses.
	LOC_Os05g02050	*Os05g0111100*	–	Zinc finger, Tim10/DDP-type family protein.
MQTL6.1	*LOC_Os06g03860*	*Os06g0129400*	*OsSPX-MFS3*	Splicing variant of *SPX-MFS* protein 3. Vacuolar phosphate efflux transporter, Pi homeostasis.
MQTL6.2	*LOC_Os06g04920*	*Os06g0141200*	*OsZFP1*	Putative zinc finger protein, Negative regulation of salt stress response.
	LOC_Os06g05110	*Os06g0143000*	*SodB*	Iron-superoxide dismutase;Splicing variant of the iron-superoxide dismutase.
	LOC_Os06g05160	*Os06g0143700*	*OsSultr3;4*	*SULTR*-like phosphorus distribution transporter, Control of the allocation of phosphorus to the grain.
MQTL6.4	*LOC_Os06g11450*	*Os06g0218300*	*OsRING342*	Zinc finger, *RING*-type domain containing protein.
	LOC_Os06g11860	*Os06g0222400*	*OsERF#120*	Similar to *DRE*-binding protein 2.
	LOC_Os06g11980	*Os06g0223700*	*OsFLZ20*	*FCS*-like zinc finger (FLZ) protein 20, Submergence response.
	LOC_Os06g12160	*Os06g0225900*	–	Similar to ATP binding/ATPase/nucleoside-triphosphatase/nucleotide binding.
	LOC_Os06g12370	*Os06g0229000*	*OsFtsH6*	Similar to FtsH protease (VAR2) (Zinc dependent protease).
	LOC_Os06g12610	*Os06g0232300*	*PIN1C*	PIN protein, Auxin efflux carrier, Auxin transport and signaling, "Root, shoot and inflorescence development".
	LOC_Os06g12640	*Os06g0232700*	–	Similar to SWIM zinc finger family protein.
MQTL6.5	*LOC_Os06g16060*	*Os06g0271600*	*OsRING141*	Zinc finger, *RING/FYVE/PHD*-type domain containing protein.
	LOC_Os06g17280	*Os06g0283200*	*OsRFP*	Zinc finger, *RING/FYVE/PHD*-type domain containing protein.
	LOC_Os06g17410	*Os06g0284500*	*OsDof20*	Zinc finger, Dof-type family protein.
MQTL6.6	*LOC_Os06g23530*	*Os06g0343100*	–	Similar to ATP-dependent helicase DHX8 (RNA helicase HRH1) (DEAH-box protein 8). Similar to predicted protein.
	LOC_Os06g24850	*Os06g0355300*	*OsIAA22*	Similar to Auxin-responsive protein IAA22.
MQTL6.7	*LOC_Os06g35960*	*Os06g0553100*	*HSfC2b*	Similar to Heat stress transcription factor C-2b.
	LOC_Os06g36210	*Os06g0556200*	*OsAAP12B*	Similar to amino acid permease 1.
	LOC_Os06g36560	*Os06g0561000*	*OsMIOX*	Myo-inositol oxygenase, Drought stress tolerance.
	LOC_Os06g37450	*Os06g0571800*	*OsGATA16*	Similar to GATA transcription factor 20. Similar to GATA transcription factor 3 (AtGATA-3).
	LOC_Os06g37750	*Os06g0575400*	*SDRLK-5*	S-Domain receptor like kinase-5, Response to drought in a tolerant genotype, Response to submergence.

(Continued)

10.3389/fpls.2023.1226297

TABLE 4 Continued

MQTL	MSU ID	RAP ID	Symbol	Description
MQTL6.8	*LOC_Os06g47290*	*Os06g0687400*	–	Similar to auxin-independent growth promoter-like protein. (Os06t0687400-01);Similar to auxin-independent growth promoter-like protein.
	LOC_Os06g47590	*Os06g0691100*	*OsERF#121*	Pathogenesis-related transcriptional factor and ERF domain containing protein.
	LOC_Os06g47840	*Os06g0693500*	–	Zinc finger, C2H2-like domain containing protein.
MQTL8.2	*LOC_Os08g05030*	*Os08g0145600*	–	Similar to cDNA clone:J023091L02, full insert sequence.
	LOC_Os08g23410	*Os08g0323400*	*OsISC42*	Similar to Rubredoxin (Rd).
MQTL9.1	*LOC_Os09g23140*	*Os09g0394600*	–	Endonuclease/exonuclease/phosphatase domain containing protein.
MQTL10.2	*LOC_Os10g06030*	*Os10g0151100*	*OsWAK103*	Similar to Protein kinase domain containing protein, expressed.
MQTL10.3	*LOC_Os10g16974*	*Os10g0317900*	*OsCYP75B4*	Chrysoeriol 5'-Hydroxylase, Flavonoid B-ring hydroxylase, Tricin biosynthesis; Similar to Flavonoid 3-monooxygenase.
MQTL11.1	*LOC_Os11g03420*	*Os11g0128300*	*OsMIF1*	Mini zinc finger protein, A member of the *ZF-HD* (zinc finger-homeodomain) family, Negative regulation of deep sowing tolerance, Mesocotyl elongation.
MQTL11.3	*LOC_Os11g10590*	*Os11g0211800*	*OsDT11*	Cysteine-rich peptide, Short-chain peptide, ABA-dependent drought tolerance.
MQTL11.4	*LOC_Os11g29870*	*Os11g0490900*	*OsWRKY72*	*WRKY* transcription factor 72, ABA response with respect to germination and abiotic stresses, ABA signaling and auxin transport
	LOC_Os11g30484	*Os11g0498400*	–	Zinc finger, C2H2-like domain containing protein.
	LOC_Os11g30560	*Os11g0499600*	*drp7*	Hydroxysteroid dehydrogenase, Cuticle formation, Lipid homeostasis, Submergence tolerance.
	LOC_Os11g31340	*Os11g0512100*	*ONAC127*	NAC (NAM, ATAF1/2, CUC2) transcription factor, Heat stress response, Regulation of grain filling.
	LOC_Os11g32100	*Os11g0523700*	*OsbHLH002*	bHLH transcription factor, Positive regulation of chilling tolerance, Control of stomatal initiation, Regulation of mature stoma differentiation.
	LOC_Os11g32110	*Os11g0523800*	*OsARF1*	Similar to Isoform 3 of Auxin response factor 23. Auxin response factor 1.
	LOC_Os11g32290	*Os11g0525900*	–	Zinc finger, GRF-type domain containing protein.
MQTL11.6	*LOC_Os11g36480*	*Os11g0573200*	–	Similar to Zinc knuckle family protein, expressed.
	LOC_Os11g36960	*Os11g0578100*	*OsDjC76*	Heat shock protein DnaJ, N-terminal domain containing protein.
	LOC_Os11g37000	*Os11g0578500*	*OsDjC77*	Heat shock protein DnaJ family protein.
MQTL12.1	*LOC_Os12g01530*	*Os12g0106000*	*OsFER2*	Ferritin, Iron storage protein, Iron homeostasis.
	LOC_Os12g02100	*Os12g0112300*	–	ADP/ATP carrier protein domain containing protein.
	LOC_Os12g02450	*Os12g0116700*	*OsWRKY64*	*WRKY* transcription factor 64, Response to the rice pathogens, Regulation of root elongation under iron excess, Iron stress tolerance.
	LOC_Os12g02980	*Os12g0123500*	–	Similar to Apyrase precursor (EC 3.6.1.5) (ATP-diphosphatase) (Adenosine diphosphatase) (ADPase) (ATP-diphosphohydrolase).
	LOC_Os12g03670	*Os12g0130500*	*SDRLP-6*	S-Domain receptor like protein-6, Response to submergence.
	LOC_Os12g03830	*Os12g0132500*	*OsZIFL9*	Similar to Major facilitator superfamily antiporter.
	LOC_Os12g03950	*Os12g0133300*	*OsZIFL13*	Similar to Carbohydrate transporter/sugar porter/transporter.
MQTL12.2	*LOC_Os12g07700*	*Os12g0176200*	*OsISC14*	Similar to Nitrogen fixation like protein.
	LOC_Os12g08070	*Os12g0181300*	–	Similar to TRAF-type zinc finger family protein.
	LOC_Os12g08090	*Os12g0181500*	*OsAAP11A*	Amino acid permease, Transport of amino acids.
	LOC_Os12g08130	*Os12g0181600*	*OsAAP11B*	Amino acid transporter, transmembrane domain containing protein.
	LOC_Os12g08780	*Os12g0189500*	*OsYUCCA11*	Flavin-containing monooxygenase, Auxin biosynthesis, Endosperm development, Regulation of grain filling.
	LOC_Os12g08820	*Os12g0190100*	–	Similar to Auxin-independent growth promoter-like protein.
	LOC_Os12g09300	*Os12g0194900*	*OsAAP10B*	Amino acid permease, A member of the amino acid transporter (AAT) family, Regulation of tillering and grain yield, Regulation of neutral amino acid transport
	LOC_Os12g09590	*Os12g0197700*	–	Region of unknown function, putative Zinc finger, XS and XH domain containing protein.

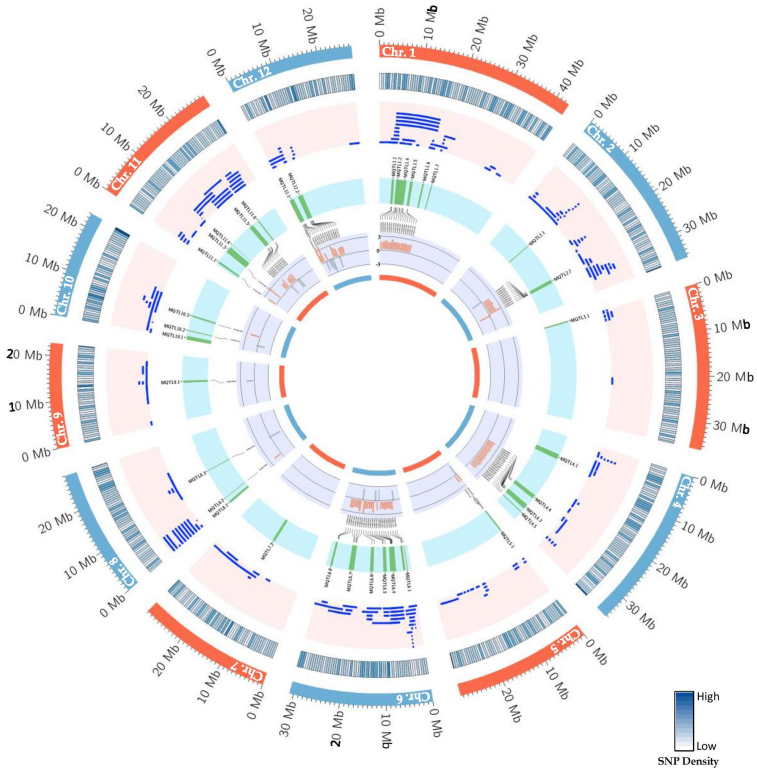


FIGURE 6
Schematic representation of the distribution patterns of PUE QTLs, MQTLs, and PUE CGs on rice chromosomes. The outermost circle represents the rice karyotype. The second circle outlines the consensus map SNP density (500kb window). The third circle display the initial PUE QTLs used in the MQTL analysis. The fourth inner circle represents the MQTLs supported by QTLs from at least 2 independent studies and with a PVE value of ≥ 5%. The fifth and sixth inner circle represent the P-responsive CGs and the fold change in gene expression (P non-supplied vs. control treatment), respectively.

interest. Our results are consistent with the observations in previous MQTL studies on various traits in cereal crops (Chen et al., 2017; Venske et al., 2019; Soriano et al., 2021; Miao et al., 2022).

The successful integration of pleiotropic QTLs into breeding programs has been documented in rice (Ookawa et al., 2010; Yan et al., 2011; Vishnukiran et al., 2020). In this study, we categorized QTLs identified under field conditions, based on traits well-distributed throughout the rice growth stages. These included seedling stage (RT, RSR, RRP), vegetative stage (SDW, BM, IPT), and reproductive stage traits (SPC and YLD). This enabled us to identify PUE MQTLs that may have pleiotropic effects, as observed in MQTL1.5, MQTL1.6, and MQTL1.7, on multiple traits including BM, SCP, RSR, IPT, YLD, and RRP. This can aid in a more efficient improvement of PUE in rice through the accumulation of beneficial PUE MQTL alleles. Among the QTLs used in this study, genomic regions regulating root traits under P deficient conditions were the least abundant across rice chromosomes, which can be attributed to the challenges in establishing reliable and high-throughput root phenotyping techniques under field conditions (Heuer et al., 2017), as well as the lack of genetic diversity (Ismail et al., 2007). It is worth noting that the number of PUE CGs underlying MQTL2.7, MQTL4.3, and MQTL11.4 are relatively large despite their narrow CI. This is unlike the observations in previous MQTL studies in rice, wherein the number of CGs underlying an MQTL had a strong positive correlation with the size of the CI (Daware et al., 2017; Islam et al., 2019; Khahani et al., 2021; Selamat and Nadarajah, 2021; Aloryi et al., 2022; Anilkumar et al., 2022; Joshi et al., 2023). These MQTLs have the potential to be effective genomic targets for use in the MAB after being validated in a wide range of genetic backgrounds and environments by utilizing the aforementioned linked flanking markers (Table 3).

We identified CGs underlying the PUE MQTLs using GO terms directly related to PUE traits, as well as for the secondary PUE traits. We further selected genes that could be validated through the root expression data under control and P deficient/non-supplied conditions. Most of the PUE CGs (83%) were upregulated (Figure 6F), implying that genes underlying the PUE MQTLs confer an active response to nutrient deficiency, rather than conservation of resources. This agrees with the pattern of gene regulation under nutrient deficiency in previous studies in various plants (López-Bucio et al., 2003; Chen et al., 2022; Wang M et al., 2019). The results of our GO analysis give an insight into the complexity of PUE in rice. CGs underlying PUE MQTLs were heavily enriched with genes involved in amino acid transmembrane transport, organic acid transport, and response to auxin. Amino acid transporters in rice, although not reported to be directly involved in PUE, have been associated with the regulation of flowering time and defense against abiotic stresses and pathogen attack (Guo et al., 2021). In particular, the genes involved in the amino acid transport pathways function in various plant species

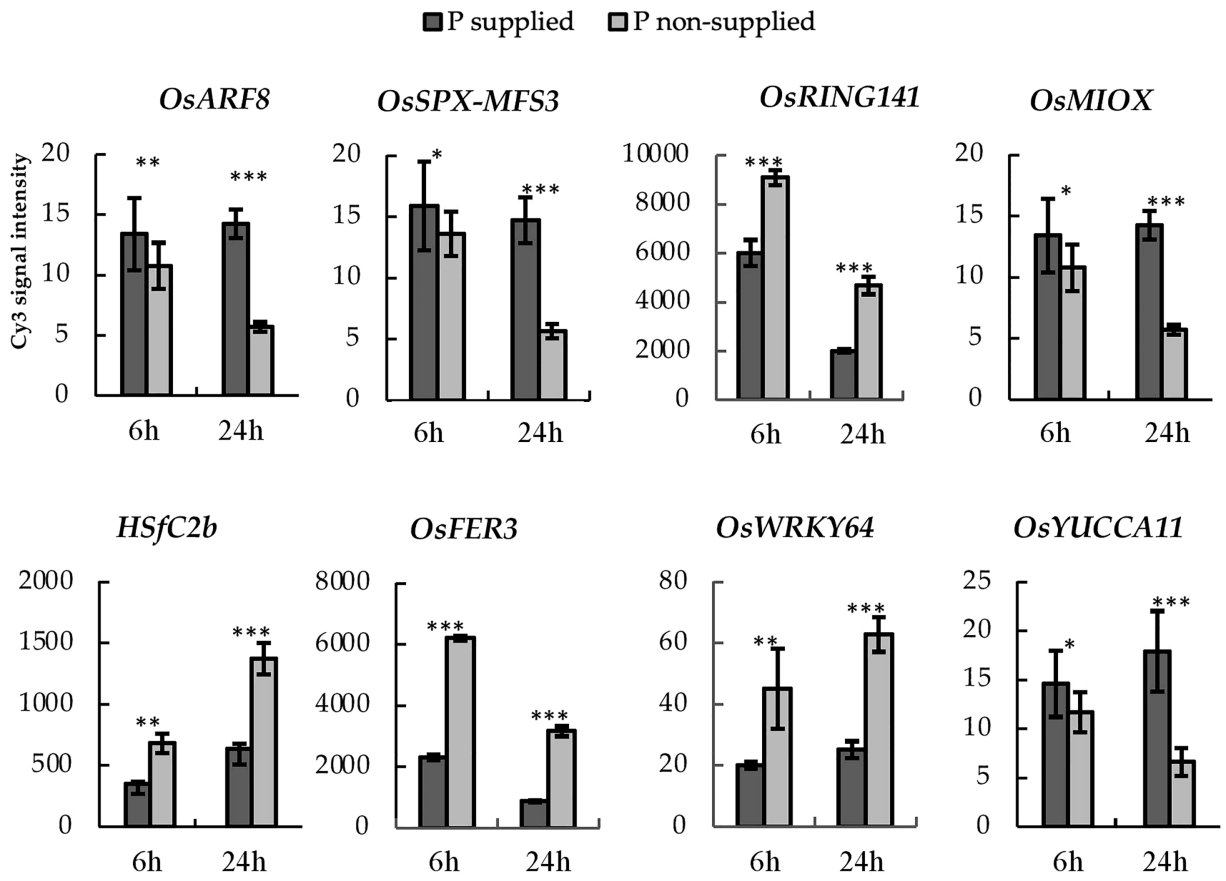


FIGURE 7
PUE-MQTL candidate genes showing significant responses to P application at 6- and 24-hours post-treatment. Asterisks denote significance using two-tailed t-test, *P<0.10; **P<0.05, ***P<0.01.

defending against both biotic and abiotic stresses through the regulation of the salicylic acid pathway (Kan et al., 2017), including drought, salinity, UV radiation, heavy metals, and pathogens (Szabados and Savouré, 2010). In plants, organic acid transporters are upregulated by Al and/or P deficiency (Yu et al., 2016). This is caused by enhanced phosphorylation levels of the plasma membrane H+-ATPase, which creates an electrochemical potential across the cell membrane. This leads to an increase in the activity of the organic acid transporters and the passive release of organic anions from root tips. Ultimately, this process causes organic

TABLE 5 PUE-related candidate genes used for the haplotype analysis in the rice 3K genome panel.

MSU ID	RAP ID	MQTL	Description/function	Gene Symbol	Reference
LOC_Os02g41800	*Os02g0628600*	*MQTL2.7*	Similar to auxin response factor 8	*OsARF8*	–
LOC_Os06g03860	*Os06g0129400*	*MQTL6.1*	Splicing variant of SPX-MFS protein 3, vacuolar phosphate efflux transporter, Pi homeostasis	*OsSPX-MFS3*	Wang et al., 2015
LOC_Os06g16060	*Os06g0271600*	*MQTL6.5*	RING-type E3 ubiquitin ligase 141; Zinc finger, PHD-type domain containing protein	*OsRING141*	Park et al., 2019
LOC_Os06g36560	*Os06g0561000*	*MQTL6.7*	Myo-inositol oxygenase, drought stress tolerance	*OsMIOX*	Duan et al., 2012
LOC_Os06g35960	*Os06g0553100*	*MQTL6.7*	Similar to heat stress transcription factor C-2b.	*HSfC2b*	Xiang et al., 2013
LOC_Os12g01530	*Os12g0106000*	*MQTL12.1*	Ferritin 2, iron storage protein, iron homeostasis	*OsFER2*	Paul et al., 2012
LOC_Os12g02450	*Os12g0116700*	*MQTL12.1*	Similar to *WRKY* transcription factor 64, response to the rice pathogens; iron stress tolerance	*OsWRKY64*	Viana et al., 2017
LOC_Os12g08780	*Os12g0189500*	*MQTL12.2*	Flavin-containing monooxygenase, auxin biosynthesis, Endosperm development, regulation of grain filling	*OsYUCCA11*	Xu et al., 2021

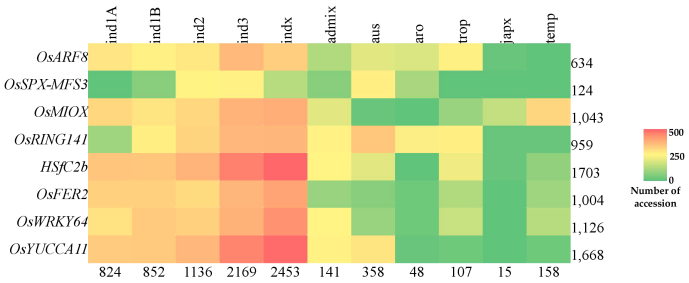


FIGURE 8
Abundance of superior PUE CG haplotypes in the rice 3K rice genome panel. Values indicate the number of accessions.

acids to exude from the roots and form a stable complex with Al, which allows P to become soluble for plant assimilation (Ligaba et al., 2004). Plant root architecture undergoes adaptive changes, including the inhibition of primary root growth and the increase in the number and length of lateral roots (Péret et al., 2011) modulated by the change in sensitivity of auxin receptors such as *TIR1*, under a P deficient condition (Wu et al., 2023). Upregulation of auxin receptors degrades auxin repressors, releasing auxin response factor, *ARF19*, which then leads to the activation of genes related to lateral root morphogenesis (Pérez-Torres et al., 2008).

We further narrowed down the PUE CGs to eight genes and investigated their natural variation in a diverse set of rice (3K RGP). These genes included *OsARF8*, *OsSPX-MFS3*, *OsRING141*, *OsMIOX*, *HsfC2b*, *OsFER2*, *OsWRKY64*, and *OsYUCCA11* (Figure 7). *OsSPX-MFS3* is a member of the rice SPX-MFS family which mediates Pi transport between the cytosol and vacuole (Yang et al., 2017). *OsSPX-MFS3* plays a major role in the transport of Pi from the cytosol to vacuole (Guo et al., 2023) and is downregulated under P deficiency. The regulation pattern of *OsSPX-MFS3* is consistent with the observation in the microarray data utilized in this study. It is worth noting that, except for *OsSPX-MFS3*, the rest of the CGs did not have GO terms directly related to PUE but were nevertheless associated to secondary PUE traits and other abiotic stresses. Auxin is one of the phytohormones that regulate root architecture modifications in response to Pi deficiency (Nacry et al., 2005). In this study, we identified two CGs, namely *OsARF8* and *OsYUCCA11*, implicated in auxin response and which were significantly downregulated under P non-supplied condition. *OsARF8* is a member of the ARF family implicated in the crosstalk between auxin signaling and P status (Wang S et al., 2014).

Genes under the zinc-finger family may play an important role in the regulation tolerance to multiple stresses in rice (Deng et al., 2018). The present study identified a P-deficiency-upregulated zinc-finger gene, *OsRING141*, among the genes subtending MQTL6.5. Similarly, previous studies found two C2H2- type zinc finger protein genes, ZOS3-12 and ZOS5-08, to be responsive to P and N deficiencies (Huang et al., 2012). MQTL12.1 harbors two genes modulating iron (Fe) homeostasis (*OsFER2*) and iron stress tolerance (*OsWRKY64*). Both genes were significantly upregulated in the P non-supplied condition. Previous studies have shown that the reduction of Fe concentration can lead to the recovery of primary root elongation under low P conditions, suggesting that Fe may play a role in the Pi deficiency-induced reduction of primary root growth (PRG) (Ward et al., 2008). Moreover, research on Arabidopsis has demonstrated that when Fe is deposited in the root tip meristem, it can trigger the accumulation of reactive oxygen species (ROS) and callose, likely through LPR1-dependent redox signaling (Müller et al., 2015). The accumulation of ROS and callose can interfere with cell-to-cell communication that is essential for maintaining the stem cell niche and inhibiting PRG (Müller et al., 2015).

MQTL6.7 harbors two abiotic stress-related genes, namely *OsMIOX* (drought) and *HSfC2b* (heat). A previous study reported that the overexpression of *OsMIOX* in transgenic rice greatly improved growth performance under drought conditions by decreasing oxidative damage (Duan et al., 2012). Severe phosphorus deficiency can lead to changes in the photosynthetic apparatus, such as decreased rates of carbon dioxide assimilation, reduced expression of photosynthesis-related genes, and photoinhibition at the photosystem II level. These changes can potentially cause photo-oxidative stress, which can damage the plant's cells. By preventing oxidative damage, the plant may be able to better tolerate the effects of phosphorus deficiency on its photosynthetic apparatus (Hernández and Munné-Bosch, 2015). Therefore, as *OsMIOX* may prevent oxidative damage, it might be beneficial to phosphorus-starved plants. A heat shock factor (HSF) member, *HSfC2b*, harbored within MQTL6.7, was upregulated under P deficient condition. The first attempt to elucidate the genome wide expression of these HSFs was conducted by Chauhan et al. (2011). HSF encoding genes showed significant upregulation in abiotic stresses such as cold, drought, and salinity. Our study is the first to show the link between an HSF gene and PUE in rice. In Arabidopsis, the overexpression of HSFs in transgenic plants confer simultaneous tolerances to multiple abiotic stresses such as heat and anoxia (Banti et al., 2010). Nevertheless, further study is needed to elucidate the relationship between HSF and PUE in rice.

The use of NGS-based genotyping methods have aided in the identification of SNPs associated with agronomic traits in rice. However, the applicability of SNPs in breeding programs is constrained by their bi-allelic nature by cross breeding, the presence of uncommon alleles, and the abundance of linkage drag

(Annicchiarico et al., 2017). Considering the gene haplotypes for genome-wide analysis will help overcome SNP marker's limitation. Haplotypes are specific combinations of jointly inherited nucleotides or DNA markers from polymorphic sites in the same chromosomal segment (Stephens et al., 2001; Lu et al., 2010). Haplotype-based breeding holds a promise in accumulating beneficial alleles for a trait of interest in living organisms. This can be achieved through the identification of haplotypes for different genes and utilizing them through techniques such as allele mining, pyramiding, or GS. This approach has resulted in genetic gains in crops (Abbai et al., 2019; Anandan et al., 2022; Du et al., 2022). We identified subspecific-wise potential donors of the superior haplotypes for PUE CGs: *OsARF8*, *OsSPX-MFS3*, *OsRING141*, *OsMIOX*, *HsfC2b*, *OsFER2*, *OsWRKY64*, and *OsYUCCA11* (Table 6). We inferred the superior haplotypes for these genes based on the haplotype of beneficial allele donors (Kasalath and IR20) in the initial QTL studies used for the meta-QTL analysis. The frequencies of superior haplotypes differed among sub-populations, with particular groups exhibiting a higher prevalence (Figure 8; Table S9). Similarly, previous studies have suggested that the distribution of haplotypes is influenced by evolutionary and population genetic factors, such as rates of mutation and recombination, as well as selection pressures (Magwa et al., 2016; Sinha et al., 2020). Superior haplotypes for almost all the CGs were predominant in *indica* subspecies. The frequency of the superior haplotypes in *japonica* was considerably smaller, compared to that of other subspecies. As mentioned, *OsSPX-MFS3* was completely absent in *japonica* varieties in 3K RGP. This implies a big opportunity to improve PUE in the *japonica* varieties using the donors identified in this study. Our results suggest that *indica* varieties are a rich source of PUE CG superior haplotypes and could be utilized in the genomics-assisted breeding programs for PUE in rice. In the case of the *japonica*, it is necessary to generate pre-breeding lines that encompass the superior PUE CG haplotypes.

We identified potential CGs associated with PUE as well as superior haplotypes that could be used to accumulate beneficial alleles to improve PUE in rice. However, it should be noted that the PUE CG mining pipeline used in the present study was limited by the availability of P-supplied and P-non-supplied root transcriptome data. Initially, we identified 273 PUE genes (Table S1), but we could only analyze 238 CGs (Table S2) that had gene expression data in the RiceXPro database. Therefore, a complete set of transcriptome data would provide offer higher precision in the identification of PUE CGs in rice. Another limitation of our pipeline is the limited scope of the reference genome used in both the annotation of genes and the root expression analysis. We utilized gene annotation from the IRGSP v. 1.0 (Nipponbare) reference genome, a widely used reference genome for rice research, to identify PUE CGs. Here, it was not possible for us to identify genes that were absent in the reference genome and therefore were possibly neglected in our analysis. For instance, the P uptake gene *OsPSTOL1* harbored in the P uptake major QTL *Pup1* is absent in the reference genome Nipponbare, as well as in the genomes of several other commonly used rice varieties (Chin et al., 2011; Gamuyao et al., 2012). Similarly, *Sub1A*, a gene conferring submergence tolerance in rice, was not found in the fully sequenced genome of Nipponbare (Xu et al., 2006). The limited

TABLE 6 Suggested donors based on a haplotype-based selection for the pyramiding of favorable PUE CG alleles.

Subspecies or varietal group	Line	*OsARF8*	*OsSPX-MFS3*	*OsRING141*	*OsMIOX*	*HSfC2b*	*OsFER2*	*OsWRKY64*	*OsYUCCA11*
Superior haplotype controls	IR20	+		+			+	+	+
	Kasalath	+	+		+	+	+		
indica	IRGC 132300	+	+	+	+	+	+	+	+
	IRGC 132300	+	+	+	+	+	+	+	+
	IRIS 313-9415	+	+	+	+	+	+	+	+
	IRGC 126983	+	+	+	+	+	+	+	+
aus	IRGC 127181	+	+		+	+			+
	IRGC 135811			+	+		+	+	
japonica	IRGC 128463	+			+	+			
	IRGC 135811			+	+	+		+	+
	IRGC 121959						+	+	+

'+' represents the presence of superior haplotype of each gene Figure 1. Phenotypic trait classes and chromosome-wise distribution of QTLs utilized in the MQTL analysis for PUE in rice.

availability of gene information from the reference genome may fail to precisely capture the genetic variability linked to PUE. Consequently, it is important to generate supplementary reference genomes, or a "rice meta-genome," that encompass diverse rice varieties and species to address this limitation. Nevertheless, our approach provides breeders with valuable information regarding the selection of optimal donors for their desired traits from the gene bank. However, the assumed genetic gains resulting from the accumulation of superior haplotypes of PUE CGs require validation in practical breeding programs.

5 Conclusion

We identified 38 meta-QTLs (MQTLs) for phosphorus use efficiency (PUE) that were supported by multiple QTLs from independent studies, which had a phenotypic variation explained (PVE) value of at least 5%. We subjected the 38 PUE MQTLs to candidate gene (CG) mining. The genomic regions associated with PUE MQTLs were found to be enriched with genes involved in the transmembrane transport of amino acids and organic acids, as well as genes involved in the response to auxin. Some superior haplotypes containing eight CGs for PUE could be considered for the genomics-assisted breeding in rice.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

IN and JC conceptualized and designed the manuscript. JC and WZ supervised the study. IN and PM curated data and performed analysis. IN wrote the manuscript. JC, WZ, PM, N-HS, J-HH, SY, and WJ reviewed end edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Half of this work was supported by grants from the framework of the international cooperation program managed by the National Research Foundation of Korea (NRF-2021K1A3A1A61002988) and National Key R&D Program of China (2022YFE0198100). Half of this work was supported by the Rural Administration of Korea (PJ017008).

Acknowledgments

We are grateful to Ms. Priskila Tolangi (Food Crops Molecular Breeding Laboratory, Department of Integrative Biological Sciences and Industry, Sejong University) on her efforts in the mining of related literatures used in 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/fpls.2023.1226297/full#supplementary-material

References

Abbai, R., Singh, V. K., Nachimuthu, V. V., Sinha, P., Selvaraj, R., Vipparla, A. K., et al. (2019). Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant Biotechnol. J.* 17 (8), 1612–1622. doi: 10.1111/pbi.13087

Ai, P., Sun, S., Zhao, J., Fan, X., Xin, W., Guo, Q., et al. (2009). Two rice phosphate transporters, *OsPht1;2 and OsPht1;6*, have different functions and kinetic properties in uptake and translocation. *Plant J.* 57 (5), 798–809. doi: 10.1111/j.1365-313X.2008.03726

Akaike, H. (1987). Factor analysis and AIC. *Psychometrica* 52, 317–332. doi: 10.1007/BF02294359

Aloryi, K. D., Okpala, N. E., Amo, A., Bello, S. F., Akaba, S., and Tian, X. (2022). A meta-quantitative trait loci analysis identified consensus genomic regions and candidate genes associated with grain yield in rice. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1035851

Anandan, A., Panda, S., Sabarinathan, S., Travis, A. J., Norton, G. J., and Price, A. H. (2022). Superior haplotypes for early root vigor traits in rice under dry direct seeded low nitrogen condition through genome wide association mapping. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.911775

Anilkumar, C., Sah, R. P., Muhammed Azharudheen, T. P., Behera, S., Singh, N., Prakash, N. R., et al. (2022). Understanding complex genetic architecture of rice grain weight through QTL-meta analysis and candidate gene identification. *Sci. Rep.* 12 (1). doi: 10.1038/s41598-022-17402-w

Anis, G. B., Zhang, Y., Wang, H., Li, Z., Wu, W., Sun, L., et al. (2018). Genomic regions analysis of seedling root traits and their regulation in responses to phosphorus deficiency tolerance in CSSL population of elite super hybrid rice. *Int. J. Mol. Sci.* 19 (5). doi: 10.3390/ijms19051460

Annicchiarico, P., Nazzicari, N., Pecetti, L., ROmani, M., Ferrari, B., Wei, Y., et al. (2017). GBS-based genomic selection for pea grain yield under severe terminal drought. *Plant Genome* 10 (2). doi: 10.3835/plantgenome2016.07.0072

Arcade, A., Labourdette, A., Falque, M., Mangin, B., Chardon, F., Charcosset, A., et al. (2004). BioMercator: Integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* 20 (14), 2324–2326. doi: 10.1093/bioinformatics/bth230

Banti, V., Mafessoni, F., Loreti, E., Alpi, A., and Perata, P. (2010). The heat-inducible transcription factor *HsfA2* enhances anoxia tolerance in Arabidopsis. *Plant Physiol.* 152 (3), 1471–1483. doi: 10.1104/pp.109.149815

Caliński, T., and Harabasz, J. (1974). A dendrite method foe cluster analysis. *Commun. Stat* 3 (1), 1–27. doi: 10.1080/03610927408827101

Chauhan, H., Khurana, N., Agarwal, P., and Khurana, P. (2011). Heat shock factors in rice (*Oryza sativa* L.): Genome-wide expression analysis during reproductive development and abiotic stress. *Mol. Genet. Genomics* 286 (2), 171–187. doi: 10.1007/s00438-011-0638-8

Chen, L., An, Y., Li, Y. X., Li, C., Shi, Y., Song, Y., et al. (2017). Candidate loci for yield-related traits in maize revealed by a combination of metaQTL analysis and regional association mapping. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02190

Chen, W. W., Zhu, H. H., Wang, J. Y., Han, G. H., Huang, R. N., Hong, Y. G., et al. (2022). Comparative physiological and transcriptomic analyses reveal altered fe-deficiency responses in tomato epimutant colorless non-ripening. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.796893

Chin, J. H., Gamuyao, R., Dalid, C., Bustamam, M., Prasetiyono, J., Moeljopawiro, S., et al. (2011). Developing rice with high yield under phosphorus deficiency: *Pup1* sequence to application. *Plant Physiol.* 156 (3), 1202–1216. doi: 10.1104/pp.111.175471

Collard, B. C., and Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B: Biol. Sci.* 363 (1491), 557–572. doi: 10.1098/rstb.2007.2170

Darvasi, A., and Seller, M. (1997). A simple method to calculate resolving power and confidence interval of QTL map location. *Behav. Genet.* 27, 125–132. doi: 10.1023/A:1025685324830

Daware, A. V., Srivastava, R., Singh, A. K., Parida, S. K., and Tyagi, A. K. (2017). Regional association analysis of metaQTLs delineates candidate grain size genes in rice. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00807

Deng, Q. W., Luo, X. D., Chen, Y. L., Zhou, Y., Zhang, F. T., Hu, B. L., et al. (2018). Transcriptome analysis of phosphorus stress responsiveness in the seedlings of Dongxiang wild rice (*Oryza rufipogon* Griff.). *Biol. Res.* 51 (1). doi: 10.1186/s40659-018-0155-x

Dobermann, A., and Fairhurst, T. (2000). "Phosphorus deficiency," in *Rice: Nutrient Disorders and Nutrient Management* (Los Banos, Philippines: International Rice Research Institute), 60–71.

Du, H., Qin, R., Li, H., Du, Q., Li, X., Yang, H., et al. (2022). Genome-wide association studies reveal novel loci for herbivore resistance in wild soybean (*Glycine soja*). *Int. J. Mol. Sci.* 23 (14). doi: 10.3390/ijms23148016

Duan, J., Zhang, M., Zhang, H., Xiong, H., Liu, P., Ali, J., et al. (2012). *OsMIOX*, a myo-inositol oxygenase gene, improves drought tolerance through scavenging of reactive oxygen species in rice (*Oryza sativa* L.). *Plant Sci.* 196, 143–151. doi: 10.1016/j.plantsci.2012.08.003

Fu, Y., Zhong, X., Pan, J., Liang, K., Liu, Y., Peng, B., et al. (2019). QTLs identification for nitrogen and phosphorus uptake-related traits using ultra-high density SNP linkage. *Plant Sci.* 288. doi: 10.1016/j.plantsci.2019.110209

Gamuyao, R., Chin, J. H., Pariasca-Tanaka, J., Pesaresi, P., Catausan, S., Dalid, C., et al. (2012). The protein kinase *Pstol1* from traditional rice confers tolerance of phosphorus deficiency. *Nature* 488 (7412), 535–539. doi: 10.1038/nature11346

Ge, S. X., Jung, D., Jung, D., and Yao, R. (2020). ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36 (8), 2628–2629. doi: 10.1093/bioinformatics/btz931

Goffinet, B., and Gerber, S. (2000). Quantitative trait loci: A meta-analysis. *Genetics* 155, 463–473. doi: 10.1093/genetics/155.1.463

Guo, N., Zhang, S., Gu, M., and Xu, G. (2021). Function, transport, and regulation of amino acids: What is missing in rice? *Crop J.* 9 (3), 530–542. doi: 10.1016/j.cj.2021.04.002

Guo, R., Zhang, Q., Ying, Y., Liao, W., Liu, Y., Whelan, J., et al. (2023). Functional characterization of the three *Oryza sativa* SPX-MFS proteins in maintaining phosphate homoeostasis. *Plant Cell Environ.* 46 (4), 1264–1277. doi: 10.1111/pce.14414

Haefele, S. M., Nelson, A., and Hijmans, R. J. (2014). Soil quality and constraints in global rice production. *Geoderma* 235–236, 250–259. doi: 10.1016/j.geoderma.2014.07.019

Hernández, I., and Munné-Bosch, S. (2015). Linking phosphorus availability with photo-oxidative stress in plants. *J. Exp. Bot.* 66 (10), 2889–2900. doi: 10.1093/jxb/erv056

Heuer, S., Gaxiola, R., Schilling, R., Herrera-Estrella, L., López-Arredondo, D., Wissuwa, M., et al. (2017). Improving phosphorus use efficiency: a complex trait with emerging opportunities. *Plant J.* 90 (5), 868–885. doi: 10.1111/tpj.13423

Hu, B., Wu, P., Liao, C. Y., Zhang, W. P., and Ni, J. J. (2001). QTLs and epistasis underlying activity of acid phosphatase under phosphorus sufficient and deficient condition in rice (*Oryza sativa* L.). *Plant Soil* 230, 99–105. doi: 10.1023/A:1004809525119

Huang, J., Zhao, X., Weng, X., Wang, L., and Xie, W. (2012). The rice B-box zinc finger gene family: genomic identification, characterization, expression profiling and diurnal analysis. *PloS One* 7 (10). doi: 10.1371/journal.pone.0048242

Irfan, M., Aziz, T., Maqsood, M. A., Bilal, H. M., Siddique, K. H. M., and Xu, M. (2020). Phosphorus (P) use efficiency in rice is linked to tissue-specific biomass and P allocation patterns. *Sci. Rep.* 10 (1). doi: 10.1038/s41598-020-61147-3

Islam, M. S., Ontoy, J., and Subudhi, P. K. (2019). Meta-analysis of quantitative trait loci associated with seedling-stage salt tolerance in rice (*Oryza Sativa* L.). *Plants* 8 (2). doi: 10.3390/plants8020033

Ismail, A. M., Heuer, S., Thomson, M. J., and Wissuwa, M. (2007). Genetic and genomic approaches to develop rice germplasm for problem soils. *Plant Mol. Biol.* 65 (4), 547–570. doi: 10.1007/s11103-007-9215-2

Joshi, G., Soe, Y. P., Palanog, A., Hore, T. K., Nha, C. T., Calayugan, M. I., et al. (2023). Meta-QTL s and haplotypes for efficient zinc biofortification of rice. *Plant Genome*. doi: 10.1002/tpg2.20315

Kale, R. R., Durga Rani, C. V., Anila, M., Mahadeva Swamy, H. K., Bhadana, V. P., Senguttuvel, P., et al. (2021). Novel major QTLs associated with low soil phosphorus tolerance identified from the Indian rice landrace, Wazuhophek. *PloS One* 16 (7), e0254526. doi: 10.1371/journal.pone.0254526

Kan, C. C., Chung, T. Y., Wu, H. Y., Juo, Y. A., and Hsieh, M. H. (2017). Exogenous glutamate rapidly induces the expression of genes involved in metabolism and defense responses in rice roots. *BMC Genomics* 18 (1). doi: 10.1186/s12864-017-3588-7

Khahani, B., Tavakol, E., Shariati, V., and Rossini, L. (2021). Meta-QTL and ortho-MQTL analyses identified genomic regions controlling rice yield, yield-related traits and root architecture under water deficit conditions. *Sci. Rep.* 11 (1). doi: 10.1038/s41598-021-86259-2

Khan, G. A., Bouraine, S., Wege, S., Li, Y., Carbonnel, M., Berthomieu, P., et al. (2014). Coordination between zinc and phosphate homeostasis involves the transcription factor *PHR1*, the phosphate exporter *PHO1*, and its homologue *PHO1; H3* in Arabidopsis. *J. Exp. Bot.* 65 (3), 871–884. doi: 10.1093/jxb/ert444

Koide, Y., Pariasca-Tanaka, J., Rose, T., Fukuo, A., Konisho, K., Yanagihara, S., et al. (2013). Qtls for phosphorus deficiency tolerance detected in upland nerica varieties. *Plant Breed.* 132 (3), 259–265. doi: 10.1111/pbr.12052

Kokaji, H., and Shimizu, A. (2022). An indica rice cultivar "Habataki" segment on chromosome 6 improves low-phosphorus tolerance. *J. Crop Res.* 67, 1–6.

Kumar, A., Sandhu, N., Venkateshwarlu, C., Priyadarshi, R., Yadav, S., Majumder, R. R., et al. (2020). Development of introgression lines in high yielding, semi-dwarf genetic backgrounds to enable improvement of modern rice varieties for tolerance to multiple abiotic stresses free from undesirable linkage drag. *Sci. Rep.* 10 (1). doi: 10.1038/s41598-020-70132-9

Kumari, S., Sharma, N., and Raghuram, N. (2021). Meta-analysis of yield-related and N-responsive genes reveals chromosomal hotspots, key processes and candidate genes for nitrogen-use efficiency in rice. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.627955

Ligaba, A., Yamaguchi, M., Shen, H., Sasaki, T., Yamamoto, Y., and Matsumoto, H. (2004). Phosphorus deficiency enhances plasma membrane H+-ATPase activity and citrate exudation in greater purple lupin (*Lupinus pilosus*). *Funct. Plant Biol.* 31 (11), 1075–1083. doi: 10.1071/FP04091

Liu, D. (2021). Root development responses to phosphorus nutrition. *J. Integr. Plant Biol.* 63, 1065–1090. doi: 10.1111/jipb.13090

López-Bucio, J., Cruz-Ramırez, A., and Herrera-Estrella, L. (2003). The role of nutrient availability in regulating root architecture. *Curr. Opin. Plant Biol.* 6 (3), 280–287. doi: 10.1016/S1369-5266(03)00035-9

Lu, J., Qian, Y., Li, Z., Yang, A., Zhu, Y., Li, R., et al. (2010). Mitochondrial haplotypes may modulate the phenotypic manifestation of the deafness-associated 12S rRNA 1555A>G mutation. *Mitochondrion* 10 (1), 69–81. doi: 10.1016/j.mito.2009.09.007

Lynch, J. P. (2011). Root phenes for enhanced soil exploration and phosphorus acquisition: Tools for future crops. *Plant Physiol.* 156 (3), 1041–1049. doi: 10.1104/pp.111.175414

Magwa, R. A., Zhao, H., and Xing, Y. (2016). Genome-wide association mapping revealed a diverse genetic basis of seed dormancy across subpopulations in rice (*Oryza sativa* L.). *BMC Genet.* 17 (1). doi: 10.1186/s12863-016-0340-2

Mansueto, L., Fuentes, R. R., Chebotarov, D., Borja, F. N., Detras, J., Abriol-Santos, J. M., et al. (2016). SNP-Seek II: A resource for allele mining and analysis of big genomic data in Oryza sativa. *Curr. Plant Biol.* 7–8, 16–25. doi: 10.1016/j.cpb.2016.12.003

Miao, Y., Jing, F., Ma, J., Liu, Y., Zhang, P., Chen, T., et al. (2022). Major genomic regions for wheat grain weight as revealed by QTL linkage mapping and meta-analysis. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.802310

Ming, F., Zheng, X., and Mi, G. (2000). Identification of quantitative trait loci affecting tolerance to low phosphorus in rice (*Oryza Sativa* L.). *Chin. Sci. Bull.* 45, 520–525. doi: 10.1007/BF02887097

Mori, A., Fukuda, T., Vejchasarn, P., Nestler, J., Pariasca-Tanaka, J., and Wissuwa, M. (2016). The role of root size versus root efficiency in phosphorus (P) acquisition of rice. *J. Exp. Bot.* 67 (4), 1179–1189. doi: 10.1093/jxb/erv557

Müller, J., Toev, T., Heisters, M., Teller, J., Moore, K. L., Hause, G., et al. (2015). Iron-dependent callose deposition adjusts root meristem maintenance to phosphate availability. *Dev. Cell* 33 (2), 216–230. doi: 10.1016/j.devcel.2015.02.007

Nacry, P., Canivenc, G., Muller, B., Azmi, A., Van Onckelen, H., Rossignol, M., et al. (2005). A role for auxin redistribution in the responses of the root system architecture to phosphate starvation in Arabidopsis. *Plant Physiol.* 138 (4), 2061–2074. doi: 10.1104/pp.105.060061

Nagelkerke, N. J. D. (1991). A note on a general definition of the coef-ficient of determination. *Biometrika* 78, 691–692. doi: 10.1093/biomet/78.3.69

Navea, I. P., Dwiyanti, M. S., Park, J., Kim, B., Lee, S., Huang, X., et al. (2017). Identification of quantitative trait loci for panicle length and yield related traits under different water and P application conditions in tropical region in rice (*Oryza sativa* L.). *Euphytica* 213 (2). doi: 10.1007/s10681-016-1822-z

Navea, I. P., Han, J. H., Shin, N. H., Lee, O. N., Kwon, S. W., Choi, I. R., et al. (2022). Assessing the Effect of a Major Quantitative Locus for Phosphorus Uptake (*Pup1*) in Rice (*O. sativa* L.) Grown under a Temperate Region. *Agric. (Switzerland)* 12 (12). doi: 10.3390/agriculture12122056

Ni, J. J., Wu, P., Senadhira, D., and Huang, N. (1998). Mapping QTLs for phosphorus deficiency tolerance in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 97, 1361–1369. doi: 10.1007/s001220051030

Ogawa, S., Selvaraj, M. G., Fernando, A. J., Lorieux, M., Ishitani, M., McCouch, S., et al. (2014). N- and P-mediated seminal root elongation response in rice seedlings. *Plant Soil* 375 (1–2), 303–315. doi: 10.1007/s11104-013-1955-y

Ookawa, T., Hobo, T., Yano, M., Murata, K., Ando, T., Miura, H., et al. (2010). New approach for rice improvement using a pleiotropic QTL gene for lodging resistance and yield. *Nat. Commun.* 1 (8). doi: 10.1038/ncomms1132

Park, Y. C., Lim, S. D., Moon, J. C., and Jang, C. S. (2019). A rice really interesting new gene H2-type E3 ligase, *OsSIRH2-14*, enhances salinity tolerance *via* ubiquitin/26S proteasome-mediated degradation of salt-related proteins. *Plant Cell Environ.* 42 (11), 3061–3076. doi: 10.1111/pce.13619

Paul, S., Ali, N., Gayen, D., Datta, S. K., and Datta, K. (2012). Molecular breeding of *Osfer 2* gene to increase iron nutrition in rice grain. *GM Crops Food* 3 (4), 310–316. doi: 10.4161/gmcr.22104

Péret, B., Clément, M., Nussaume, L., and Desnos, T. (2011). Root developmental adaptation to phosphate starvation: better safe than sorry. *Trends Plant Sci.* 16 (8), 442–450. doi: 10.1016/j.tplants.2011.05.006

Pérez-Torres, C. A., López-Bucio, J., Cruz-Ramírez, A., Ibarra-Laclette, E., Dharmasiri, S., Estelle, M., et al. (2008). Phosphate availability alters lateral root development in Arabidopsis by modulating auxin sensitivity *via* a mechanism involving the *TIR1* auxin receptor. *Plant Cell* 20 (12), 3258–3272. doi: 10.1105/tpc.108.058719

Ping, M. U., HUANG, C., Jun-Xia, L. I., Li-Feng, L. I. U., and Zi-Chao, L. I. (2008). Yield trait variation and QTL mapping in a DH population of rice under phosphorus deficiency. *Acta Agro. Sin.* 34 (7), 1137–1142. doi: 10.3724/sp.j.1006.2008.01137

Raghavan, C., Mauleon, R., Lacorte, V., Jubay, M., Zaw, H., Bonifacio, J., et al. (2017). Approaches in characterizing genetic structure and mapping in a rice multiparental population. *G3: Genes Genomes Genet.* 7 (6), 1721–1730. doi: 10.1534/g3.117.042101

Ranaivo, H. N., Lam, D. T., Ueda, Y., Pariasca Tanaka, J., Takanashi, H., Ramanankaja, L., et al. (2022). QTL mapping for early root and shoot vigor of upland rice (*Oryza sativa* L.) under P deficient field conditions in Japan and Madagascar. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1017419

Rose, T. J., Rose, M. T., Pariasca-Tanaka, J., Heuer, S., and Wissuwa, M. (2011). The frustration with utilization: why have improvements in internal phosphorus utilization efficiency in crops remained so elusive? *Front. Plant Sci.* 2. doi: 10.3389/fpls.2011.00073

Rose, T. J., and Wissuwa, M. (2012). "Rethinking internal phosphorus utilization efficiency: a new approach is needed to improve PUE in grain crops," in *Advances in Agronomy*, vol. 116 . Ed. D. L. Sparks (Burlington: Elsevier Inc), 185–217.

Sandhu, N., Pruthi, G., Prakash Raigar, O., Singh, M. P., Phagna, K., Kumar, A., et al. (2021). Meta-QTL analysis in rice and cross-genome talk of the genomic regions controlling nitrogen use efficiency in cereal crops revealing phylogenetic relationship. *Front. Genet.* 12. doi: 10.3389/fgene.2021.807210

Selamat, N., and Nadarajah, K. K. (2021). Meta-analysis of quantitative traits loci (QTL) identified in drought response in rice (*Oryza sativa* l.). *Plants* 10 (4). doi: 10.3390/plants10040716

Shimizu, A., Yanagihara, S., Kawasaki, S., and Ikehashi, H. (2004). Phosphorus deficiency-induced root elongation and its QTL in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 109 (7), 1361–1368. doi: 10.1007/s00122-004-1751-4

Shin, N. H., Lee, O. N., Han, J. H., Song, K., Koh, H. J., Yoo, S. C., et al. (2021). The effect of water level in rice cropping system on phosphorus uptake activity of *Pup1* in a *Pup1+Sub1* breeding line. *Plants* 10 (8). doi: 10.3390/plants10081523

Sinha, P., Singh, V. K., Saxena, R. K., Khan, A. W., Abbai, R., Chitikineni, A., et al. (2020). Superior haplotypes for haplotype-based breeding for drought tolerance in pigeon pea (*Cajanus cajan* L.). *Plant Biotechnol. J.* 18 (12), 2482–2490. doi: 10.1111/pbi.13422

Soriano, J. M., Colasuonno, P., Marcotuli, I., and Gadaleta, A. (2021). Meta-QTL analysis and identification of candidate genes for quality, abiotic and biotic stress in durum wheat. *Sci. Rep.* 11 (1). doi: 10.1038/s41598-021-91446-2

Sosnowski, O., Charcosset, A., and Joets, J. (2012). Biomercator V3: An upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics* 28 (15), 2082–2083. doi: 10.1093/bioinformatics/bts313

Stephens, M., Smith, N. J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68 (4), 978–989. doi: 10.1086/319501

Szabados, L., and Savouré, A. (2010). Proline: a multifunctional amino acid. *Trends Plant Sci.* 15 (2), 89–97. doi: 10.1016/j.tplants.2009.11.009

Thomson, M. J., Singh, N., Dwiyanti, M. S., Wang, D. R., Wright, M. H., Perez, F. A., et al. (2017). Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. *Rice* 10 (1). doi: 10.1186/s12284-017-0181-2

Van Kauwenbergh, S. J. (2010). *World phosphate rock reserves and resources* (Muscle Shoals: Ifdc), 48.

Venske, E., Dos Santos, R. S., Farias, D. D. R., Rother, V., Da Maia, L. C., Pegoraro, C., et al. (2019). Meta-analysis of the QTLome of Fusarium head blight resistance in bread wheat: refining the current puzzle. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00727

Veyrieras, J. B., Goffinet, B., and Charcosset, A. (2007). MetaQTL: A package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinf.* 8. doi: 10.1186/1471-2105-8-49

Viana, V. E., Marini, N., Finatto, T., Ezquer, I., Busanello, C., dos Santos, R. S., et al. (2017). Iron excess in rice: From phenotypic changes to functional genomics of *WRKY* transcription factors. *Genet. Mol. Res.* 16 (3). doi: 10.4238/gmr16039694

Vishnukiran, T., Neeraja, C. N., Jaldhani, V., Vijayalakshmi, P., Raghuveer Rao, P., Subrahmanyam, D., et al. (2020). A major pleiotropic QTL identified for yield components and nitrogen content in rice (Oryza sativa L.) under differential nitrogen field conditions. *PloS One* 15 (10), e0240854. doi: 10.1371/journal.pone.0240854

Wang, K., Cui, K., Liu, G., Xie, W., Yu, H., Pan, J., et al. (2014). Identification of quantitative trait loci for phosphorus use efficiency traits in rice using a high-density SNP map. *BMC Genet.* 15 (1). doi: 10.1186/s12863-014-0155-y

Wang, W., Ding, G. D., White, P. J., Wang, X. H., Jin, K. M., Xu, F. S., et al. (2019). Mapping and cloning of quantitative trait loci for phosphorus efficiency in crops: opportunities and challenges. *Plant Soil* 439, 91–112. doi: 10.2307/48703803

Wang, M., Kawakami, Y., and Bhullar, N. K. (2019). Molecular analysis of iron deficiency response in hexaploid wheat. *Front. Sustain. Food Syst.* 3. doi: 10.3389/fsufs.2019.00067

Wang, C., Yue, W., Ying, Y., Wang, S., Secco, D., Liu, Y., et al. (2015). Rice SPX-Major Facility Superfamily3, a vacuolar phosphate efflux transporter, is involved in maintaining phosphate homeostasis in rice. *Plant Physiol.* 169 (4), 2822–2831. doi: 10.1104/pp.15.01005

Wang, S., Zhang, S., Sun, C., Xu, Y., Chen, Y., Yu, C., et al. (2014). Auxin response factor (*OsARF12*), a novel regulator for phosphate homeostasis in rice (Oryza sativa). *New Phytol.* 201 (1), 91–103. doi: 10.1111/nph.12499

Ward, J. T., Lahner, B., Yakubova, E., Salt, D. E., and Raghothama, K. G. (2008). The effect of iron on the primary root elongation of Arabidopsis during phosphate deficiency. *Plant Physiol.* 147 (3), 1181–1191. doi: 10.1104/pp.108.118562

Wasaki, J., Yonetani, R., Shinano, T., Kai, M., and Osaki, M. (2003). Expression of the OsPI1 gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status. *New Phytol.* 158 (2), 239–248. doi: 10.1046/j.1469-8137.2003.00748.x

Weller, J. I., and Soller, M. (2004). An analytical formula to estimate confidence interval of QTL location with a saturated genetic map as a function of experimental design. *Theor. Appl. Genet.* 109 (6), 1224–1229. doi: 10.1007/s00122-004-1664-2

Wissuwa, M., Kondo, K., Fukuda, T., Mori, A., Rose, M. T., Pariasca-Tanaka, J., et al. (2015). Unmasking novel loci for internal phosphorus utilization efficiency in rice germplasm through genome-wide association analysis. *PloS One* 10 (4). doi: 10.1371/journal.pone.0124215

Wissuwa, M., Yano, M., and Ae, N. (1998). Mapping of QTLs for phosphorus-deficiency tolerance in rice (Oryza sativa L.). *Theor. Appl. Genet.* 97, 777–783. doi: 10.1007/s001220050955

Wu, T., Wang, C., Han, B., Liu, Z., Yang, X., Wang, W., et al. (2023). Emerging roles of inositol pyrophosphates in signaling plant phosphorus status and phytohormone signaling. *Plant Soil*, 1–19. doi: 10.1007/s11104-023-05976-x

Xiang, J., Ran, J., Zou, J., Zhou, X., Liu, A., Zhang, X., et al. (2013). Heat shock factor *OsHsfB2b* negatively regulates drought and salt tolerance in rice. *Plant Cell Rep.* 32 (11), 1795–1806. doi: 10.1007/s00299-013-1492-4

Xu, K., Xu, X., Fukao, T., Canlas, P., MaghIrang-Rodriguez, R., Heuer, S., et al. (2006). *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442 (7103), 705–708. doi: 10.1038/nature04920

Xu, X., Zhiguo, E., Zhang, D., Yun, Q., Zhou, Y., Niu, B., et al. (2021). *OsYUC11*-mediated auxin biosynthesis is essential for endosperm development of rice. *Plant Physiol.* 185 (3), 934–950. doi: 10.1093/PLPHYS/KIAA057

Yan, W. H., Wang, P., Chen, H. X., Zhou, H. J., Li, Q. P., Wang, C. R., et al. (2011). A major QTL, *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Mol. Plant* 4 (2), 319–330. doi: 10.1093/mp/ssq070

Yang, S. Y., Huang, T. K., Kuo, H. F., and Chiou, T. J. (2017). Role of vacuoles in phosphorus storage and remobilization. *J. Exp. Bot.* 68 (12), 3045–3055. doi: 10.1093/jxb/erw481

Yang, H., Zhang, X., Gaxiola, R. A., Xu, G., Peer, W. A., and Murphy, A. S. (2014). Over-expression of the Arabidopsis proton-pyrophosphatase *AVP1* enhances transplant survival, root mass, and fruit development under limiting phosphorus conditions. *J. Exp. Bot.* 65 (12), 3045–3053. doi: 10.1093/jxb/eru149

Yi, K., Wu, Z., Zhou, J., Du, L., Guo, L., Wu, Y., et al. (2005). *OsPTF1*, a novel transcription factor involved in tolerance to phosphate starvation in rice. *Plant Physiol.* 138 (4), 2087–2096. doi: 10.1104/pp.105.063115

Yu, W., Kan, Q., Zhang, J., Zeng, B., and Chen, Q. (2016). Role of the plasma membrane H+-ATPase in the regulation of organic acid exudation under aluminum toxicity and phosphorus deficiency. *Plant Signaling Behav.* 11 (1), e1106660. doi: 10.1080/15592324.2015.1106660

Zhang, J., Xiang, C., Zhang, J., Ren, J., Liu, Z., and Wang, C. (2014). Mapping QTL controlling yield traits using low phosphorus tolerance selected backcrossing introgression lines of rice (*Oryza sativa* L.). *Chin. Agric. Sci. Bull.* 30, 56–65.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 October 2023
DOI 10.3389/fpls.2023.1256186

OPEN ACCESS

EDITED BY
Shouvik Das,
Regional Centre for Biotechnology (RCB), India

REVIEWED BY
Janardanan Sreekumar,
Central Tuber Crops Research Institute (ICAR), India
Guosheng Han,
Xiangtan University, China
Shashi Rawat,
ICAR-Central Institute of Agricultural Engineering, India

*CORRESPONDENCE
Sunil Archak
sunil.archak@icar.gov.in

RECEIVED 10 July 2023
ACCEPTED 01 September 2023
PUBLISHED 09 October 2023

CITATION
Sinha D, Dasmandal T, Paul K, Yeasin M, Bhattacharjee S, Murmu S, Mishra DC, Pal S, Rai A and Archak S (2023) MethSemble-6mA: an ensemble-based 6mA prediction server and its application on promoter region of LBD gene family in Poaceae. *Front. Plant Sci.* 14:1256186. doi: 10.3389/fpls.2023.1256186

MethSemble-6mA: an ensemble-based 6mA prediction server and its application on promoter region of LBD gene family in Poaceae

Dipro Sinha[1,2], Tanwy Dasmandal[1,2,3], Krishnayan Paul[2,4], Md Yeasin[1], Sougata Bhattacharjee[2,4,5], Sneha Murmu[1], Dwijesh Chandra Mishra[1], Soumen Pal[1], Anil Rai[6] and Sunil Archak[7]*

[1]ICAR-Indian Agricultural Statistics Research Institute, Delhi, India, [2]Graduate School, ICAR-Indian Agricultural Research Institute, Delhi, India, [3]ICAR-National Bureau of Fish Genetic Resources, Lucknow, India, [4]ICAR-National Institute for Plant Biotechnology, Delhi, India, [5]ICAR-Indian Agricultural Research Institute, Hazaribagh, Jharkhand, India, [6]Indian Council of Agricultural Research, Delhi, India, [7]ICAR-National Bureau of Plant Genetic Resources, Delhi, India

The Lateral Organ Boundaries Domain (LBD) containing genes are a set of plant-specific transcription factors and are crucial for controlling both organ development and defense mechanisms as well as anthocyanin synthesis and nitrogen metabolism. It is imperative to understand how methylation regulates gene expression, through predicting methylation sites of their promoters particularly in major crop species. In this study, we developed a user-friendly prediction server for accurate prediction of 6mA sites by incorporating a robust feature set, viz., Binary Encoding of Mono-nucleotide DNA. Our model, MethSemble-6mA, outperformed other state-of-the-art tools in terms of accuracy (93.12%). Furthermore, we investigated the pattern of probable 6mA sites at the upstream promoter regions of the LBD-containing genes in *Triticum aestivum* and its allied species using the developed tool. On average, each selected species had four 6mA sites, and it was found that with speciation and due course of evolution in wheat, the frequency of methylation have reduced, and a few sites remain conserved. This obviously cues gene birth and gene expression alteration through methylation over time in a species and reflects functional conservation throughout evolution. Since DNA methylation is a vital event in almost all plant developmental processes (e.g., genomic imprinting and gametogenesis) along with other life processes, our findings on epigenetic regulation of LBD-containing genes have dynamic implications in basic and applied research. Additionally, MethSemble-6mA (http://cabgrid.res.in:5799/) will serve as a useful resource for a plant breeders who are interested to pursue epigenetic-based crop improvement research.

KEYWORDS

6mA, ensemble model, DNA methylation, MethSemble-6mA, LBD gene, wheat, poaceae, prediction

1 Introduction

Plant architecture is an important trait that distinguishes domesticated plant types from wild ones and enables breeders to choose the most productive types for agriculture. One of the major gene families that influence plant architecture is the family of Lateral Organ Boundaries Domain (LBD) genes. In plant systems, LBD genes have a wide role from embryonic development to stress resistance (Zhao et al., 2023). The role of LBD-containing genes has been reported in various agricultural crops including ideotype in rice (Zhao et al., 2023); drought tolerance in maize (Jiao et al., 2022), tomato (Liu et al., 2020), and potato (Liu et al., 2019); salt tolerance in switch grass (Guan et al., 2023); and multiple abiotic stress tolerance in wheat (Wang et al., 2021) and cotton (Yu et al., 2020).

In addition to studying the structure and function of genes, it is essential to unravel gene regulation to achieve expected plant architecture and better quantity and quality of produce. Genes in eukaryotes are regulated, among others, by epigenetic factors including stress memory. Conrad Waddington introduced the field of epigenetics in 1942, which explores heritable and reversible alterations in gene expression without modifications to the DNA sequence (Waddington, 2012). The involvement of these modifications in plants' gene regulatory mechanism has been revealed in recent studies for both biotic (Ashapkin et al., 2020) and abiotic stresses (Saraswat et al., 2017). Epigenetic mechanisms encompass diverse biological processes such as DNA methylation, histone modification, and chromosome remodeling, among which DNA methylation is considered a fundamental and widely distributed epigenetic process in various animal genomes, directly impacting gene expression (Ratel et al., 2006). DNA methylation can be categorized based on the location of methylation as N^6-methyladenine (6mA) (O'Brown and Greer, 2016), N^4-methylcytosine (4mC), and 5-methylcytosine (5mC) (Zhou et al., 2018; Lv et al., 2019; Lv et al., 2020). Although many studies have been performed on methylated cytosine, the potency of 6mA methylation is yet to be unveiled thoroughly. As per the literature, 6mA plays a vital role in basic cell functions such as replication (Campbell and Kleckner, 1990), transcription (Robbins-Manke et al., 2005), and repair (Pukkila et al., 1983). Although its presence can be observed in all three kingdoms of life, the distribution pattern of 6mA sites throughout the genome does not occur randomly, making it essential to accurately identify the specific locations of 6mA positions across the entire genome.

In recent years, there has been significant progress in high-throughput sequencing techniques, enabling the study of DNA 6mA modifications on a genome-wide scale. For instance, a method combining bisulfite sequencing with methyl-DNA immunoprecipitation was developed to identify 6mA sites in eukaryotes (Pomraning et al., 2009). Another efficient technique involved capillary electrophoresis and laser-induced fluorescence to quantify global adenine methylation of DNA (Krais et al., 2010). Additionally, the single-molecule real-time (SMRT) sequencing technology was utilized to detect genome-wide positions of 4mC and 6mA throughout the entire genome (Flusberg et al., 2010). Mass spectrometry analysis and 6mA immunoprecipitation followed by sequencing (IP-seq) were also employed to decipher 6mA sites in the rice genome (Zhou et al., 2018). However, these approaches have three significant limitations: time-consuming, labor-intensive, and expensive.

For bypassing these issues, *in silico* prediction tools can provide a faster and more reliable alternative to these *in vitro* methods. The ground-breaking research originated in 2019 when Chen et al. introduced a classifier called i6mA-Pred, based on a support vector machine (SVM), which was developed using a feature space consisting of nucleotide frequencies and nucleotide chemical properties. This classifier was trained and evaluated on a benchmark rice dataset comprising 880 6mA sites and 880 non-6mA sites obtained from the rice genome (Chen et al., 2019). So far, a few other attempts have been made to predict these modifications in plants, viz., iDNA6mA (Tahir et al., 2019), SDM6A (Basith et al., 2019), iDNA6mA-Rice (Lv et al., 2019), SNNRice6mA (Yu and Dai, 2019), i6mA-DNCP (Kong and Zhang, 2019), and i6mA-Caps (Rehman et al., 2022) for rice (Huang et al., 2020) and *Arabidopsis* (Wang and Yan, 2018). A major shortcoming of these tools is that they are based on single species, and cross-species performance is not well known. Only Meta-i6mA (Hasan et al., 2021), proposed in 2021, has provided a significant result in cross-species.

Keeping these lacunas in mind, in our first attempt, we developed an ensemble-based model called EpiSemble (Sinha et al., 2023) based on two model plant species, viz., *Oryza* and *Arabidopsis*. It outperformed the state-of-the-art tools for all the evaluation measures. In the present study, we improved our model by adding more robust features set, viz., binary encoded nucleotide frequencies, and we could achieve better performance in terms of accuracy while applying it in cross-species analysis. In order to study the regulation of complex genes and gene families, we chose LBD genes, as they have a crucial role in both developmental and stress conditions and are conserved across species (Wang et al., 2021; Xu et al., 2021). We investigated the distribution pattern of the 6mA sites in the promoter region, in the four species of *Triticum*, one species of *Oryza*, and one species of *Arabidopsis*. Here we report the mapping the 6mA sites of selected LBD domain-containing genes to understand their functionality and conservation across species.

2 Materials and methods

Section 2.1 deals with the materials and methods used for the construction of the prediction model, and Section 2.2 deals with the materials and methods employed for the analysis of methylation of promoter regions of LBD genes.

2.1 Construction of prediction model

The prediction model consisted of three modules: vectorization of DNA fragments, feature set optimization using a hybrid feature selection module, and finally, ensemble modeling. The aim was to

include as many features as possible in the model to obtain higher prediction accuracy.

2.1.1 Dataset description

All the datasets was downloaded from public domains. For the training of the machine learning models, we selected benchmark datasets of RiceLv (sub sp) and *Arabidopsis* thaliana (Hasan et al., 2021) and the test dataset was of rice Nipponbare (http://www.elabcaas.cn/smep/index.html). The details of positive and negative 6mA samples in these datasets are given in Table 1. Size of all the positive and negative DNA fragments was 41 bp (Wang et al., 2021). The total rice and *Arabidopsis* dataset was used for training purposes, and the Nipponbare dataset was used for testing purposes. This approach provides a validation of the application of the developed model in intra-specific species.

2.1.2 Feature extraction of DNA sequences

DNA sequences need to be vectorized before the machine learning module can be applied. For this, five feature vectors were considered. Dinucleotide frequency (DNF) has been proven to be an efficient feature for converting short DNA fragments (Hasan et al., 2021). It also helps to reduce the time complexity of the computation, as it comprises less vector space, i.e., 16 in comparison with higher-order nucleotide frequencies (tri-, tetra-, hexa-, etc.). Another feature is nucleotide chemical properties (NCPs), where the bases give a score based on their physio-chemical properties like ring structures (single or double), hydrogen bonds (two or three), and base composition (amino or keto). Based on this, the four bases are represented as (1, 1, 1), (0, 0, 1), (1, 0, 0), and (0, 1, 0) for adenine, cytosine, guanine, and thymine, respectively (Chen et al., 2019). It can be also noticed that guanine–cytosine (GC) content varies in different DNA fragments based on their roles. In this study, a log-transformed GC content was used. Transformation is performed to reduce the weight of the GC content feature compared to other feature sets (Sinha et al., 2022). Another recently used feature is the Average Mutual Information Profile (AMIP), where the AMI measures the level of "information" that can be obtained from the relationship between two random variables, X and Y (X and Y are the two DNA sequences here) (Bauer et al., 2008). In the context of genomic sequences, X and Y represent nucleotide bases. Therefore, the proposed genomic signature is a vector where each entry corresponds to the AMI between nucleotides that are a certain number of positions apart. The AMI profile provides a summary of the statistical dependencies between nucleotides at different distances within the sequence (Eq. 1). To put it simply, the AMI profile is a way to represent the amount of shared information between nucleotides separated by specific distances in a genomic sequence.

$$MI_k = \sum_{X\in S}\sum_{Y\in S} p_k(X,Y) log \frac{p_k(X,Y)}{p(X)p(Y)}. \quad (1)$$

Here, p_k is the probability of two nucleotides occurring together at k distance apart.

Along with these features, another robust feature for encoding DNA sequences, Binary Encoding of Mono-nucleotide DNA (MBED), was incorporated into the model. MBED was proven to be an efficient representer in the case of cross-species. In this, the four nucleotides A, C, G, and T are represented as (1000), (0010), (0100), and (0001) respectively (Wang et al., 2021).

2.1.3 Selection of informative features

The feature set contains both relevant and irrelevant features. Irrelevant features may lead to improper training of the models. In previous studies, feature selection techniques like Maximum Relevance Maximum Distance (MRMD) (Chen et al., 2019) and Sequential Forward Selection (SFS) (Basith et al., 2019) were implemented. To obtain more robust features, a hybrid feature selection module was used in this study combining random forest and stepwise regression (Chen and Howard, 2015).

2.1.4 Machine learning models

Based on the performance of these two datasets, three machine learning models were chosen, viz., SVM (Cortes et al., 1995), random forest (RF) (Breiman, 2001), and gradient boosting (GB) (Friedman, 2001). SVM tends to find the optimum hyperplane between 6mA and non-6mA sequences. In this study, radial function $[k\ (x_i, x_j) = (-\gamma|x_i - x_j)^2]$ was taken as a kernel with the "C-classification" type. R-package "e1071" was used to implement SVM. For RF, the number of trees was set to 500 with five splits. R-package "randomForest" was used for RF. In the case of GB, the stochastic method was chosen with an interaction depth of 3 and shrinkage value of 0.1 and 150 ntrees. It was implemented using the "gbm" R-package.

2.1.5 Ensemble model

To integrate the prediction from each model, ensemble voting was used. Here, an instance with the highest vote, which means with at least two predictions favoring it, was chosen. The final prediction was computed as follows using the prediction score (Eq. 2):

$$Prediction = \frac{1}{M}\sum_{i=1}^{M} P_i, \quad (2)$$

TABLE 1 Description of the datasets used for model construction and validation.

Data Class	Training data		Testing data (Nipponbare)
	RiceLv	*Arabidopsis*	
Positive	154,000	31,873	5,000
Negative	154,000	31,873	10,000
Total	308,000	63,746	15,000

where M is the number of ML models, and P_i is the prediction value, i.e., whether it is methylated or not of each ML model. The workflow of the model is given in Figure 1.

2.1.6 Performance evaluation

As in the other experiments, evaluation metrics for the classifiers included sensitivity, specificity, accuracy, and Matthews' correlation coefficient (MCC) (Basith et al., 2019; Lv et al., 2019; Yu and Dai, 2019; Huang et al., 2020; Kha et al., 2022; Rehman et al., 2022; Sinha et al., 2023).

The proportion of positively tagged cases that are projected to be positive is termed as sensitivity.

$$\text{Sensitivity} = \frac{P_+^+}{P_+^+ + P_+^-}$$

The proportion of negatively tagged cases that are projected to be negative is termed as specificity.

$$\text{Specificity} = \frac{P_-^-}{P_-^- + P_-^+}$$

The ratio of successfully identified cases to all test data points is known as accuracy.

$$\text{Accuracy} = \frac{P_+^+ + P_-^-}{P_+^+ + P_-^- + P_-^+ + P_+^-}$$

Between the actual and predicted series, there is a correlation known as the MCC. It returns numbers between −1 and +1. A value of 0 is similar to a random forecast, while a coefficient of −1 signifies a full difference between the prediction and the observation. A coefficient of +1 denotes a flawless prediction. The MCC can be calculated directly from the confusion matrix by the following formula:

$$\text{MCC} = \frac{P_+^+ \times P_-^- - P_-^+ \times P_+^-}{\sqrt{(P_+^+ + P_-^+)(P_+^+ + P_+^-)(P_-^- + P_-^+)(P_-^- + P_+^-)}}$$

where

P_+^+ = Instances that are true and predicted as true.
P_-^- = Instances that are false and predicted as false.
P_-^+ = Instances that are false but predicted as true.
P_+^- = Instances that are true but predicted as false.

The receiver operating characteristic (ROC) curve was also used to evaluate the effectiveness of this strategy. One of the most important measures of a binary classifier's effectiveness is the area under the ROC curve (AUC), which is determined by graphing the true positive rate (sensitivity) against the false-positive rate (1 – specificity). Better predictions are produced when the value is closer to 1, whereas a value of 0.5 indicates random prediction.

2.2 Promoter analysis of LBD genes

Upstream promoter regions (1.5 kb) of LBD-containing genes of six crop species including monocot species (*T. aestivum* (Wang et al., 2021), *Triticum dicoccoides*, *Triticum urartu*, *Aegilops tauschii*, and *Oryza sativa*) and dicot species (*Arabidopsis thaliana* (Xu et al., 2021)) were selected for this study (Table 2). The basis of selection was to capture representative species (*O. sativa* acted as a reference for the Poaceae family, while *A. thaliana* acted as a non-grass comparison). These sequences are available with high sequencing coverage, which will provide a more informative prediction of our analysis. In order to understand the dynamics of 6mA sites through evolution and speciation, our findings on wheat were emphasized, and therefore, four out of six species were selected from the Triticeae family.

2.2.1 Extraction of promoter region

Promoter regions of these genes were extracted from the Ensembl Plants database using the Biomart tool (http://plants.ensembl.org/info/data/biomart/index.html). Promoter regions were selected from their respective databases like *Triticum aestivum* genes (IWGSR) for *T. aestivum*, *Triticum dicoccoides* genes (WEWSEQv.1.0) for *T. dicoccoides*, *Triticum urartu* genes

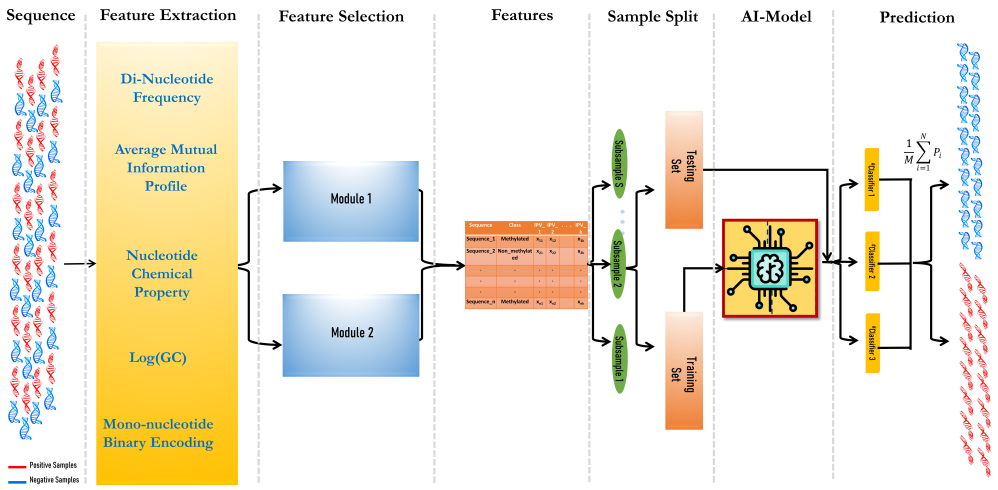


FIGURE 1
Workflow of the MethSemble-6mA model.

TABLE 2 Species-wise frequency of identified LBD genes.

Species	Number of LBD genes
Triticum aestivum	94
Triticum dicoccoides	49
Triticum urartu	27
Aegilops tauschii	29
Oryza sativa	37
Arabidopsis thaliana	43

(Tu2.0) for *T. urartu*, *Aegilops tauschii* genes (Aetv4.0) for *A. tauschii*, *Oryza sativa indica* group genes (ASM465v1) for *O. sativa*, and *Arabidopsis thaliana* genes for *A. thaliana*. The 1.5-kb upstream region of each gene from the transcription start site (TSS) was extracted.

2.2.2 Pre-processing of promoter sequences

The ambiguous bases (containing "N") were trimmed out from promoter sequences using the seqkit tool (https://bioinf.shenwei.me/seqkit/). Then, the promoters from all six species were fragmented into 41-bp sequences using the "split fasta" function of sequence Manipulation Suite (https://www.bioinformatics.org/sms2/split_fasta.html) separately. The resulting fasta file was fed into the MethSemble-6mA server for the prediction of the sequences containing 6mA sites.

2.2.3 Prediction of 6mA sites in the promoter of LBD genes

For the prediction of the 6mA sites in the promoter region of the LBD genes, a bidirectional approach was taken into consideration. The prediction was performed by taking both rice and *Arabidopsis* as model plants followed by taking the intersection (Eq. 3) of the two predictions. As in the dataset, both types of data are present, viz., Poaceae and Brassicaceae; these two were used concurrently, which means prediction has been performed by taking rice and *Arabidopsis* as model plants one by one. This will provide a robust prediction and reduce the chance of obtaining false-positive sites.

$$6mA\ \ (LBD) = Model_{Rice}\ \cap\ Model_{Arabidopsis} \quad (3)$$

2.2.4 Mapping of 6mA sites and phylogenetic analysis

Predicted 6mA sites found in upstream promoter elements of the LBD-containing genes were filtered for the highest number of sites present and were mapped using the MapChart tool. Standalone BLASTn search was performed using NCBI BLAST+ tool using *T. aestivum* LBD-containing genes under study as query and *T. dicoccoides*, *T. urartu*, and *A. tauschii* LBD-containing genes as database. The obtained hits were filtered with 100% identity for multiple sequence alignment (ClustalW) following phylogenetic analysis (MLM) using MEGAXI (Tamura et al., 2021). The generated Newick file was used for visualizing the tree using the online iTOL tool (https://itol.embl.de/).

2.2.5 Pipeline of 6mA site prediction of LBD-containing genes

The pictorial representation pipeline used for analysis of the promoters of LBD genes is given as Figure 2.

3 Results

3.1 Prediction of 6mA sites

The prediction was performed based on both rice and *Arabidopsis*. The results are as follows.

3.1.1 Feature space analysis

The feature sets DNF, NCP, AMIP, log-transformed GC content, and MBED resulted in a total vector space of 124 dimensions. The top 40 features were taken from each feature selection module (SwR and RF) (Chen and Howard, 2015), and common features predicted in each module were taken for model development. We found that the final feature space consisted of three features from DNF, eight from NCP, seven from AMIP, one from log-transformed GC content, and nine from MBED, which makes a total feature space of 28 dimensions. Clearly, it can be seen that MBED has the highest contribution in terms of constructing the final feature space, followed by NCP.

3.1.2 Prediction of 6mA sites

Testing was performed using the Nipponbare dataset. It was found that the RF performed better, in terms of accuracy, sensitivity, specificity, MCC, and AUC, than SVM and GB in both cases, i.e., when trained with the rice dataset (Supplementary Table 1) and when trained with the *Arabidopsis* dataset (Supplementary Table 2). The ensemble model also performed better than the existing models in terms of accuracy, specificity, and MCC, while Meta-i6mA exhibited more sensitivity (Figure 3A; Supplementary Table 3). Also, in terms of AUC, MethSemble-6mA outperformed the other two state-of-the-art tools (Figure 3B). To ascertain the superiority of the MethSemble-6mA model, TOPSIS analysis was performed.

3.1.3 TOPSIS analysis

The performances of multiple tools, assessed on various evaluation criteria, were tested using the TOPSIS method, which aids in multiple-criteria decision-making (MCDM) (Barretta et al., 2023). The R-package "topsis" was employed for this analysis. This approach enabled the determination of the best-ranked tools based on their similarity to the ideal solution across the multiple criteria considered. MethSemble-6mA secured the top rank followed by Meta-i6mA and EpiSemble (Table 3). The TOPSIS analysis was performed based on the evaluation measures.

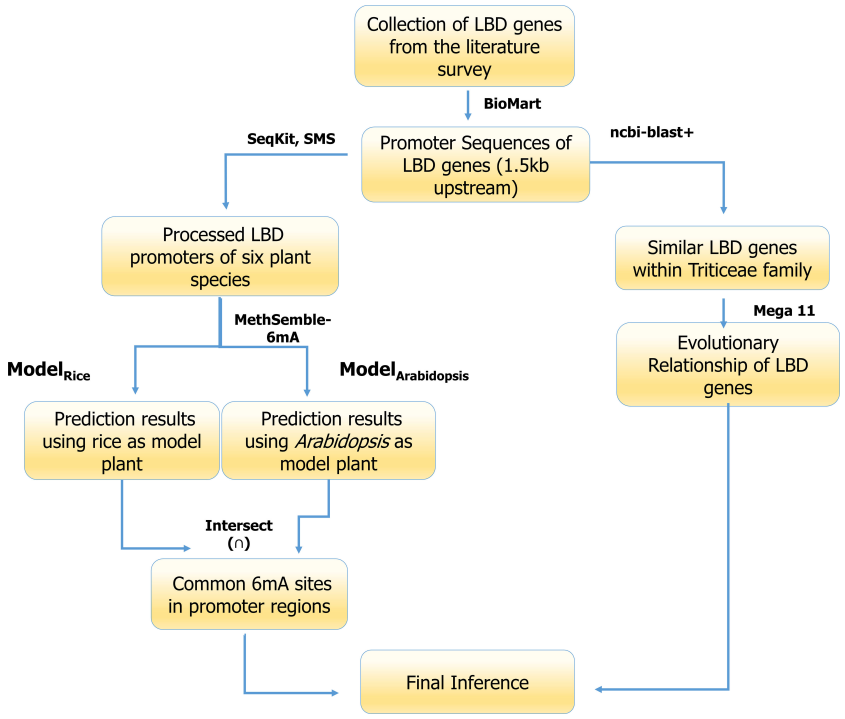


FIGURE 2
6mA methylation site prediction pipeline used in the current study.

TABLE 3 Ranking of the models using the TOPSIS method.

	Score	Rank
MethSemble-6mA	**0.67**	**1**
Meta-i6mA	0.59	2
EpiSemble	0.20	3

Output of our tool showing comparative superior performance highlighted as bold.

3.1.4 MethSemble-6mA server

A user-friendly server was built for hassle-free implementation of the model. The interface is given in Figure 4.

- File format: Provide the input file in fasta format.
- Sequence details: The sequence length of the input must be exactly 41nt, and it should not contain any ambiguous bases ("N").
- Number of inputs: Users can provide multiple sequences in a multifasta file.
- Accession name: The accession name must be different for each sequence in the multifasta file.

3.2 6mA analysis of promoter region of LBD genes

3.2.1 6mA prediction in LBD gene promoters

After the fragmentation and removal of ambiguous sequences, a total of 9,614 fragments were obtained from the selected promoters

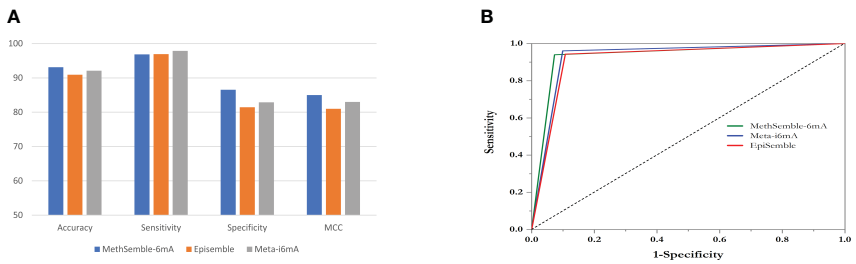


FIGURE 3
Performance of MethSemble-6mA with existing models. **(A)** Bar Plot of Performance Metrics **(B)** ROC Curve.

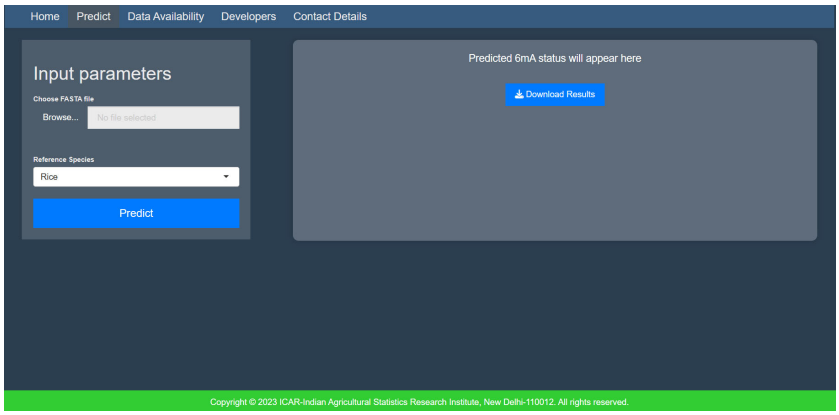


FIGURE 4
Homepage of MethSemble-6mA.

of 279 LBD genes. When prediction was performed by taking rice as a model plant, 1,246 fragments out of 9,614 fragments were found to be methylated. However, when prediction was performed by taking *Arabidopsis* as a model plant, 1,173 fragments out of 9,614 fragments were found to be methylated. After taking the common methylation sites, a total number of 1026 methylated fragments were obtained. Species-wise frequency of 6mA sites is given in Table 4.

3.2.2 Distribution pattern of the 6mA sites in LBD gene promoters

To study the methylation pattern and number of frequencies of 6mA sites, we plotted a histogram, which represents the frequency of the number of promoters that contain a certain number of 6mA sites. We found that most promoters contain three 6mA sites, while the highest number of 6mA sites was found in the promoter region of *Aegilops tauschii* (AET1Gv20706400). The species-wise distribution is given in Figure 5. The enrichment analysis of the LBD genes having 6mA sites yielded position-specific nucleotide abundance (Supplementary Figure 1).

3.2.3 Map of 6mA sites predicted in the promoter regions of selected LBD-containing genes

6mA sites at upstream promoter regions of the LBD-containing genes under study with the highest occurring frequency were plotted for visualization. The highest frequency obtained was as follows: *O. sativa* (BGIOSGA001373; 8), *A. thaliana* (AT4G37540, AT5G66870, AT4G22700, AT3G27650, and AT2G30340; 5), *T. aestivum* (TraesCS5A02G284000 and TraesCS3D02G340000; 10), *T. urartu* (TuG1812S0002083700.01, TuG1812G0100002612.01, and TuG1812G0500005367.01; 5), *T. dicoccoides* (TRIDC2BG028050 and TRIDC6BG069150; 8), and *A. tauschii* (AET1Gv20706400; 12).

TABLE 4 Frequency of 6mA sites in upstream promoter regions of LBD-containing genes in selected species under study.

Species	Frequency of 6mA sites
Triticum aestivum	322
Triticum dicoccoides	188
Triticum urartu	145
Aegilops tauschii	134
Oryza sativa	126
Arabidopsis thaliana	134

3.2.4 Phylogenetic analysis of LBD-containing genes of *Triticum* species under study

We shortlisted 100% identity genes with LBD domain identified using BLASTn search. A phylogenetic tree was generated to interpret the evolutionary conservation of 6mA methylation pattern over the course of evolution and speciation in *Triticum* (Figure 6). We classified the tree with five clades and analysed for 6mA site conservation. *T. urartu* had the least commonality with other wheat species under study (Figure 6). We took a few clade genes for further analysis of the 6mA position and details are given in the discussion section.

4 Discussion

LBD genes play a crucial role in plant developmental biology. To our knowledge, there are a few tools available to date that work on multiple species. Our proposed model, MethSemble-6mA, was trained based on both rice and *Arabidopsis* datasets. This model was found to be efficient in testing cross-species based on evaluation measures like accuracy and specificity (Supplementary Tables 1, 2). Adding a robust feature like MBED helps to identify the 6mA sites in cross-species. In the case of *Arabidopsis*, the sensitivity is relatively low (Supplementary Table 2), which reflects that, although it can detect true-positive (TP) instances, improvement can be performed to decrease false-positive (FP) instances. In our opinion, TP poses a particular pattern in methylation states, as it is not random, and true negative (TN) instances lack that. Further investigation can be performed to reduce the FP instances to make

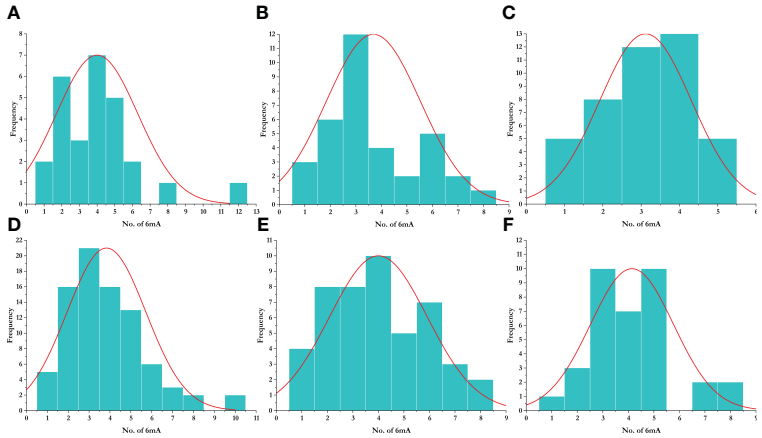


FIGURE 5
Distribution of 6mA sites in different species. **(A)** *Aegilops tauschii*, **(B)** *Arabidopsis thaliana*, **(C)** *Oryza sativa*, **(D)** *Triticum aestivum*, **(E)** *Triticum dicoccoides*, and **(F)** *Triticum urartu*.

the model more robust. However, MethSemble-6mA is still outperforming other existing cross-species tools in terms of accuracy, specificity, and MCC. TOPSIS analysis of the models based on the evolutionary measures also depicts the same outcome.

Employing Methsemble-6mA, we predicted the 6mA sites based on both the model plant species and took the common sites having the 6mA sites. This will reduce the FP instances but will provide a more precise outcome about the probable 6mA sites. We found that the methylation rate per gene in the Triticeae family (greater than 3.83) is higher than in *Oryza sativa i.e.,* 3.68 and in *Arabidopsis thaliana i.e.,* 3.11. We also observed that, with evolution, the rate of methylation is decreasing within Triticeae family; as we can see the rate of methylation in *Triticum urartu i.e.,* 4.14, and *Aegilops tauschii i.e.,* 4.0 (wild type wheat) is higher than the *Triticum*

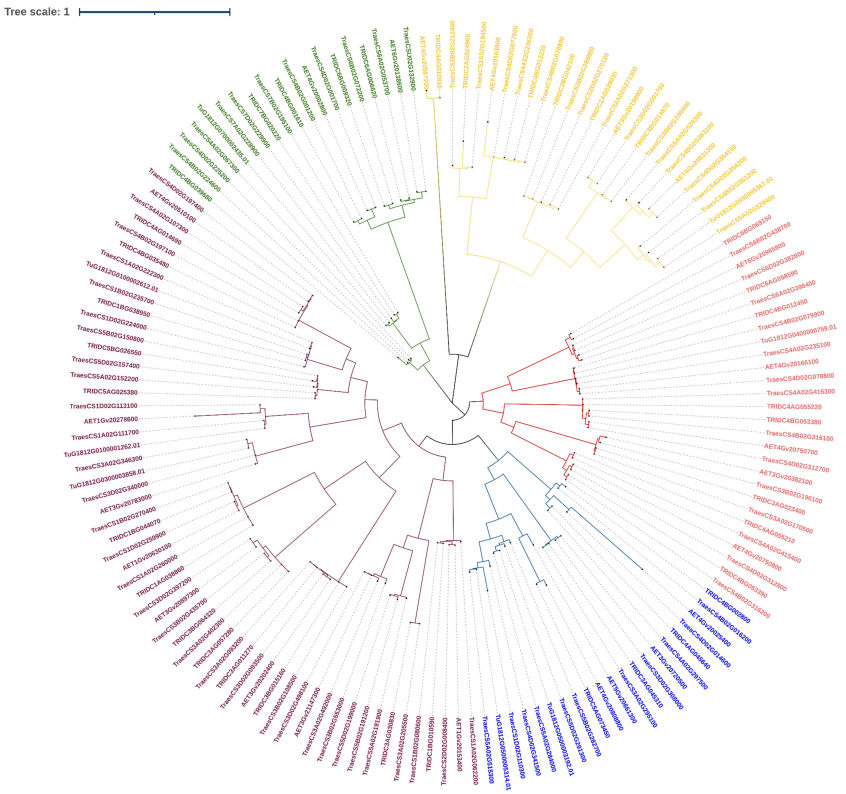


FIGURE 6
Phylogeny of LBD genes of *Triticum* species.

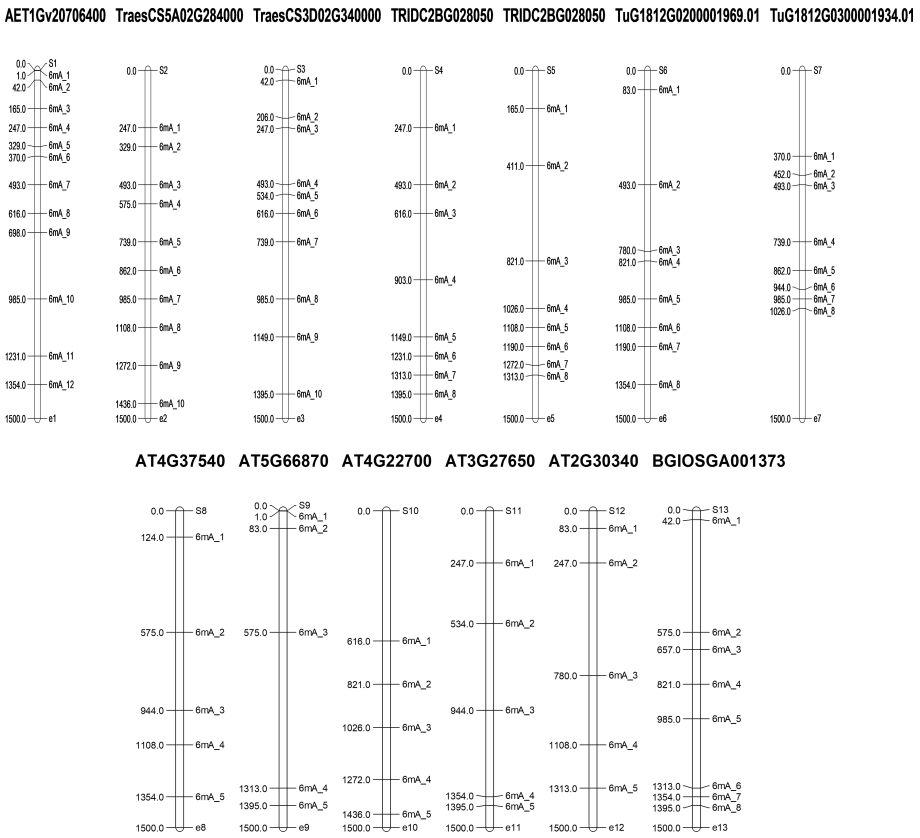


FIGURE 7
Map of 6mA sites predicted in the promoter regions of selected LBD-containing genes.

dicoccoides (cultivated wheat ancestors) *i.e.,* 3.91 and *Triticum aestivum i.e.,* 3.83 (cultivated wheat). It will be interesting to further analyse correlation between ploidy and methylation rate.

In addition to the species-specific differences, we also observed that the methylation rate is much higher in the 700–1,300-bp upstream (from TSS) compared to other parts of the promoter (Figure 7; Supplementary Material 1) except for AET1Gv20706400, where initial 700 bases are rich in 6mA sites and TraesCS5A02G284000, where 6mA sites are evenly distributed throughout the promoter region.

In order to validate our findings, we selected closely related LBD gene pairs under study were selected. These were obtained through all against all blast hits with 100% identity and their predicted 6mA sites in promoter regions were analyzed critically. It was observed that in wheat, with polyploidization, speciation and domestication, 6mA methylation sites have reduced. Moreover, similarity in a few methylation sites also indicates their conservation across species. For example, two phylogenetically closed genes, TRIDC6BG069150 (from *T. dicoccoides*) and TraesCS6B02G438700 (from *T. aestivum*), were analyzed, and it was found that the promoter of the former has eight 6mA sites, the promoter of the latter one has five sites, and two sites, 1026-1066 and 1313-1353, were conserved for both the genes. Again, from the same clade, AET6Gv20965900 (from *A. tauschii*) promoter has six 6mA sites of which 1026-1066 site is conserved in all these three species. Similar inferences can be made using other genes from another clade. For example, promoters of TraesCS3B02G108500, TRIDC3BG015160, and AET3Gv20202400 genes have four, four, and one 6mA sites, respectively, while the first and last two genes have a conserved site at base position at 165-205 and 1313-1353, respectively. This again indicates the conservation of 6mA sites across the domestication of wheat, and it can be an important finding to understand speciation, domestication and gene evolution over time. Moreover, it help understand how gene regulation has changed through changing methylation sites through speciation. Ultimate objective is to identify screening strategy to select climate ready genotypes.

5 Conclusion

Targeting the methylation sites in gene regulatory elements to investigate gene expression patterns and genome imprinting mechanisms is always intriguing to obtain more robust information about functional epigenetic sites in the genome. An improved model for predicting 6mA sites, more specifically for Poaceae and Brassicaceae family crops, has been delivered to be utilized in molecular biology research. Nevertheless, a dedicated user-friendly server was developed for easy implementation of the proposed model. While analyzing our results, we found that through speciation and evolution, gene methylation (6mA) in regulatory sequences has changed (reduced frequency of occurrence). This could be an interesting and fundamental factor

to be investigated thoroughly to answer the mechanism of gene birth as well as gene regulation. Moreover, the epigenetic control of growth and development along with stress tolerance and disease resistance could be addressed in the near future using the results obtained from our proposed model.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

DS: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Writing – original draft. TD: Data curation, Formal Analysis, Methodology, Software, Writing – original draft. KP: Data curation, Formal Analysis, Writing – original draft. MY: Formal Analysis, Software, Writing – original draft. SB: Conceptualization, Data curation, Formal Analysis, Writing – review & editing. SM: Conceptualization, Formal Analysis, Writing – original draft. DM: Conceptualization, Funding acquisition, Methodology, Writing – review & editing. SP: Software, Writing – review & editing. AR: Writing – review & editing. SA: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Funding

The study was partly supported by the ICAR-National Fellow Project on PGR Informatics (grant no. 1006528).

Acknowledgments

The authors acknowledge the facilities provided by the ICAR-Indian Agricultural Statistics Research Institute, ICAR-National Bureau of Plant Genetic Resources, and ICAR-Indian Agricultural Research Institute. Sunil Archak was supported by the ICAR National Fellow Project. Dipro Sinha was supported by the ICAR-Indian Agricultural Statistics Research Institute fellowship.

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.

The handling editor SD declared a past co-authorship with the author(s) DM at the time of review.

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/fpls.2023.1256186/full#supplementary-material

References

Ashapkin, V. V., Kutueva, L. I., Aleksandrushkina, N. I., and Vanyushin, B. F. (2020). Epigenetic mechanisms of plant adaptation to biotic and abiotic stresses. *Int. J. Mol. Sci.* 21 (20), 1–32. doi: 10.3390/IJMS21207457

Barretta, R., Taherdoost, H., and Madanchian, M. (2023). Multi-criteria decision making (MCDM) methods and concepts. *Encyclopedia* 3 (1), 77–87. doi: 10.3390/ENCYCLOPEDIA3010006

Basith, S., Manavalan, B., Shin, T. H., and Lee, G. (2019). SDM6A: A web-based integrative machine-learning framework for predicting 6mA sites in the rice genome. *Mol. Ther. - Nucleic Acids* 18, 131–141. doi: 10.1016/j.omtn.2019.08.011

Bauer, M., Schuster, S. M., and Sayood, K. (2008). The average mutual information profile as a genomic signature. *BMC Bioinf.* 9 (1), 1–11. doi: 10.1186/1471-2105-9-48

Breiman, L. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32. doi: 10.1023/A:1010933404324

Campbell, J. L., and Kleckner, N. (1990). *E. coli* oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* 62 (5), 967–979. doi: 10.1016/0092-8674(90)90271-F

Chen, F. H., and Howard, H. (2015). An alternative model for the analysis of detecting electronic industries earnings management using stepwise regression, random forest, and decision tree. *Soft Computing* 20 (5), 1945–1960. doi: 10.1007/S00500-015-1616-6

Chen, W., Lv, H., Nie, F., and Lin, H. (2019). i6mA-Pred: identifying DNA N6-methyladenine sites in the rice genome. *Bioinf. (Oxford England)* 35 (16), 2796–2800. doi: 10.1093/BIOINFORMATICS/BTZ015

Cortes, C., Vapnik, V., and Saitta, L. (1995). Support-vector networks. *Mach. Learn.* 20 (3), 273–297. doi: 10.1007/BF00994018

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., et al. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7 (6), 461–465. doi: 10.1038/NMETH.1459

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Stat* 29 (5), 1189–1232. doi: 10.1214/AOS/1013203451

Guan, C., Wu, B., Ma, S., Zhang, J., Liu, X., Wang, H., et al. (2023). Genome-wide characterization of LBD transcription factors in switchgrass (*Panicum virgatum* L.) and the involvement of PvLBD12 in salt tolerance. *Plant Cell Rep.* 42 (4), 735–748. doi: 10.1007/S00299-023-02989-9/METRICS

Hasan, M. M., Basith, S., Khatun, M. S., Lee, G., Manavalan, B., and Kurata, H. (2021). Meta-i6mA: an interspecies predictor for identifying DNA N6-methyladenine sites of plant genomes by exploiting informative features in an integrative machine-learning framework. *Briefings Bioinf.* 22 (3), 1–16. doi: 10.1093/BIB/BBAA202

Huang, Q., Zhang, J., Wei, L., Guo, F., and Zou, Q. (2020). 6mA-ricePred: A method for identifying DNA N 6-methyladenine sites in the rice genome based on feature fusion. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.00004/BIBTEX

Jiao, P., Wei, X., Jiang, Z., Liu, S., Guan, S., and Ma, Y. (2022). Zmlbd2 A maize (*Zea mays* L.) lateral organ boundaries domain (LBD) transcription factor enhances drought tolerance in transgenic *arabidopsis thaliana*. *Front. Plant Sci.* 13. doi: 10.3389/FPLS.2022.1000149/BIBTEX

Kha, Q. H., Ho, Q. T., and Le, N. Q. K. (2022). Identifying SNARE proteins using an alignment-free method based on multiscan convolutional neural network and PSSM profiles. *J. Chem. Inf. Modeling* 62 (19), 4820. doi: 10.1021/ACS.JCIM.2C01034

Kong, L., and Zhang, L. (2019). i6mA-DNCP: computational identification of DNA N6-methyladenine sites in the rice genome using optimized dinucleotide-based features. *Genes* 10 (10), 1–13. doi: 10.3390/GENES10100828

Krais, A. M., Cornelius, M. G., and Schmeiser, H. H. (2010). Genomic N6-methyladenine determination by MEKC with LIF. *ELECTROPHORESIS* 31 (21), 3548–3551. doi: 10.1002/ELPS.201000357

Liu, H., Cao, M., Chen, X., Ye, M., Zhao, P., Nan, Y., et al. (2019). Genome-wide analysis of the lateral organ boundaries domain (LBD) gene family in *solanum tuberosum*. *Int. J. Mol. Sci.* 20 (21), 1–23. doi: 10.3390/IJMS20215360

Liu, L., Zhang, J., Xu, J., Li, Y., Guo, L., Wang, Z., et al. (2020). CRISPR/cas9 targeted mutagenesis of sllbd40, A lateral organ boundaries domain transcription factor, enhances drought tolerance in tomato. *Plant Sci.* 301, 110683. doi: 10.1016/J.PLANTSCI.2020.110683

Lv, H., Dao, F. Y., Guan, Z. X., Zhang, D., Tan, J. X., Zhang, Y., et al. (2019). iDNA6mA-rice: A computational tool for detecting N6-methyladenine sites in rice. *Front. Genet.* 10. doi: 10.3389/FGENE.2019.00793

Lv, H., Dao, F. Y., Zhang, D., Guan, Z. X., Yang, H., Su, W., et al. (2020). iDNA-MS: an integrated computational tool for detecting DNA modification sites in multiple genomes. *IScience* 23 (4), 100991. doi: 10.1016/J.ISCI.2020.100991

O'Brown, Z. K., and Greer, E. L. (2016). N6-methyladenine: A conserved and dynamic DNA mark. *Adv. Exp. Med. Biol.* 945, 213–246. doi: 10.1007/978-3-319-43624-1_10

Pomraning, K. R., Smith, K. M., and Freitag, M. (2009). Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods (San Diego Calif.)* 47 (3), 142–150. doi: 10.1016/J.YMETH.2008.09.022

Pukkila, P. J., Peterson, J., Herman, G., Modrich, P., and Meselson, M. (1983). Effects Of High Levels of DNA adenine methylation on Methyl-Directed Mismatch Repair in. *Escherichia coli. Genet.* 104 (4), 571–582. doi: 10.1093/GENETICS/104.4.571

Ratel, D., Ravanat, J. L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *BioEssays: News Rev. Molecular Cell. Dev. Biol.* 28 (3), 309–315. doi: 10.1002/BIES.20342

Rehman, M. U., Tayara, H., Zou, Q., and Chong, K. T. (2022). i6mA-caps: A capsulenet-based framework for identifying DNA N6-methyladenine sites. *Bioinformatics* 38 (16), 3885–3891. doi: 10.1093/BIOINFORMATICS/BTAC434

Robbins-Manke, J. L., Zdraveski, Z. Z., Marinus, M., and Essigmann, J. M. (2005). Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient *escherichia coli*. *J. Bacteriology* 187 (20), 7037. doi: 10.1128/JB.187.20.7027-7037.2005

Saraswat, S., Yadav, A. K., Sirohi, P., and Singh, N. K. (2017). Role of epigenetics in crop improvement: water and heat stress. *J. Plant Biol.* 3 (60), 231–240. doi: 10.1007/S12374-017-0053-8

Sinha, D., Dasmandal, T., Yeasin, MD., Mishra, D. C., Rai, A., and Archak, S. (2023). EpiSemble: A novel ensemble-based machine-learning framework for prediction of DNA N6-methyladenine sites using hybrid features selection approach for crops. *Curr. Bioinf.* 18 (7), 587–597. doi: 10.2174/1574893618666230316151648

Sinha, D., Sharma, A., Mishra, D. C., Rai, A., Lal, S. B., Kumar, S., et al. (2022). MetaConClust - unsupervised binning of metagenomics data using consensus clustering. *Curr. Genomics* 23 (2), 137. doi: 10.2174/1389202923666220413114659

Tahir, M., Tayara, H., and Chong, K. T. (2019). iDNA6mA (5-step rule): Identification of DNA N6-methyladenine Sites in The Rice Genome by Intelligent Computational Model via Chou's 5-step Rule. *Chemometrics Intelligent Lab. Syst.* 189, 96–101. doi: 10.1016/J.CHEMOLAB.2019.04.007

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38 (7), 3022–3027. doi: 10.1093/MOLBEV/MSAB120

Waddington, C. H. (2012). The epigenotype. 1942. *Int. J. Epidemiol.* 41 (1), 10–13. doi: 10.1093/IJE/DYR184

Wang, X., and Yan, R. (2018). RFAthM6A: A new tool for predicting M6a sites in arabidopsis thaliana. *Plant Mol. Biol.* 96 (3), 327–337. doi: 10.1007/S11103-018-0698-9/TABLES/5

Wang, Y., Zhang, P., Guo, W., Liu, H., Li, X., Zhang, Q., et al. (2021). A deep learning approach to automate whole-genome prediction of diverse epigenomic modifications in plants. *New Phytol.* 232 (2), 880–897. doi: 10.1111/NPH.17630

Wang, Z., Zhang, R., Cheng, Y., Lei, P., Song, W., Zheng, W., et al. (2021). Genome-wide identification, evolution, and expression analysis of LBD transcription factor family in bread wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.721253/BIBTEX

Xu, J., Hu, P., Tao, Y., Song, P., Gao, H., and Guan, Y. (2021). Genome-wide identification and characterization of the lateral organ boundaries domain (Lbd) gene family in polyploid wheat and related species. *PeerJ* 9, e11811. doi: 10.7717/PEERJ.11811/SUPP-10

Yu, H., and Dai, Z. (2019). SNNRice6mA: A deep learning method for predicting DNA N6-methyladenine sites in rice genome. *Front. Genet.* 10. doi: 10.3389/FGENE.2019.01071/BIBTEX

Yu, J., Xie, Q., Li, C., Dong, Y., Zhu, S., and Chen, J. (2020). Comprehensive characterization and gene expression patterns of LBD gene family in gossypium. *Planta* 251 (4), 1–16. doi: 10.1007/S00425-020-03364-8/FIGURES/8

Zhao, D., Chen, P., Chen, Z., Zhang, L., Wang, Y., and Xu, L. (2023). Genome-wide analysis of the LBD family in rice: gene functions, structure and evolution. *Comput. Biol. Med.* 153, 106452. doi: 10.1016/J.COMPBIOMED.2022.106452

Zhou, C., Wang, C., Liu, H., Zhou, Q., Liu, Q., Guo, Y., et al. (2018). Identification and analysis of adenine N 6-methylation sites in the rice genome. *Nat. Plants* 4 (8), 554–563. doi: 10.1038/S41477-018-0214-X

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 November 2023
DOI 10.3389/fpls.2023.1267281

OPEN ACCESS

EDITED BY
Rongrong Zhai,
Zhejiang Academy of Agricultural Sciences, China

REVIEWED BY
Changquan Zhang,
Yangzhou University, China
Xiangjin Wei,
China National Rice Research Institute (CAAS), China

*CORRESPONDENCE
Xiaoli Shu
shuxl@zju.edu.cn

†These authors have contributed equally to this work

RECEIVED 26 July 2023
ACCEPTED 16 October 2023
PUBLISHED 03 November 2023

CITATION
Luo M, Gong W, Zhang S, Xie L, Shi Y, Wu D and Shu X (2023) Discrepancies in resistant starch and starch physicochemical properties between rice mutants similar in high amylose content.
Front. Plant Sci. 14:1267281.
doi: 10.3389/fpls.2023.1267281

Discrepancies in resistant starch and starch physicochemical properties between rice mutants similar in high amylose content

Mingrui Luo[1,2†], Wanxin Gong[1†], Siyan Zhang[1], Lanyu Xie[2], Yitao Shi[2], Dianxing Wu[1,3] and Xiaoli Shu[1,3*]

[1]State Key Laboratory of Rice Biology, Key Laboratory of the Ministry of Agriculture and Rural Affairs for Nuclear Agricultural Sciences, Zhejiang University, Hangzhou, China, [2]Life Science and Technology Center, China National Seed Group Co., Ltd., Wuhan, China, [3]Hainan Institute of Zhejiang University, Yazhou Bay Science and Technology City, Sanya, China

The content of resistant starch (RS) was considered positively correlated with the apparent amylose content (AAC). Here, we analyzed two Indica rice mutants, RS111 and Zhedagaozhi 1B, similar in high AAC and found that their RS content differed remarkably. RS111 had higher RS3 content but lower RS2 content than Zhedagaozhi 1B; correspondingly, cooked RS111 showed slower digestibility. RS111 had smaller irregular and oval starch granules when compared with Zhedagaozhi 1B and the wild type. Zhedagaozhi 1B showed a B-type starch pattern, different from RS111 and the wild type, which showed A-type starch. Meantime, RS111 had more fa and fb1 but less fb3 than Zhedagaozhi 1B. Both mutants showed decreased viscosity and swelling power when compared with the parents. RS111 had the lowest viscosity, and Zhedagaozhi 1B had the smallest swelling power. The different fine structures of amylopectin between RS111 and Zhedagaozhi 1B led to different starch types, gelatinization properties, paste viscosity, and digestibility. In addition to enhancing amylose content, modifications on amylopectin structure showed great potent in breeding rice with different RS2 and RS3 content, which could meet the increasing needs for various rice germplasms.

KEYWORDS

Rice (*Oryza sativa* L.), resistant starch, amylose, amylopectin, physiochemical properties

Introduction

Resistant starch (RS) is a kind of starch and starch degradation products that are resistant to be digested in the small intestine but can be fermented in the large intestine by microbial flora to produce metabolites such as short-chain fatty acids (SCFAs). RS plays a beneficial role in controlling blood sugar and regulating intestinal flora (Englyst et al., 1992; Birt et al., 2013). With the development of living standards and the change in lifestyle, the number of the type 2 diabetics is increasing. High RS diet can help to prevent diabetes and

reduce calorie intake, which may be helpful for weight management (Birt et al., 2013). Therefore, it is essential to develop functional foods rich in RS. According to the feature and botanical origin, RS can be classified into five subtypes, namely, RS1, RS2, RS3, RS4, and RS5. RS2 is made up of native starch granules and usually presents in raw food such as green bananas and raw potatoes (Englyst and Cummings, 1987; Faisant et al., 1995). RS3 consists of retrograded starch, mainly the recrystallized amylose, and is formed during the cooling and retrogradation of gelatinized starch and cooked foods stored at room or low temperature (Noah et al., 1998).

Rice is the staple food for more than half of the population worldwide and the primary source of nutrition and carbohydrates for many people. RS2 and RS3 is the main RS type that exists in raw and cooked rice, respectively (Ding et al., 2019). Hot cooked rice is regarded as a typically high-GI food, generally contained low RS content with less than 3% (Frei et al., 2003), which is not enough to confer the health benefits (Shu et al., 2009). Breeding rice high in RS is an effective way to improve public health. Previous studies have found that amylose content, the ratio of amylose to amylopectin and fine structure of amylopectin, had obvious impacts on RS content (Sajilata et al., 2006; Chung et al., 2011; Zhu et al., 2011). Several genes related to starch synthesis have been found to regulate RS formation in rice. *Waxy* (*Wx*), which encodes granule-bound starch synthase (GBSS) I, can regulate RS formation through impacting amylose synthesis (Jeon et al., 2010). *BEIIb* plays an important role in forming short amylopectin chains and the deficiency of *BEIIb* (*ae* mutant) in rice can significantly increase in RS content by increasing the number of B2 and B3 branched chains (Butardo et al., 2011; Abe et al., 2014; Nakata et al., 2018; Baysal et al., 2020). In addition, *SSIIIa* involved in the elongation of the B2 and B3 chains in amylopectin (Fujita et al., 2007) has been verified to be a critical gene responsible for the RS synthesis in rice, loss of function of *SSIIIa* gives rise to a high RS content in rice grain (Zhou et al., 2016). These results provided opportunities to increase RS content of rice through genetic improvement.

Up to date, enhancing amylose content is an important approach to develop high-RS rice and functional foodstuffs (Shen et al., 2022a). Some mutants or varieties with enhanced amylose content also had elevated RS, such as RS111 (Yang et al., 2006), Gaomi 2 (Kang et al., 2003), Jiangtangdao 1 (Yang et al., 2012), TRS (Wei et al., 2010), and b10 (Zhou et al., 2016). However, the palatability of rice high in RS with high amylose content were generally poor; moreover, the yield was also reduced (Tao et al., 2019; Miura et al., 2021). Thus, balancing the quality, yield, and RS content is important for high-RS rice breeding. Improving the quality of restorer and sterile lines is the key for traditional rice hybrid breeding, while the quality of sterile is majorly determined by its maintainer. In previous design breeding, we simultaneously improved restorer R7954 and maintainer II-32B, both parents of super hybrid rice IIyou7954 (II-32A×R7954), and isolated and developed two high RS mutants, RS111 (*ssIIIa* mutant) and Zhedagaozhi 1B (unknown mutant), respectively, which were similar in high amylose but different in RS properties. Clarifying discrepancies in RS and starch physicochemical properties of rice similar in high amylose content will provide some guidance for high-RS rice breeding and diversified healthy and functional foods.

Materials and methods

Rice materials

Elite restorer R7954 and leading commercial maintainer II-32B (Shu et al., 2007), high-RS mutants RS111, and Zhedagaozhi 1B were used in this study. RS111 and Zhedagaozhi 1B were derived, respectively, from R7954 and II-32B by irradiated with 300 Gy ^{60}Co-γ rays. RS111 is an *ssIIIa* mutant (Zhou et al., 2022), while Zhedagaozhi 1B is not (data not provided) and might be a novel high AAC/RS mutant. All accessions were grown in the experiment farms of Zhejiang University (Hangzhou, China, 120.2E, 30.3N) in June 2022. The mature seed were harvested in late October 2022.

Preparation of rice flour and starch

The rice grains were air-dried to achieve the moisture content of about 12%. The samples were dehulled using a Satake Rice Machine (Satake Co., Tokyo, Japan), milled to white rice using a Satake Rice Machine (Satake Co., Hiroshima, Japan), and then ground into flour (CT 293, Foss, Sweden) and passed through a 100-mesh sieve.

Starch was extracted according to a previous method described by Yang et al. (2006) with minor modifications. Briefly, rice flour (30 g) was suspended with 150 mL 0.2% sodium hydroxide solution and then shaken for 12h. The starch fraction was recovered by centrifugation at 4000 rpm for 10 min and repetitive scrubbed until the pH reached 7.0. The starch was dried at 40°C in a vacuum oven for 72h, passed through a 150-μm sieve and stored in a drier till used.

Apparent amylose content, total starch, resistant starch, and lipid content

Apparent amylose content (AAC) was determined by the simplified assay as Yang et al. (2006). AAC standard samples (2%, 8%, 17%, 22%, and 28.5%) were provided by the China National Rice Research Institute. Total starch (TS) content was also measured according to the method described by Yang et al. (2006). The contents of RS2 in raw sample and RS3 in cooked sample were measured as the previous studies (Gong et al., 2021). Free lipid was measured by Hanon SOX406 fat analyzer (Hanon Group, Jinan, China) following the principle of Soxhlet extraction method [extraction solvent, petroleum ether (AR, bp, 30–60°C)].

Scanning electron microscopy

Starch sample was directly adhered to double-sided adhesive tape mounted on an aluminum stub and coated under vacuum with platinum for 50 s (IB-5 ion coater, Eiko Co.). All coated samples were observed with a Hitachi SU8010 cold field emission SEM (Tokyo, Japan) operated at 3.0 kV.

X-ray diffraction

X-ray diffraction pattern was performed with copper Kα radiation on an x-ray diffractometer (D8 Advance, Bruker, Germany). The scanning regions of the diffraction angle 2θ were 4°–40°, which covered most of the significant diffraction peaks of the starch crystallites, with a 0.02°-step size and a count time of 0.2 s. The crystallinity was calculated according to the method of Hayakawa et al. (1997).

Particle size distribution

Granule size of starches was detected using a laser light scattering particle size analyzer (LS13320, Beckman Coulter, California, USA). Starch (50 mg) was mixed with 15 mL of ddH_2O and then sonicated for 5 min. The sample was loaded to the sample port by drops until 8%–12% obscuration.

Fourier transform infrared spectroscopy

Starch granule short-range ordered structure was detected with the Fourier transform infrared (FTIR) spectrometer (NICOLET iS50FT-IR, Thermo Fisher Scientific, USA). The scanning range was from 4,000 cm^{-1} to 400 cm^{-1} with 32 scans at 4 cm^{-1}. FTIR spectra were treated in peak half-width of 19 cm^{-1} and the resolution enhancement factor of 1.9. Then the relative absorbances at 1,045, 1,022, and 995 cm^{-1} were extracted from the deconvoluted spectra and were measured from the baseline to the peak height according with Man et al. (2012).

Amylopectin fraction and amylopectin chain-length distribution determination

Amylopectin fractionation was conducted as Kong et al. (2008). The chain-length distribution of the de-branched sample was analyzed with the Carbo-Pac PA-100 column (250 mm × 4 mm, with a guard column) using a high-performance anion-exchange chromatography system (Dionex ICS-5000+, Sunnyvale, CA, USA) coupled with a BioLC gradient pump and a pulsed amperometric detector. The samples were then eluted with a flow rate of 1 mL/min. The eluents used were A (150 mM sodium hydroxide) and B (150 mM sodium hydroxide containing 500 mM sodium acetate). A mixed elution gradient with eluent A and eluent B was set as follows: 0–10 min, 93%–82% A; 10–19 min, 82%–78% A; 19–70 min, 78%–62% A; 70–70.1 min, 62%–50% A; 70.1–85 min, 50% A; 85–85.1 min, 50%–93%. The system was stabilized by elution at 93% A for 5 min between runs.

Thermal properties measured with DSC

Thermal properties of starch were determined by a differential scanning calorimeter (DSC) (Q20, TA Instruments, New Castle, DE, USA) equipped with DSC standard and dual sample cells. Starch (2.0 mg, dry weight basis) was weighed into an aluminum pan, and 6 μL of MilliQ water was added (Kong et al., 2015). The pan was hermetically sealed and equilibrated at room temperature for 12h, then scanned from 40°C to 11°C at a heating rate of 10°C/min with an empty sealed pan as a reference. The onset gelatinization temperature (GT) (*To*), peak temperature (*Tp*), conclusion temperature (*Tc*), and enthalpy change (ΔH) were analyzed with a Universal Analysis Program, Version 4.4A. The range of GT was calculated as $R = Tc - Tp$.

Gelatinization properties in urea

Gelatinization properties was determined according to the method of Nishi et al. (2001) with minor modifications. Briefly, rice flour (10 mg) was mixed with 1 mL of 5M urea in an Eppendorf tube and incubated at 25°C for 24h. After centrifugation, sample was allowed to stand for 1h. Swelling power was calculated by subtracting the volume of supernatant from the urea solution (1 mL).

Pasting viscosity

Pasting properties of rice flour were analyzed using a Rapid Visco Analyzer (RVA Tecmaster, Perten, Hägersten, Sweden) according to the methods described by Yang et al. (2006). The peak viscosity (PV), trough viscosity (TV), final viscosity (FV), and paste temperature (PT) were recorded; breakdown viscosity (BV = PV − TV) and setback viscosity (SV = FV − TV) were calculated.

In vitro digestion study

Digestion property was studied using a standard protocol as reported by Shen et al. (2022b) with minor modifications. Rice flour of 0.5 g was mixed with 4.5 mL of water and boiled for 30 min to ensure full gelatinization. After cooling to the room temperature, the rice was homogenized for 30 s using an Ultra Turrax T25 IKA Homogenizer to mimic human chewed. After blending, crushed sample was digested through three phases (oral, gastric, and intestinal phase). During the simulated intestinal phase, totally 10 time points (0, 10, 20, 30, 45, 60, 90, 120, 150, and 180 min) was chosen for sampling. The digested starch content in the supernatant was determined by glucose analysis. Uncooked rice flour was digested in the same method above but without cooking.

The percentage of the starch hydrolysis was calculated by the following equation (1):

$$SH(\%) = \frac{Sh}{Si} = \frac{Gp}{Si} \times 0.9 \quad (1)$$

Where SH is the percentage of starch hydrolysis, Sh is the amount of hydrolyzed starch at different time points in intestinal phase, Si is the initial amount of starch in the samples, and Gp is the amount of glucose released due to starch hydrolysis. The conversion factor of 0.9 was used based on the molecular weight ratio of a starch monomer to glucose.

The experiment data were fitted to a first-order model according to equation (2), as previous reported by Goñi et al. (1997):

$$C_t = C_0 + C_\infty(1 - e^{-kt}) \quad (2)$$

where C_t, C_0, and C_∞ are the percentage of digested starch at time t, 0, and infinite time, respectively, and k is a pseudo first-order rate constant. Box Lucas model as the nonlinear curve fit model in OriginPro2022 (Origin Lab corporation, Northampton, MA, USA) was used for estimating k and C_∞ value. Initial digestion rate in the first 20 min (IRR_{20}) was calculated by equation (3) as described by Kan et al. (2020):

$$IRR_{20} = (C_{20} - C_0)/20 \quad (3)$$

where C_{20} and C_0 are the percentage of digested starch at 20 min and 0 min.

Statistical analysis

All the tests were conducted in triplicate at least. The statistics analysis was conducted using Duncan with Tukey's multiple comparison tests following one-way analysis of variance with SPSS version 20.0 (IBM, Chicago, USA).

Results and discussions

Apparent amylose and resistant starch content

The AAC of R7954 and II-32B were both around 25% and significantly lower than those of their mutants. Meantime, R7954 and II-32B had similar RS content (Table 1). The AAC of mutant RS111 and Zhedagaozhi 1B were also similar, which were 33.85% and 34.70%, respectively, indicating that both of RS111 and Zhedagaozhi 1B were typically high amylose mutants when compared to common rice. However, obvious discrepancies in RS contents were found between two mutants. The contents of RS3 in RS111 was 4.78%, significantly higher than that in Zhedagaozhi 1B which was 2.46%, while the content of RS2 in RS111 was only 1.30%, significantly lower than that in Zhedagaozhi 1B which was as high as 17.37%. The contents of RS2 and RS3 in the flour of RS111 and Zhedagaozhi 1B were all significantly higher than those in the parents (Table 1). AAC showed positive correlation to both RS2 and RS3. Several high-amylose rice mutants also showed high-RS contents (Kang et al., 2003; Wei et al., 2010; Zhang et al., 2011; Yang et al., 2012; You et al., 2022).

Considering that other components in the flour might also affect the RS content, the RS content in isolated starch were also determined. The RS2 content in isolated starch was higher than those in four for all samples. The RS3 content in the isolated starches was a little lower than those in flours for R7954, II-32B, and RS111 while higher for Zhedagaozhi 1B. Meantime, the contents of RS3 in mutant RS111 were 3.67%, lower than that in Zhedagaozhi 1B starch (4.06%). The higher RS2 in isolated starch indicated that the other components in the flour might reduce the attack of amylase to the starch granules. Although the difference in RS3 content between starch and flour might because of the existence of lipids in rice flour, Zhedagaozhi 1B had significantly higher lipids content (Table 1). Although it has been reported that lipid can

TABLE 1 Physiochemical properties of flours and starches*.

		RS111	Zhedagaozhi 1B	R7954	II-32B
Flour	Apparent amylose (%)	33.85 ± 0.79a	34.70 ± 0.91a	25.82 ± 0.22b	25.48 ± 0.55b
	RS2 (%)	1.30 ± 0.12b	17.37 ± 0.68a	0.37 ± 0.02c	0.75 ± 0.12c
	RS3 (%)	4.78 ± 0.35a	2.46 ± 0.32b	1.39 ± 0.12c	1.27 ± 0.09c
	Swelling volume (mL)	0.139 ± 0.001c	0.066 ± 0.002d	0.233 ± 0.006a	0.153 ± 0.002b
	Lipids (%)	0.65 ± 0.02a	1.25 ± 0.01b	0.64 ± 0.03a	0.66 ± 0.01a
Starch	RS2 (%)	1.58 ± 0.10b	19.23 ± 0.49a	1.65 ± 0.12b	1.54 ± 0.07b
	RS3 (%)	3.67 ± 0.04b	4.06 ± 0.10a	1.17 ± 0.03c	1.11 ± 0.05c
	Mean diameter (μm)	6.12 ± 0.00d	7.75 ± 0.05b	8.13 ± 0.07a	6.52 ± 0.07c
	Relative crystallinity (%)	19.67 ± 0.28c	19.55 ± 0.19c	24.01 ± 0.22b	28.57 ± 0.38a
	IR ratio 1045/1022 (cm^{-1})	0.58 ± 0.00d	0.68 ± 0.02c	0.79 ± 0.02b	0.85 ± 0.02a
	IR ratio 1022/995 (cm^{-1})	1.86 ± 0.00d	1.60 ± 0.00c	1.84 ± 0.00b	1.99 ± 0.00a

*The results were expressed as mean ± SD (n = 3). The same small letters indicated that there were no significant differences among the starches or flours of rice ($p < 0.05$).

interact with amylose to form starch-lipid complexes and contribute to the RS, the formation of starch-lipid complexes will compete with amylose retrogradation (Cui and Oates, 1999), which would impede retrogradation of the remained amylose and reduced the RS content in turn (Perera et al., 2010). Zhou et al. (2023) also found that the existence of lipids in indica rice with high amylose reduced the RS3 content.

Morphology of starch granules

The four varieties showed obvious different morphology of grain (Figure 1A) and starch granules (Figure 1B). The starch granules of R7954 and II-32B were uniform and almost polygonal, most starch granules of Zhedagaozhi 1B also showed polyhedron, and some were nearly spherical and some with small pinholes on their surfaces, which might be caused by the unsaturated grains. Although, starch granules of RS111 were irregular and small, with some round granules (Figure 1B). Starch with more round and spherical granules were found to have higher RS in cooked rice, while starch with a higher percentage of polyhedral and angular granules usually have higher RS in raw rice (Ding et al., 2019; You et al., 2022). Wei et al. (2010) also found that TRS with higher RS2 had large and non-angular rounded bodies compared with its parent. That was consistent with our results that RS111 with irregular and smaller starch granules had significantly lower content of RS2. However, Zhedagaozhi 1B which had both polyhedral and oval starch granules had significant higher RS2 than all other samples, indicating other structural properties might also be responsible for the difference in RS between RS111 and Zhedagaozhi 1B.

Crystalline structures of rice starch

The external chains of amylopectin interact with each other and water to form crystalline structure. The arrangement of the crystals of the granules gives rise to A, B, and C three types (Buléon et al., 1998). The x-ray diffraction pattern of mutant RS111 and two parents showed typical A-type with strong peaks at around 2θ 15.2°, 17.2°, 18°, 20°, and 23° (Figure 2A). However, the X-ray diffraction (XRD) pattern of Zhedagaozhi 1B showed B-type with the typical peak at 5.6° and weak peak at 18°. The relative crystallinity of mutant RS111 was 19.67%, similar with that of Zhedagaozhi 1B (19.55%), while significantly lower than those of their parents R7954 (24.01%) and II-32B (28.57%). It has been found that B-type crystalline structure could form large "blocklets" bound in crystalline layers of granules (Guo et al., 2017); the dense internal structure can enhance the resistance of digestion to a certain extent. Starch with B-type diffraction pattern (ungelatinized) have been reported to have higher levels of RS and exhibit lower starch digestibility (Magallanes-Cruz et al., 2017) than the A- and C-type starch. That can explain Zhedagaozhi 1B with B-type starch had higher RS2 (Table 1). Short-range ordered structure of starch granule can be observed by FT-IR, the ratio of absorption at 1,045/1,022 cm^{-1} and 1,022/995 cm^{-1} can reflect the degree of order structure and the proportion of amorphous to ordered carbohydrate structure in starch granule external region, respectively (Sevenou et al., 2002). All four varieties showed similar FTIR profiles without new peaks appeared (Figure 2B). RS111 and Zhedagaozhi 1B showed lower 1,045/1,022 cm^{-1} ratio than their parents respectively (Table 1), indicating that there were less short-range order helical structures in the external region of two mutants. Meantime, RS111 and Zhedagaozhi 1B also showed different short-range ordered structure, RS111 had higher ratio of 1,045/1,022 cm^{-1} and lower ratio of 1,022/995 cm^{-1} than Zhedagaozhi 1B (Table 1).

AAC had been confirmed to be negatively correlated with the degree of particle surface order (Sun et al., 2022) and RC positively related to the degree of particle surface ordering (Zou et al., 2022). The lower 1,045/1,022 cm^{-1} in both mutants reflected as the lower crystallinity and consistent with their higher AAC (Table 1). However, RS111 and Zhedagaozhi 1B had similar AAC and

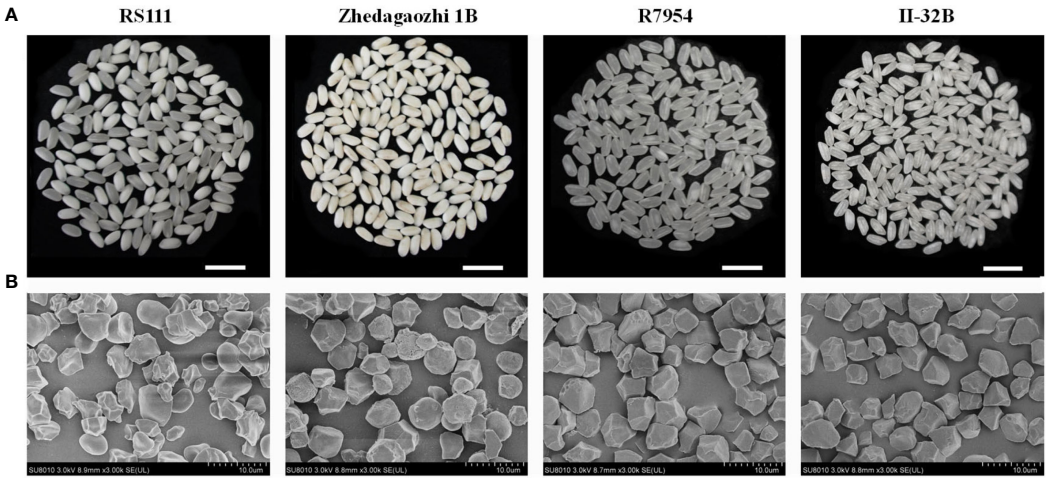


FIGURE 1
(A) The grain morphology of milled rice. **(B)** Scanning electron micrographs (SEM) of flour rice starch granules.

10.3389/fpls.2023.1267281

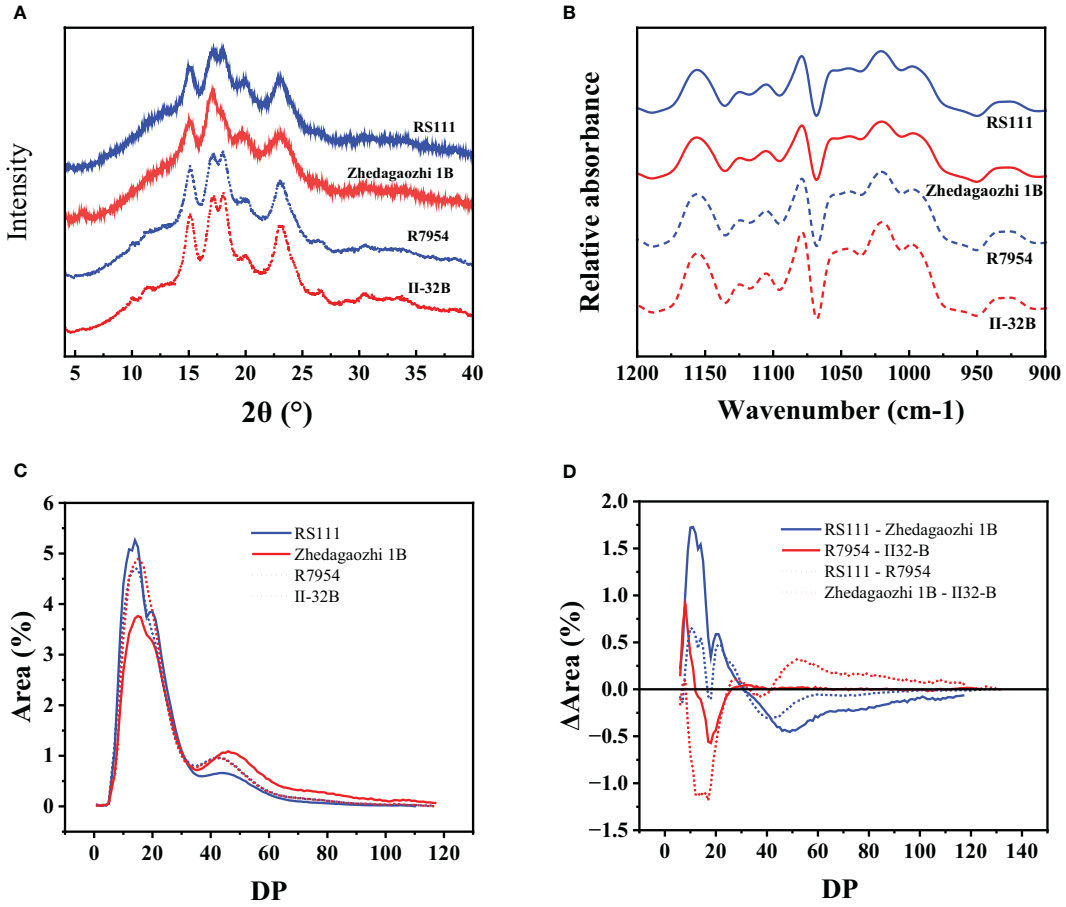


FIGURE 2
(A) XRD patterns of four rice starches. Data have been offset for clarify. **(B)** Fourier transform infrared (FTIR) spectroscopy spectra of four rice starches. **(C)** Chain length distribution patterns of amylopectin for four starches. **(D)** Comparison of the chain length distribution patterns of amylopectin between RS111 and Zhedagaozhi 1B.

similar RC; the different short-range ordered structure might be due to the different fine structure of amylopectin, as in addition to AAC, fine amylopectin structure also influenced the ordered structure greatly.

Chain-length distributions of amylopectin

Based on the DP, the side chains of amylopectin can be classified into fa (DP ≤ 12), fb1 (13 ≤ DP ≤ 24), fb2 (25 ≤ DP ≤ 36), and fb3 (DP ≥ 37), corresponding to fractions A, B1, B2, and B3 and longer chains, respectively (Hanashiro et al., 1996). R7954 had more proportions of fa and less fb1 than II-32B, but the proportions of fb2 and fb3 were similar. RS111 had significantly higher proportion of fa and fb1 but lower of fb3 than Zhedagaozhi 1B (Table 2), with significant increases in DP ≤ 30 (Figure 2D). Although, the differences in chain length distribution between R7954 and II-32B were obviously smaller than those between RS111 and Zhedagaozhi 1B (Figure 2D). When compared with their parents, RS111 had significant increases in fa, fb1 and decrease in fb3, while Zhedagaozhi 1B had significantly

TABLE 2 Branch chain-length distribution of rice starch amylopectins*.

	Peak DP	fa DP6-12(%)	fb1 DP13-24(%)	fb2 DP25-36(%)	fb3 DP ≥ 37(%)	fa/fb1	fa/fb2	fa/fb3
RS111	14	21.77 ± 0.18a	48.56 ± 0.26a	14.65 ± 0.34a	14.91 ± 0.74c	0.45	1.49	1.46
Zhedagaozhi 1B	15	13.22 ± 0.17d	39.20 ± 0.06c	14.25 ± 0.31a	33.19 ± 0.05a	0.34	0.93	0.40
R7954	14	18.89 ± 0.60b	44.68 ± 0.03b	14.31 ± 0.60a	22.01 ± 0.03b	0.42	1.32	0.86
II-32B	15	16.61 ± 0.49c	48.70 ± 0.59a	13.44 ± 0.51a	21.14 ± 0.61b	0.34	1.24	0.79

*The results were expressed as mean ± SD (n = 3). The same small letters indicated that there were no significant differences among the starches or flours of rice ($p < 0.05$).

decreased proportions of fa, fb1 and fb2, but increased fb3 (Table 2). Meantime, Chain-length distributions (CLDs) of amylopectin showed a peak DP (degree of polymerization) at 14 and 15 for RS111 and Zhedagaozhi 1B, respectively. Zhedagaozhi 1B showed significantly different CLDs from RS111, R7954, and II-32B.

The differential CLDs between RS111 and Zhedagaozhi 1B might explain the differential short ordered structure (Table 1), which might be responsible majorly by the amylopectin structure. The proportion of fa, fb1 showed negative impacts on RS2, while the proportion of fb3 showed positive correlation with RS2. Ramadoss et al. (2018) also found RS2 content showed significantly positive correlation with medium and long chains (DP > 12). Amylopectin with intermediate and longer can pack into intact or dense structures, which showed stronger resistance to enzymatic hydrolysis (Chi et al., 2021). However, the chain length distribution did not show direct correlation with RS3, Zhedagaozhi 1B with the lowest fa and fb1 still had higher RS3 content, inconsistent with our previous reports that RS3 showed significantly positive correlation with fa (Shu et al., 2007). That might, because RS3 was formed majorly after retrogradation, the ordered hierarchical starch structures were progressively disrupted during cooking. The reassembled structures involving amylose–amylose, amylose–lipids, and amylopectin–amylose during cooking, cooling might be more important for RS3 content (Chi et al., 2021), especially for Zhedagaozhi 1B.

Thermal characteristics of rice flour and starch

Thermal properties of native starches and flours were summarized in Table 3. The onset temperature (*To*), peak GT (*Tp*), conclusion temperature (*Tc*), and enthalpy of gelatinization (ΔH_{gel}) varied significantly among four materials. The *To*, *Tp*, *Tc*, and ΔH_{gel} of Zhedagaozhi 1B were higher than those of RS111 for both starch and flour, which was consistent with the difference between their parents. When compared with their parents, RS111 had lower *To*, *Tp*, *Tc*, and ΔH_{gel}, while Zhedagaozhi 1B had higher *To*, *Tp*, *Tc*, and ΔH_{gel}. Meantime, the value of *R*(*Tc*-*To*)of mutant Zhedagaozhi 1B was significantly higher than other varieties, and the differences of thermal characteristics between starch and flour were more significant for Zhedagaozhi 1B, which may greatly be influenced by the non-starch components such as lipids. GT has been found to be negatively correlated with DP6-11 and positively correlated with DP12-24 (Bao et al., 2009). RS111 with the highest proportion of fa had the lowest GT while Zhedagaozhi 1B had the lowest proportion of fa had the highest GT (Tables 2, 3). Furthermore, Zhedagaozhi 1B with the highest GT had the highest content of RS2 while RS111 with the lowest GT had the highest RS3, in parts line with previous study that DSC thermal parameters showed a negative correlation with RS in cooked rice and retrogradation rice but positive correlation with RS in raw milled rice (You et al., 2022).

Thermal parameters can reflect the double helix unfolding and crystallite fusion in the starch crystalline region during the gelatinization process of starch. *ΔH* primarily reflected the loss of double helical order (Li et al., 2017). The highest GT and *ΔH* of II32-B indicated the gelatinization of II32-B needed more energy, which might result from its lower proportion of amorphous to ordered carbohydrate structure in starch granule external region (Table 1). The lower thermal temperature and *ΔH* in RS111 may be majorly due to its higher proportion of fa (Table 2 and Figure 2D), lower RC, and higher proportion of amorphous to ordered carbohydrate structure (Table 1).

Gelatinization properties of rice starch

When gelatinized in 5 M urea, four varieties showed different degrees of gelatinization (Figure 3A). R7954 and II-32B swelled obviously in urea, whereas the starch of RS111 and Zhedagaozhi 1B were hardly gelatinized, especially Zhedagaozhi 1B, which was scarcely gelatinized. Meanwhile, the solubility of Zhedagaozhi 1B determined by the volume of sediment was significantly lower than that of RS111 (Table 1). These results were consistent with the chain length distribution of amylopectin that starch with more short chain was easier to be gelatinized (Li, 2022). The lowest swelling and gelatinizing of Zhedaogaozhi 1B might explain for its highest RS2 content as swelling power showed negative correlation with RS (Singh et al., 2010). However, Zhedaogaozhi 1B had less RS3

TABLE 3 Thermal properties of four rice *.

		To (°C)	*Tp* (°C)	*Tc* (°C)	△*Hgel* (J/g)	*R* (°C)
Flour	RS111	57.19 ± 0.18c	63.81 ± 0.07d	71.06 ± 0.15d	7.60 ± 0.31b	13.87 ± 0.33b
	Zhedagaozhi 1B	70.31 ± 0.85a	80.94 ± 0.11a	88.54 ± 0.30a	9.00 ± 0.44a	18.23 ± 1.08a
	R7954	59.27 ± 0.04b	65.71 ± 0.14c	73.58 ± 0.25c	9.52 ± 0.05a	14.31 ± 0.21b
	II-32B	70.05 ± 0.07a	75.04 ± 0.24b	79.71 ± 0.21b	9.55 ± 0.22a	9.65 ± 0.22c
Starch	RS111	58.55 ± 0.13d	63.89 ± 0.26d	69.68 ± 0.21d	10.43 ± 0.35c	11.13 ± 0.32b
	Zhedagaozhi 1B	63.33 ± 0.39b	77.64 ± 0.13a	89.51 ± 0.24a	16.33 ± 1.47a	26.19 ± 0.49a
	R7954	61.94 ± 0.07c	65.49 ± 0.13c	71.05 ± 0.31c	13.55 ± 0.21b	9.11 ± 0.26c
	II-32B	68.79 ± 0.78a	73.77 ± 0.25b	79.58 ± 0.35b	15.83 ± 0.56a	10.80 ± 0.54b

*The results were expressed as mean ± SD (n = 3). To, onset temperature; Tp, peak temperature; Tc, conclusion temperature; ΔH_{gel}, enthalpy of gelatinization; R, Tc − To.

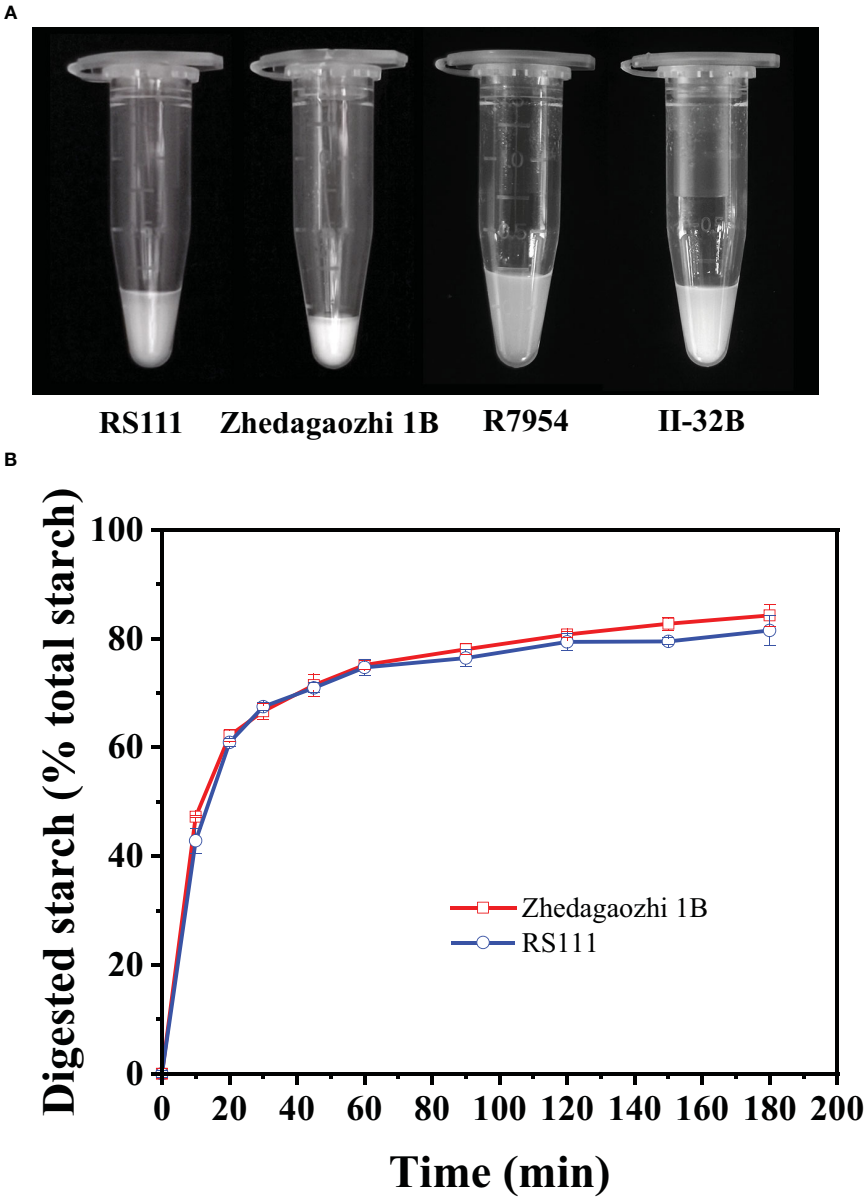


FIGURE 3
(A) Effects of 5M urea on the swelling of starch granules from the endosperm of two genotypes. **(B)** *In-vitro* starch hydrolysis curves of the two rice varieties cooked in different forms. Error bars represent the standard deviation of the mean of triplicate digestion.

content than RS111, although it is hard to be gelatinized in urea, that might be because of the gelatinization in urea cannot reflect completely the gelatinization in boiling water and the reassembled structures during cooking and cooling played more important role in RS3 formation (Chi et al., 2021).

Pasting properties of rice flours and starches

The RVA profiles showed significant differences among four rice varieties (Figures 4A, B). All RVA parameters, namely, PV, TV, breakdown viscosity, FV, and setback viscosity of Zhedagaozhi 1B were dramatically higher than that of RS111, both in starch and flour (Table 4). While the pasting temperature (PT) of RS1111 flour was higher than that of Zhedagaozhi 1B, and PT showed significantly positive correlation with RS3 (Table 3). Srichuwong and Jane (2007) found that starches consisting of amylopectin with more short A chains (DP 6–12) had a lower PT and PV and larger breakdown value since the short branch chains could not provide strong interaction to maintain the integrity of the swollen granules. The current results were partially consistent with this study, suggesting that there were other factors affected the pasting properties of rice starch between two mutants.

10.3389/fpls.2023.1267281

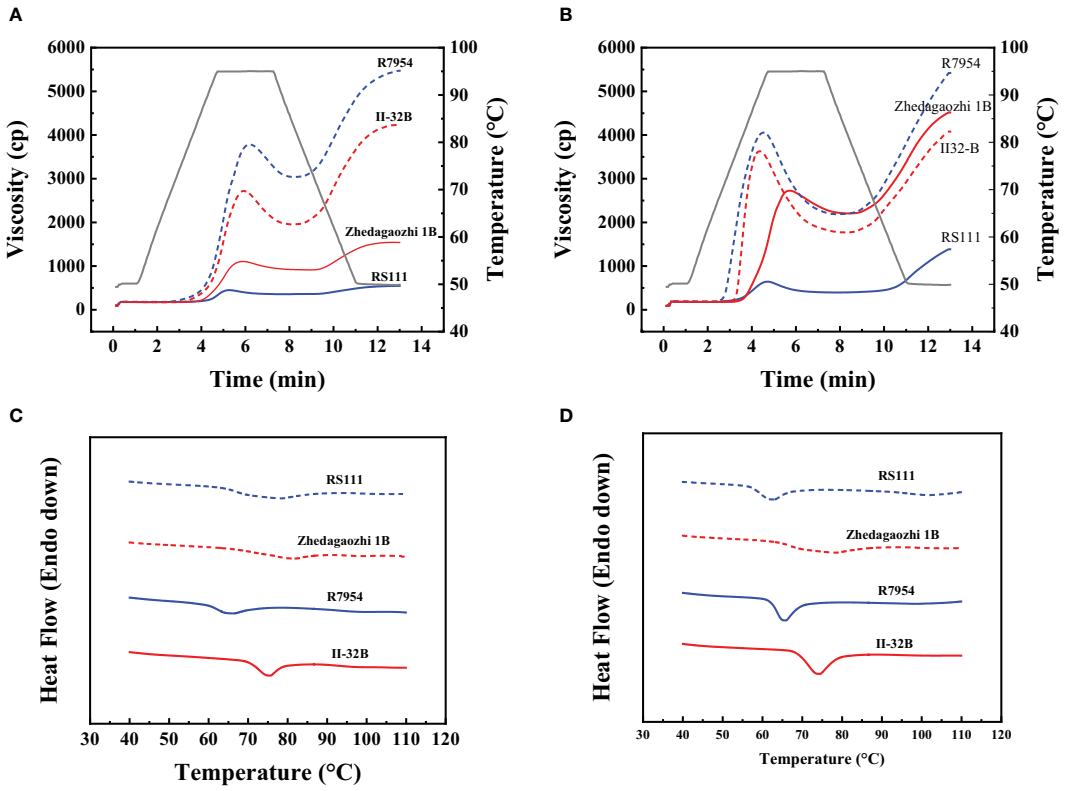


FIGURE 4
(A) Pasting properties of rice flours measured by Rapid Visco Analyzer (RVA). **(B)** Pasting properties of rice starches measured by RVA. **(C)** Thermal properties of the four rice fours measured by DSC. **(D)** Thermal properties of the four rice starches measured by DSC.

Digestion properties

Cooked flours were homogenized to mimic chewing and then digested using the standard INFOGEST2.0 method (Brodkorb et al., 2019). The digestion data of two mutants both showed good fit with a first-order equation ($R^2 > 0.86$). RS111 showed a lower digestion extent than that of Zhedagaozhi 1B (Figure 4B and Table 5). The digestion extent (C_∞) and the digestion rate (k) of RS111 was significantly lower than that of Zhedagaozhi 1B (Table 5). Considering the calculated values of C_∞ were affected by the digestion time, the experimentally determined values of C_{120} were also presented, which were similar as C_∞ (Table 5). Although Holm et al. (1988) found that the degree of starch gelatinization is an important determinant for the rate of starch hydrolysis *in vitro* (Holm et al., 1988) and Zhedagaozhi 1B was scarcely gelatinized, gelatinized and retrograded starch can benefit the formation of RS, which has been reported to be decisive for the glycemic index of rice (Kumar et al., 2018). The RS3 content should be the major

TABLE 4 Pasting viscosity of four rice *.

		PV (cP)	TV (cP)	BV (cP)	FV (cP)	SV (cP)	PT (°C)
Flour	RS111	452.7 ± 5.5d	361.7 ± 5.7d	91.0 ± 2.0b	547.7 ± 4.0d	186.0 ± 8.7d	92.3 ± 0.8a
	Zhedagaozhi 1B	1115.5 ± 14.8c	923.5 ± 16.3c	192.0 ± 1.4b	1540.0 ± 2.8c	616.5 ± 19.1c	87.1 ± 0.6b
	R7954	3811.5 ± 38.9a	3033.0 ± 11.3a	778.5 ± 50.2a	5539.5 ± 96.9a	2506.5 ± 108.2a	81.1 ± 0.0c
	II-32B	2749. 0 ± 89.1b	2009.0 ± 59.3b	740.0 ± 52.3a	4221.3 ± 36.7b	2212.3 ± 56.9b	79.2 ± 1.2c
Starch	RS111	623.7 ± 20.0d	388.3 ± 7.6c	235.3 ± 12.7c	1424.7 ± 39.0d	1036.3 ± 42.0c	80.0 ± 1.0a
	Zhedagaozhi 1B	2746.7 ± 25.4c	2230.7 ± 21.4a	516.0 ± 17.3b	4573.3 ± 49.8b	2355.3 ± 45.0b	80.0 ± 0.5a
	R7954	4009.0 ± 66.5a	2226.0 ± 53.7a	1783.0 ± 120.2a	5362.0 ± 80.6a	3136.0 ± 134.4a	69.6 ± 0.5c
	II-32B	3615.0 ± 17.0b	1823.5 ± 74.2b	1791.5 ± 91.2a	4210.5 ± 187.4c	2387.0 ± 113.1b	76.4 ± 1.1b

*The results were expressed as mean ± SD (n = 3). The same small letters indicated that there were no significant differences among the starches or flours of rice ($p < 0.05$). PV, peak viscosity; TV, trough viscosity; BV, breakdown viscosity; FV, final viscosity; SV, setback viscosity; PT, pasting temperature.

TABLE 5 Estimated kinetic parameters for starch digestion of the two rice varieties*.

	C_∞ (%)	C_{120} (%)	k (min^{-1})	IRR_{10} (% min^{-1})
RS111	77.8 ± 0.2**	77.7 ± 0.2**	0.074 ± 0.003	4.07 ± 0.11
Zhedagaozhi 1B	79.1 ± 0.1**	79.1 ± 0.1**	0.077 ± 0.002	4.24 ± 0.06

*The results were expressed as mean ± SD (n = 3). The "**" indicates significant correlations at the 0.01 levels (2-tailed).

contributor for the digestion rate in cooked rice; thus, RS111 with higher RS3 content had slower digestion rate (Tables 1, 5).

Conclusions

RS111 and Zhedagaozhi 1B had similar AAC with around 34%, significantly higher than their wild-type parent and common rice. Although two wild types had similar properties, RS111 and Zhedagaozhi 1B conferred different physiochemical properties, such as viscosity, gelatinization parameters, swelling power, and digestibility. RS111 had higher RS3 content but lower RS2 content than Zhedagaozhi 1B; correspondingly, cooked RS111 showed slower digestibility. Differences in physiochemical properties might be due to the different starch structure between RS111 and Zhedagaozhi 1B. The B-type starch granule and high proportion of fb3 in Zhedagaozhi 1B may be responsible for its high-RS3 content, while the smaller irregular and oval starch granules, high proportion of fa in RS111 might be the major contributor to its rich RS3. As RS111 and Zhedagaozhi 1B are derived from restorer and maintainer, respectively; they can be used directly as new rice varieties, or as parents for high-RS hybrid rice breeding through improving other restorers and maintainers. Fine structure of amylopectin seemed very important for determining the RS type and content when amylose content is similar. Modifications on amylopectin structure showed great potential in breeding rice with different RS2 and RS3 content, which could meet the increasingly needs for various rice germplasms.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ML: Methodology, Data curation, Investigation, Validation, Writing – original draft. WG: Data curation, Formal Analysis, Methodology, Writing – original draft, Visualization. SZ: Investigation, Data curation, Methodology, Writing – original draft. LX: Data curation, Investigation, Writing – original draft. YS: Data curation, Investigation, Writing – original draft. DW: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing. XS: Formal Analysis, Methodology, Visualization, Writing – review & editing, Funding acquisition.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article. The authors greatly acknowledge the financial support from the National Key Research and Development Program of China (2021YFF1000202), the Functional Rice Breeding, Germplasm Enhancement, R&D and Basic Public Welfare Research Program Project (2022C02011, 2021C02063, 2021-2-032), and the National Natural Science Foundation of China (32072038).

Conflict of interest

Authors ML, LX, and YS are employed by the company China National Seed Group Co., Ltd., Wuhan.

The remaining 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

References

Abe, N., Asai, H., Yago, H., Oitome, N. F., Itoh, R., Crofts, N., et al. (2014). Relationships between starch synthase I and branching enzyme isozymes determined using double mutant rice lines. *BMC Plant Biol.* 14, 80. doi: 10.1186/1471-2229-14-80

Bao, J., Xiao, P., Hiratsuka, M., Sun, M., and Umenoto, T. (2009). Granule-bound SSIIa protein content and its relationship with amylopectin structure and gelatinization temperature of rice starch. *Starch/Stärke* 61, 431–437. doi: 10.1002/star.200800115

Baysal, C., He, W., Drapal, M., Villorbina, G., Medina, V., Capell, T., et al. (2020). Inactivation of rice starch branching enzyme IIb triggers broad and unexpected changes in metabolism by transcriptional reprogramming. *Proc. Natl. Acad. Sci. U. S. A.* 117, 26503–26512. doi: 10.1073/pnas.2014860117

Birt, D. F., Boylston, T., Hendrich, S., Jane, J. L., Hollis, J., Li, L., et al. (2013). Resistant starch: Promise for improving human health. *Adv. Nutr.* 4, 587–601. doi: 10.3945/an.113.004325

Brodkorb, A., Egger, L., Alminger, M., Alvito, P., Assunção, R., Ballance, S., et al. (2019). INFOGEST static in *vitro* simulation of gastrointestinal food digestion. *Nat. Protoc.* 14, 991–1014. doi: 10.1038/s41596-018-0119-1

Buléon, A., Colonna, P., Planchot, V., and Ball, S. (1998). Starch granules: Structure and biosynthesis. *Int. J. Biol. Macromolecules* 23, 85–112. doi: 10.1016/S0141-8130(98)00040-3

Butardo, V. M., Fitzgerald, M. A., Bird, A. R., Gidley, M. J., Flanagan, B. M., Larroque, O., et al. (2011). Impact of down-regulation of starch branching enzyme IIb in rice by artificial microRNA-and hairpin RNA-mediated RNA silencing. *J. Exp. Bot.* 62, 4927–4941. doi: 10.1093/jxb/err188

Chi, C., Li, X., Huang, S., Chen, L., Zhang, Y., Li, L., et al. (2021). Basic principles in starch multi-scale structuration to mitigate digestibility: A review. *Trends Food Sci. Technol.* 109, 154–168. doi: 10.1016/j.tifs.2021.01.024

Chung, H. J., Liu, Q., Lee, L., and Wei, D. (2011). Relationship between the structure, physicochemical properties and in *vitro* digestibility of rice starches with different amylose contents. *Food Hydrocolloids* 25, 968–975. doi: 10.1016/j.foodhyd.2010.09.011

Cui, R., and Oates, C. G. (1999). The effect of amylose-lipid complex formation on enzyme susceptibility of sago starch. *Food Chem.* 65, 417–425. doi: 10.1016/S0308-8146(97)00174-X

Ding, Y., Huang, J., Zhang, N., Rasmussen, S. K., Wu, D., and Shu, X. (2019). Physiochemical properties of rice with contrasting resistant starch content. *J. Cereal Sci.* 89, 102815. doi: 10.1016/j.jcs.2019.102815

Englyst, H. N., and Cummings, J. H. (1987). Digestion of polysaccharides of potato in the small intestine of man. *Am. J. Clin. Nutr.* 45, 423–431. doi: 10.1093/ajcn/45.2.423

Englyst, H. N., Kingman, S. M., and Cummings, J. H. (1992). Classification and measurement of nutritionally important starch fractions. *Eur. J. Clin. Nutr.* 46Suppl 2, S33–S50.

Faisant, N., Buléon, A., Colonna, P., Molis, C., Lartigue, S., Galmiche, J. P., et al. (1995). Digestion of raw banana starch in the small intestine of healthy humans: structural features of resistant starch. *Br. J. Nutr.* 73, 111–123. doi: 10.1079/BJN19950013

Frei, M., Siddhuraju, P., and Becker, K. (2003). Studies on the in *vitro* starch digestibility and the glycemic index of six different indigenous rice cultivars from the Philippines. *Food Chem.* 83, 395–402. doi: 10.1016/S0308-8146(03)00101-8

Fujita, N., Yoshida, M., Kondo, T., Saito, K., Utsumi, Y., Tokunaga, T., et al. (2007). Characterization of SSIIIa-deficient mutants of rice: The function of SSIIIa and pleiotropic effects by SSIIIa deficiency in the rice endosperm. *Plant Physiol.* 144, 2009–2023. doi: 10.1104/pp.107.102533

Gong, W., Liu, T., Zhou, Z., Wu, D., Shu, X., and Xiong, H. (2021). Physicochemical characterizations of starches isolated from Tetrastigma hemsleyanum Diels et Gilg. *Int. J. Biol. Macromolecules* 183, 1540–1547. doi: 10.1016/j.ijbiomac.2021.05.117

Goñi, I., Garcia-Alonso, A., and Saura-Calixto, F. (1997). A starch hydrolysis procedure to estimate glycemic index. *Nutr. Res.* 17, 427–437. doi: 10.1016/S0271-5317(97)00010-9

Hanashiro, I., Abe, J. I., and Hizukuri, S. (1996). A periodic distribution of the chain length of amylopectin as revealed by high-performance anion-exchange chromatography. *Carbohydr. Res.* 283, 151–159. doi: 10.1016/0008-6215(95)00408-4

Hayakawa, K., Tanaka, K., Nakamura, T., Endo, S., and Hoshino, T. (1997). Quality characteristics of waxy hexaploid wheat (Triticum aestivum L.): Properties of starch gelatinization and retrogradation. *Cereal Chem.* 74, 576–580. doi: 10.1094/CCHEM.1997.74.5.576

Holm, J., Lundquist, I., Bjork, I., Eliasson, A. C., and Asp, N. G. (1988). Degree of starch gelatinization, digestion rate of starch in *vitro*, and metabolic response in rats. *Am. J. Clin. Nutr.* 47, 1010–1016. doi: 10.1093/ajcn/47.6.1010

Jeon, J. S., Ryoo, N., Hahn, T. R., Walia, H., and Nakamura, Y. (2010). Starch biosynthesis in cereal endosperm. *Plant Physiol. Biochem.* 48, 383–392. doi: 10.1016/j.plaphy.2010.03.006

Kan, L., Oliviero, T., Verkerk, R., Fogliano, V., and Capuano, E. (2020). Interaction of bread and berry polyphenols affects starch digestibility and polyphenols bio-accessibility. *J. Funct. Foods* 68, 103924. doi: 10.1016/j.jff.2020.103924

Kang, H. J., Hwang, I. K., Kim, K. S., and Choi, H. C. (2003). Comparative structure and physicochemical properties of Ilpumbyeo, a high-quality japonica rice, and its mutant, Suweon 464. *J. Agric. Food Chemisty* 51, 6598–6603. doi: 10.1021/jf0344946

Kong, X., Bertoft, E., Bao, J., and Corke, H. (2008). Molecular structure of amylopectin from amaranth starch and its effect on physicochemical properties. *Int. J. Biol. Macromolecules* 43, 377–382. doi: 10.1016/j.ijbiomac.2008.07.018

Kong, X., Zhu, P., Sui, Z., and Bao, J. (2015). Physicochemical properties of starches from diverse rice cultivars varying in apparent amylose content and gelatinisation temperature combinations. *Food Chem.* 172, 433–440. doi: 10.1016/j.foodchem.2014.09.085

Sajilata, M. G., Singhal, R. S., and Kulkarni, P. R. (2006). Resistant starch - A review. *Compr. Rev. Food Sci. Food Saf.* 5, 1–17. doi: 10.1111/j.1541-4337.2006.tb00076.x

Kumar, A., Sahoo, U., Baisakha, B., Okpani, O. A., Ngangkham, U., Parameswaran, C., et al. (2018). Resistant starch could be decisive in determining the glycemic index of rice cultivars. *J. Cereal Sci.* 79, 348–353. doi: 10.1016/j.jcs.2017.11.013

Li, C. (2022). Recent progress in understanding starch gelatinization - An important property determining food quality. *Carbohydr. Polymers* 293, 119735. doi: 10.1016/j.carbpol.2022.119735

Li, P., He, X., Dhital, S., Zhang, B., and Huang, Q. (2017). Structural and physicochemical properties of granular starches after treatment with debranching enzyme. *Carbohydr. Polymers* 169, 351–356. doi: 10.1016/j.carbpol.2017.04.036

Magallanes-Cruz, P. A., Flores-Silva, P. C., and Bello-Perez, L. A. (2017). Starch structure influences its digestibility: A review. *J. Food Sci.* 82, 2016–2023. doi: 10.1111/1750-3841.13809

Man, J.-M., Cai, C.-H., Yan, Q.-X., Hu, M.-Z., Liu, Q.-Q., and Wei, C.-X. (2012). Applications of infrared spectroscopy in the analysis of ordered structure of starch grain. *Acta Agronomica Sin.* 38, 505–513. doi: 10.3724/sp.j.1006.2012.00505

Miura, S., Koyama, N., Crofts, N., Hosaka, Y., Abe, M., and Fujita, N. (2021). Generation and *starch characterization of non-trans*genic BEI and BEIIb Double Mutant rice (*Oryza sativa*) with ultra-high level of resistant starch. *Rice* 14, 3. doi: 10.1186/s12284-020-00441-0

Nakata, M., Miyashita, T., Kimura, R., Nakata, Y., Takagi, H., Kuroda, M., et al. (2018). MutMapPlus identified novel mutant alleles of a rice starch branching enzyme IIb gene for fine-tuning of cooked rice texture. *Plant Biotechnol. J.* 16, 111–123. doi: 10.1111/pbi.12753

Nishi, A., Nakamura, Y., Tanaka, N., and Satoh, H. (2001). Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. *Plant Physiol.* 127, 459–472. doi: 10.1104/pp.010127

Noah, L., Guillon, F., Bouchet, B., Buléon, A., Molis, C., Gratas, M., et al. (1998). Digestion of Carbohydrate from White beans (*Phaseolus vulgaris* L.) in healthy humans. *J. Nutr.* 128, 977–985. doi: 10.1093/jn/128.6.977

Perera, A., Meda, V., and Tyler, R. T. (2010). Resistant starch: A review of analytical protocols for determining resistant starch and of factors affecting the resistant starch content of foods. *Food Res. Int.* 43, 1959–1974. doi: 10.1016/j.foodres.2010.06.003

Sevenou, O., Hill, S. E., Farhat, I. A., and Mitchell, J. R. (2002). Organisation of the external region of the starch granule as determined by infrared spectroscopy. *Int. J. Biol. Macromolecules* 31, 79–85. doi: 10.1016/S0141-8130(02)00067-3

Shen, L., Li, J., and Li, Y. (2022a). Resistant starch formation in rice: Genetic regulation and beyond. *Plant Commun.* 3, 100329. doi: 10.1016/j.xplc.2022.100329

Shen, Y., Gong, W., Li, Y., Deng, J., Shu, X., Wu, D., et al. (2022b). The physiochemical and nutritional properties of high endosperm lipids rice mutants under artificially accelerated ageing. *LWT* 154, 112730. doi: 10.1016/j.lwt.2021.112730

Shu, X., Jia, L., Gao, J., Song, Y., Zhao, H., Nakamura, Y., et al. (2007). The influences of chain length of amylopectin on resistant starch in rice (Oryza sativa L.). *Starch/Staerke* 59, 504–509. doi: 10.1002/star.200700640

Shu, X., Jia, L., Ye, H., Li, C., and Wu, D. (2009). Slow digestion properties of rice different in resistant starch. *J. Agric. Food Chem.* 57, 7552–7559. doi: 10.1021/jf900988h

Singh, J., Dartois, A., and Kaur, L. (2010). Starch digestibility in food matrix: a review. *Trends Food Sci. Technol.* 21, 168–180. doi: 10.1016/j.tifs.2009.12.001

Srichuwong, S., and Jane, J.-L. (2007). Physicochemical properties of starch affected by molecular composition and structures: A review. *Food Sci. Biotechnol.* 16 663–674.

Sun, H., Fan, J., Tian, Z., Ma, L., Meng, Y., Yang, Z., et al. (2022). Effects of treatment methods on the formation of resistant starch in purple sweet potato. *Food Chem.* 367, 130580. doi: 10.1016/j.foodchem.2021.130580

Tao, K., Yu, W., Prakash, S., and Gilbert, R. G. (2019). High-amylose rice: Starch molecular structural features controlling cooked rice texture and preference. *Carbohydr. Polymers* 219, 251–260. doi: 10.1016/j.carbpol.2019.05.031

Wei, C., Qin, F., Zhou, W., Yu, H., Xu, B., Chen, C., et al. (2010). Granule structure and distribution of allomorphs in C-type high-amylose rice starch granule modified by antisense RNA inhibition of starch branching enzyme. *J. Agric. Food Chem.* 58, 11946–11954. doi: 10.1021/jf103412d

Yang, C. Z., Shu, X. L., Zhang, L. L., Wang, X. Y., Zhao, H. J., Ma, C. X., et al. (2006). Starch properties of mutant rice high in resistant starch. *J. Agric. Food Chem.* 54, 523–528. doi: 10.1021/jf0524123

Yang, R., Sun, C., Bai, J., Lou, Z., Shi, B., Zhang, J., et al. (2012). A putative gene sbe3-rs for resistant starch mutated from SBE3 for starch branching enzyme in rice (Oryza sativa L.). *PloS One* 7, e43026. doi: 10.1371/journal.pone.0043026

You, H., Liang, C., Zhang, O., Xu, H., Xu, L., Chen, Y., et al. (2022). Variation of resistant starch content in different processing types and their starch granules properties in rice. *Carbohydr. Polymers* 276, 118742. doi: 10.1016/j.carbpol.2021.118742

Zhang, G. Y., Cheng, Z. J., Zhang, X., Guo, X. P., Su, N., Jiang, L., et al. (2011). Double repression of soluble starch synthase genes SSIIa and SSIIIa in rice (Oryza sativa L.) uncovers interactive effects on the physicochemical properties of starch. *Genome* 54, 448–459. doi: 10.1139/g11-010

Zhou, Y., Cheng, Z., Jiang, S., Cen, J., Wu, D., and Shu, X. (2022). High temperature boosts resistant starch content by altering starch structure and lipid content in rice ssIIIa mutants. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1059749

Zhou, Y., Cheng, Z., JIang, S., Cen, J., Yuan, S., Yu, C., et al. (2023). Inactivation of SSIIIa enhances the RS content through altering starch structure and accumulating C18:2 in japonica rice. *Carbohydr. Polymers* 318, 121141. doi: 10.1016/j.carbpol.2023.121141

Zhu, L. J., Liu, Q. Q., Wilson, J. D., Gu, M. H., and Shi, Y. C. (2011). Digestibility and physicochemical properties of rice (Oryza sativa L.) flours and starches differing in amylose content. *Carbohydr. Polymers* 86, 1751–1759. doi: 10.1016/j.carbpol.2011.07.017

Zou, J., Li, Y., Su, X., Wang, F., Li, Q., and Xia, H. (2022). Structure and processing properties of nine yam (Dioscorea opposita thunb) starches from South China: A comparison study. *Molecules* 27, 2254. doi: 10.3390/molecules27072254

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 January 2024
DOI 10.3389/fpls.2023.1320638

OPEN ACCESS

EDITED BY
Reyazul Rouf Mir,
Sher-e-Kashmir University of Agricultural Sciences and Technology, India

REVIEWED BY
Upendra Kumar,
M. J. P. Rohilkhand University, India
Shoupu He,
Institute of Cotton Research (CAAS), China
Qian-Hao Zhu,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

*CORRESPONDENCE
Christopher A. Saski
saski@clemson.edu

†These authors contributed equally to this work

RECEIVED 12 October 2023
ACCEPTED 20 December 2023
PUBLISHED 31 January 2024

CITATION
Kangben F, Kumar S, Li Z, Sreedasyam A, Dardick C, Jones D and Saski CA (2024) Phylogenetic and functional analysis of tiller angle control homeologs in allotetraploid cotton.
Front. Plant Sci. 14:1320638.
doi: 10.3389/fpls.2023.1320638

Phylogenetic and functional analysis of tiller angle control homeologs in allotetraploid cotton

Foster Kangben[1†], Sonika Kumar[1†], Zhigang Li[1], Avinash Sreedasyam[2], Chris Dardick[3], Don Jones[4] and Christopher A. Saski[1*]

[1]Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, United States, [2]Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, United States, [3]United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Appalachian Fruit Research Station, Kearneysville, WV, United States, [4]Department of Agricultural Research, Cotton Incorporated, Cary, NC, United States

Introduction: Plants can adapt their growth to optimize light capture in competitive environments, with branch angle being a crucial factor influencing plant phenotype and physiology. Decreased branch angles in cereal crops have been shown to enhance productivity in high-density plantings. The Tiller Angle Control (TAC1) gene, known for regulating tiller inclination in rice and corn, has been found to control branch angle in eudicots. Manipulating TAC1 in field crops like cotton offers the potential for improving crop productivity.

Methods: Using a homolog-based methodology, we examined the distribution of TAC1-related genes in cotton compared to other angiosperms. Furthermore, tissue-specific qPCR analysis unveiled distinct expression patterns of TAC1 genes in various cotton tissues. To silence highly expressed specific TAC1 homeologs in the stem, we applied CRISPR-Cas9 gene editing and Agrobacterium-mediated transformation, followed by genotyping and subsequent phenotypic validation of the mutants.

Results: Gene duplication events of TAC1 specific to the Gossypium lineage were identified, with 3 copies in diploid progenitors and 6 copies in allotetraploid cottons. Sequence analysis of the TAC1 homeologs in Gossypium hirsutum revealed divergence from other angiosperms with 1-2 copies, suggesting possible neo- or sub-functionalization for the duplicated copies. These TAC1 homeologs exhibited distinct gene expression patterns in various tissues over developmental time, with elevated expression of A11G109300 and D11G112200, specifically in flowers and stems, respectively. CRISPR-mediated loss of these TAC1 homeologous genes resulted in a reduction in branch angle and altered petiole angles, and a 5 to 10-fold reduction in TAC1 expression in the mutants,

confirming their role in controlling branch and petiole angles. This research provides a promising strategy for genetically engineering branch and petiole angles in commercial cotton varieties, potentially leading to increased productivity.

KEYWORDS

branch angle, CRISPR, plant architecture, petiole angle, tac1, upland cotton

Highlight

The Tiller Angle Control (*TAC1*) gene is duplicated in the *Gossipum* lineage and regulates branch angle with significant A-subgenome expression bias. Manipulating this gene in Upland cotton can potentially improve productivity through high-density planting.

Introduction

Improving crop performance involves various factors such as optimized light capture, carbon assimilation, and photosynthetic efficiency to achieve higher yield (Kant et al., 2012; Murchie and Burgess, 2022). The precise manipulation of plant architecture, particularly branching, can significantly enhance crop management and productivity. By controlling branching and limiting plant-to-plant interference, farmers can increase planting density, photosynthesis and carbon assimilation, boost yield, and improve mechanization efficiency (McGarry et al., 2016; Fladung, 2021).

Plant architecture is defined by the spatial configuration and morphological traits of its aerial and rooting structures in three dimensions. This fundamental feature has undergone modification throughout crop domestication and is crucial to the plant's adaptability and productivity (Yang and Hwa, 2008; Cai et al., 2016). Plant architecture is the outcome of an intricate interplay between diverse regulatory mechanisms encompassing genetic programming, various hormone signaling pathways, and response to environmental cues such as light, gravity, and mechanical forces. A plant's form, stem and leaf arrangement, and overall growth patterns are all influenced by these factors (Reinhardt and Kuhlemeier, 2002; Li et al., 2022; Sun et al., 2022). In angiosperms, variations in branch angles and other structures, such as leaves and branches, are viewed as adaptive strategies in modulating light interception efficiency. The ability to intercept light under varying conditions can also affect plant growth patterns and structures (Duursma et al., 2012). A key aspect of the success of the Green Revolution was the genetic modification of cereal crops, such as wheat, to improve their productivity and growth potential through breeding and selection for short, robust stems. This led to the development of wheat varieties that could withstand damage from wind and rain (Peng et al., 1999; Reinhardt and Kuhlemeier, 2002; Song and Zhang, 2009). Likewise, the architecture and yield of corn has been substantially improved through selection for reduced tillering, upright leaf angles, and increased apical dominance compared to its ancestor, teosinte (Whipple et al., 2011; Schmidt et al., 2016). Modern breeding has focused on the development of elite cultivars with erect, compact stem architecture, upright leaf angles, reduced branching, high harvest indices, and increased seed yield (Li et al., 2013). It is now commonly recognized that branches oriented vertically near the meristem and more horizontally at the lower canopy, are considered ideal for optimal light interception and correlated with higher yields (Kaggwa-Asiimwe et al., 2013; Sun et al., 2022). Using modern tools, plant architecture, including branch and leaf orientation, can now be manipulated with genome editing tools such as CRISPR to optimize light exposure and water uptake.

The Tiller Angle Control 1 (*TAC1*) gene has been identified as a crucial regulator of upright lateral organ orientation in rice, corn, and rapeseed (Yu et al., 2007; Ku et al., 2011; Li et al., 2017). *TAC1* belongs to a small gene family, known as the IGT family, which controls the orientation of organ growth by inhibiting response to gravity in various plants, including grasses, dicots, and trees. Along with *TAC1*, the IGT family includes LAZY and DRO subfamilies that promote upward orientation of branches and downward growth of lateral roots by facilitating auxin redistribution (Duvick, 2005; Yu et al., 2007; Ku et al., 2011; Li et al., 2017). Although LAZY and DRO mechanisms have been extensively studied, little is known about *TAC1*, which has been suggested as a negative regulator of LAZY (Hollender et al., 2020). Model species such as Arabidopsis and rice have a single copy of the *TAC1* gene (Guseman et al., 2017; Hollender et al., 2020). The loss of *TAC1* function leads to a broom or pillar-like plant architecture, characterized by vertically oriented branches, leaves, inflorescence, flower buds, and tillers (Dardick et al., 2013). Under optimal light

Abbreviations: USDA-ARS, United States Department of Agriculture-Agricultural Research Services; BLAST, Basic Local Alignment Search Tool; CAS9, CRISPR-associated protein 9; CDS, Coding Sequence; COP1, Constitutive Photomorphogenesis; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; CTAB, Cetyltrimethyl Ammonium Bromide; *TAC1*, Tiller Angle Control 1.

conditions the expression of *TAC1* is up-regulated, potentially via the constitutive photomorphogenesis (*COP*1) gene, resulting in wider branch angles that enhance photoreception efficiency (Waite and Dardick, 2018).

Plants exhibiting loss of *TAC1* phenotypes occupy less space which allows for a higher planting density corresponding to an increased yield. Loss of function of *TAC1* results in a more inclined orientation of branch growth when plants are grown in prolonged darkness (Waite and Dardick, 2018; Hollender et al., 2020). In *Arabidopsis* and peach *TAC1* is primarily expressed in the apical shoots and upper sections of the stem, as well as the upper laterals. *TAC1* expression in rice is temporal peaking at 60 days after sowing and declining to a minimum at 100 days after sowing, especially during the heading stage. This leads to a decrease in leaf shading and increases photosynthetic efficiency (Yu et al., 2007; Dardick et al., 2013).

Cotton is a widely grown crop, covering 5% of the world's farmland and valued for its fiber, seeds, and oil. It's vital to the U.S. economy, which is the top cotton exporter and ranks third in production after China and India (Wang et al., 2012; Yuan et al., 2015; Shahbandeh, 2021). One approach to maximizing cotton yield is through the augmentation of planting densities (Kaggwa-Asiimwe et al., 2013; Khan et al., 2020). Current recommendations based on elite cotton varieties range from 20,000 to 45,000 cotton plants per acre (ppa). Plant populations exceeding 60,000 or falling below 20,000 ppa can result in management challenges and significant reductions in yield potential (Adams et al., 2019). Conversely, leading cotton seed suppliers often advocate for a plant density of 50,000 plants ha-1 to achieve optimal yields per hectare, resulting in a range of 700 to 1000 kg/ha. Higher plant density per unit area confers numerous advantages, including increased yield due to a greater number of plants within a given space, efficient utilization of resources such as water, nutrients, and sunlight, decreased weed growth and soil erosion, improved pest management, and enhanced microclimate conditions (Kaggwa-Asiimwe et al., 2013; Adams et al., 2019; Basu and Parida, 2021). However, under poor soil and seed conditions, higher planting population densities of up to 129,000 plants ha-1 have been reported to achieve similar yields (Norton et al., 1995; Fok, 1998). The expected cotton yield is contingent upon various factors such as the specific cotton variety cultivated, land availability, soil type, climatic conditions, planting density, available resources, nutrition, and management practices.

One possible explanation for the limited increase in yield observed in high-density cotton plantings could be linked to the prevalent spreading or horizontally oriented branching patterns seen in most cotton cultivars. In fact, small acreage farmers opt for crop varieties with narrower branch angles that can support higher plant density per area while also facilitating mechanical field management. Cotton plants exhibit two types of branches: monopodial and sympodial. Monopodial branches, also known as vegetative branches, originate from the main stem and continue to grow indefinitely throughout the plant's development. Monopodial branches primarily grow in a nearly upright position. In contrast, sympodial branches, or fruiting branches, are determinate and bear the inflorescence, which eventually develops into the cotton bolls. Sympodial branches typically grow almost laterally from the main stem and are largely responsible for bearing the reproductive buds (squares). Fruiting branch angles ranging from 36.86° to 64.56° have been reported (Shao et al., 2022). With multiple branch types, cotton plants can allocate resources efficiently to both vegetative growth and reproduction (Gore, 1935; McGarry et al., 2016). Altered branch orientation for both branches could potentially impact cotton plant architecture. Given the desire for cotton varieties with narrow branch angles, it is critical that we improve our understanding of how monopodial and sympodial branch angles are regulated and identify strategies to develop improved germplasm with branch angles better suited for very high planting densities.

This study aimed to analyze the gene content and variation of *TAC1* in the genomes of diploid and allotetraploid cottons as well as other Angiosperms. Our goals were to: (I) identify *TAC1* orthologs in cotton species and other plant species; (II) determine homeologous gene expression profiles of *TAC1* in various tissues (stem, leaf, flower, fiber, meristem and root) in the allotetraploid cotton genotype Coker312; (III) functionally profile via CRISPR/CAS9 knockout the homeologous *TAC1* copies that are the most expressed in stem tissue; and (IV) develop gene editing strategies to modify cotton branch angles.

Materials and methods

Sequence alignment and phylogenetic analysis

To uncover potential *TAC1*-related genes we conducted a blast search and motif analysis of the Ppe*TAC1* coding sequence from *Prunus persica* in the full genome sequences of *Gossypium* species (*G. hirsutum*, *G.raimondii*, *G. arboreum*, *G. barbadense*, *G. darwinii*, *G.tomentosum*) and other selected plants. For *G. hirsutum* we used the widely used and well-annotated Coker312 genotype as it is known for its regenerative capacity (Kumar et al., 2021). Coding and protein sequences were obtained from Phytozome, and the BLOSUM62 matrix was used in the blast search with an expected value of 1e-5. Amino acid sequences were concatenated for species with multiple copies using the concatenate feature in Geneious Prime software version 2023.2.1. Multiple alignments were performed with MUSCLE v5.1 (Edgar, 2004), and phylogenetic trees were constructed using the RAxML GAMMA GTR and the Neighbor-Joining method with 5,000 bootstrap replicates (Stamatakis, 2014; Kozlov et al., 2019). Genomic data for the analysis of most of the crops was retrieved from Phytozome Genome Database (Goodstein et al., 2012)

Total RNA isolation

Samples of tissue were collected 14 days after anthesis (dpa) and immediately frozen. The tissue was then ground into a fine powder using a pestle and mortar in liquid nitrogen. Total RNA was extracted from 100mg of the tissue from leaf (basal, midsection

and apical part), stem (between the 5th and 15th node and sliced into smaller pieces), meristem (apical part at 30cm downward), root (bulk roots and root tips and homogenized), flower (day before bloom), and fiber (14dpa) of the Coker 312 cotton genotype using a modified CTAB protocol method (Kumar et al., 2021). The purity and concentration of the extracted RNA were measured using a Nanodrop 8000 UV-Vis Spectrophotometer (Thermo Scientific).

Quantitative real-time PCR

Six differentially expressed *TAC1* genes (GhCoker.A08G143500, GhCoker.D08G158700, GhCoker.A11G109300, GhCoker.D11G112200, GhCoker.A12G131200 and GhCoker.D12G137700) and three biological replicates each of leaf, stem, meristem, root, flower and fiber of Coker312 were used for qPCR. For mutants three biological replicates of stem tissue were used for qPCR analysis. For cDNA synthesis 1 ug of total RNAs was extracted from leaf, stem, meristem, root, flower, and fiber. The first strand of cDNA was synthesized using the M-MuLV reverse transcriptase (New England Biolabs, USA) and primed by d(T)25-VN following the manufacturer's instructions. qPCR of gene transcripts was carried out using an iCycler iQ system (Bio-Rad, Hercules, CA, USA) in 20μL of PCR reaction solution using the Luna Universal qPCR Master Mix, New England Biolabs, USA, the SYBR®Green method was used for running the qPCR. Conditions for thermal cycling included initial denaturation at 95°C for 120 s, followed by 45 cycles of 95°C for 20 s, 60°C 30 s, and 72°C for 20s. Lastly a unique melting curve was performed from 60.0°C to 95.0°C in 0.5°C increments to amplify distinctive PCR product. One reference gene, *GhPP2A1* was used for normalization of the expression data. The reference gene was chosen because of its uniform expression between cells of different tissues and under variable experimental conditions (Artico et al., 2010). The Ct values of four technical samples for each of the three biological replications were used to estimate the relative expression of genes using the $2^{-\Delta\Delta Ct}$ equation (Schmittgen and Livak, 2008). Except for the reference genes, specific primer pairs were designed from conserved coding sequence and listed in Supplementary Table S1.

Dual guide RNA design, transformation, validation of homeologous genome edits, and prediction of off-target site editing

The modified binary vector *pCSbar* (a *bar* gene cassette was cloned into *Eco*RI-*Hin*dIIIsites of pAMBIA1300) integrating *spCAS9* genes and two sgRNAs cassettes were prepared for cotton transformation (Supplementary Figure S1). The *spCAS9* driven by a dual 35S promoter was cloned into *Eco*RI site of *pCSbar*, and then the two synthesized sgRNAs (Synbio, NJ, USA) were integrated into the *Avr*II site of *pCSbar-Cas9*, and the result was in the final construct. The two crRNAs targeted the flanking sequences of a 128-base pair region specific to the second exon of the GhCoker.D11G112200 and GhCoker.A11G109300 driven by AtU6-26t and AtU6-29t promoters, respectively. PAM sequences are highlighted. The two crRNAs share 100% similarities to conserved regions of the two homologs GhCoker.D11G112200 and GhCoker.A11G109300 but low similarities to the other four homeologs of GhCoker.A08G143500.1, GhCoker.D08G158700.1, GhCoker.D12G137700.1 and GhCoker.A12G131200.1 (Supplementary Table S2).

The *Agrobacterium* (EHA105) mediated transformation and plant regeneration protocol was adopted from a standard transformation publication (Jin et al., 2012). The regenerated plants were transplanted into one-gallon pots (Dillen Products, Middlefield, OH) containing commercial potting mixture soil (Fafard 3-B Mix, Fafard Inc., Anderson, SC, USA) and were developed under the greenhouse system conditions (Biosystems Research Complex, Clemson University) for morphological observation and materials harvesting. Three independent transgenic events (*tac1-72*, *-73* and *-74*, verified by PCR of *bar* and *Cas9*) showing the phenotype were selected to verify the genotype. The total DNA obtained from each was subjected to PCR. The primer pair: F-AGATGGGCTTGCACGAAATGTTAAG and R-CGTTTTTGGCAGGAAGAGRAGATG were carefully selected to amplify the region overlapping the target sequences of the two crRNAs in the two homologous. The PCR products were cloned into pGEM-T-Easy vector and sequenced using standard Sanger sequencing techniques. The sequencing was conducted at the Genomics Core, Biosciences at Arizona State University. The chromatograms and sequence analysis were conducted in Geneious Prime (Supplementary Figure S2, Supplementary Table S3). To confirm the *tac1-73* knockout genotype, isolated gDNA was sequenced at Hudson Alpha Institute of Biotechnology, 2x150bp paired-end reads at an average coverage of 12.58X across the genome. Reads were trimmed using Trimmomatic (Bolger et al., 2014) to remove Illumina adaptor sequences, and aligned to the *G. hirsutum* 'Coker312' genome version 1.1 (*Gossypium hirsutum* Coker v1.1, DOE-JGI, http://phytozome.jgi.doe.gov/info/GhirsutumCoker_v1_1) using Burrow-Wheeler Alignment (bwa-mem) (Li, 2013). The aligned BAM file for the *tac1-73* knockout sample was investigated using Integrative Genomics Viewer (IGV) to verify the Crispr-Cas9 target sites (Robinson et al., 2011). BAM files were prepared for genotyping using samtools to eliminate duplicate reads (Danecek et al., 2021). Off-target single nucleotide polymorphisms were genotyped using samtools integrated with Varscan2 mpileup2snp (Koboldt et al., 2012; Danecek et al., 2021). Investigation of possible off-target indels or large structural variants was performed with Varscan2 mpileup2indel and Delly's structural variant caller (Koboldt et al., 2012; Rausch et al., 2012; Danecek et al., 2021). Guide RNAs were aligned to the Coker312 genome with BLASTN (wordsize=7) and only found 8 hits with identity, Supplementary Table S4. Bedtools (Quinlan and Hall, 2010) intersect was used to determine that no variants were found in these regions when intersected with the variant call file (.VCF), data not shown.

Phenotyping angle measurements

Branch and petiole angles were measured manually with a protractor and digitally in Adobe Photoshop to estimate the angle between the main stem to the branch and petiole (BioRender.com, 2023), Supplementary Figure S3 (Sun et al.,

2016a). Data was collected from five representative sympodial branches and two monopodial branches per plant and computed mean branch and petiole angles similar to (Li et al., 2017). We also collected data on branch length, plant height, boll count. Data was collected from the T_0 mutants and wild type Coker 312 plants.

Screening and verification of mutants

Stable integration of the CRISPR/Cas9 T-DNA was determined by screening the T_1 mutants using the *BAR* gene selectable marker. The *BAR* gene has been extensively utilized and researched as a positive selectable marker for herbicide resistance, providing plants with resistance to phosphinothricin (PPT), the active ingredient in the broad-spectrum herbicide known as Basta. This functionality facilitates the elimination of non-transformed individuals and the selective advancement of transformants in plant regeneration processes (Thompson et al., 1987; Rathore et al., 1993; Hahn et al., 2017).

Statistical analysis

The analysis of variance was performed using JMP Pro version 16.1 (SAS Institute, Cary, NC, USA). Mean separation was assessed through a one-way ANOVA with Tukey's HSD test ($p<0.05$).

Results

Phylogenetic analysis of *TAC1* in the angiosperms

Copy number and phylogenetic analysis was conducted to examine the distribution of cotton *TAC1* orthologs in a range of representative angiosperms, encompassing both species with known single copies (e.g. *Prunus persica, Zea mays, Vitus vinifera, Vigna unguiculata, Theobroma cacao, Medicago trunculata*, and *Eucalyptus grandis*) and those with two copies (*Glycine max* and *Populus trichocarpa*), Table 1. Analysis revealed that *TAC1* gene copy number in *Medicago truncatula*, a basal eudicot species, is most divergent from *V. unguiculata* in the phylogenetic tree (Figure 1). Our findings revealed a separate clade for the Rosaceae family members, *Prunus persica* (1 copy) and *Malus domestica* (2 copies), respectively. The analysis also showed a separate clade for the Fabaceae family member *V. unguiculata* and *G. max* both with 1 copy each. Notably, we detected an additional *TAC1* copy in the lineage containing the diploid cotton progenitor species *G. arboreum* and *G. raimondii*, each possessing three *TAC1* copies, indicating a *TAC1* duplication event specific to this lineage (Figure 1). As anticipated, within the *Gossypium* clade, the *TAC1* gene copy number remained consistent, with three copies in both the A and D subgenomes in all examined allotetraploid cotton species (*G. hirsutum* and *G. tomentosum*) examined (Figure 1; Table 2). As expected, the A-subgenome diploid progenitor, *Gossypium arboreum*, and the D-

TABLE 1 List of putative *TAC1* orthologs in cotton and other plant species.

Genus	Species	Ploidy	Copy number	Cotyledon	Reference
Gossypium	*Arboretum*	Diploid	3	Eudicot	(Li et al., 2015; Zhang et al., 2015)
Gossypium	*hirsutum (CSX8308)*	Allotetraploid	6	Eudicot	(Phytozome, 2022b)
Gossypium	*hirsutum (Coker 312)*	Allotetraploid	6	Eudicot	(Phytozome, 2022a)
Gossypium	*hirsutum (UA48)*	Allotetraploid	6	Eudicot	(Phytozome, 2022c)
Gossypium	*raimondii*	Diploid	3	Eudicot	(Paterson et al., 2012)
Gossypium	*tomentosum*	Allotetraploid	6	Eudicot	(Chen et al., 2020)
Arabidopsis	*thaliana*	Diploid	1	Eudicot	(Cheng et al., 2017)
Prunus	*persica*	Diploid	1	Eudicot	(Verde et al., 2013)
Populus	*trichocarpa*	Diploid	2	Eudicot	(Tuskan et al., 2006)
Malus	*domestica*	Diploid	2	Eudicot	(Daccord et al., 2017)
Glycine	*max*	Diploid	2	Eudicot	(Schmutz et al., 2010)
Eucalyptus	*grandis*	Diploid	1	Eudicot	(Myburg et al., 2014)
Medicago	*truncatula*	Diploid	1	Eudicot	(Tang et al., 2014)
Theobroma	*cacao*	Diploid	1	Eudicot	(Motamayor et al., 2013)
Vigna	*unguiculata*	Diploid	1	Eudicot	(Lonardi et al., 2019)
Vitus	*vinifera*	Diploid	1	Eudicot	(Jaillon et al., 2007)
Zea	*mays*	Diploid	1	Monocot	(Zhang et al., 2009))

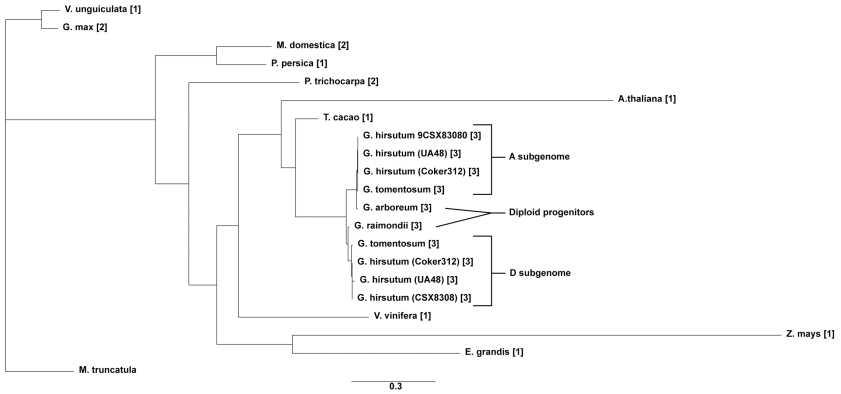


FIGURE 1
The phylogenetic tree of the TAC1 gene in cotton and other plant species. The tree was constructed by RAxML GAMMA GTR using the Neighbor-Joining method with 5000 bootstrap replicates (Stamatakis, 2014; Kozlov et al., 2019). The bar indicates an evolutionary distance of 2.0%.

subgenome diploid progenitor, *Gossypium raimondii*, are ancestral Table 1; Figure 1. Furthermore, the phylogenetic analysis also revealed a close relationship between *T.cacao* (cocoa) and the clade containing the cotton lineages highlighting the relationship of both *Theobroma cacao* and *Gossypium* spp. within the Malvaceae family.

The *TAC1* gene is close to 300 amino acids in length and is predicted to contain an NAD-dependent protein deacetylase domain from the PantherDB (Mi and Thomas, 2009), Figure 2. A multiple sequence alignment of *TAC1* from cotton and various plant species is only modestly similar with an average percent identity of 33% and only 30 (9%) identical sites (Figure 2). Alignments revealed the presence of conserved domains of the *TAC1* gene in other plant species, such as the IGT conserved motif (Figure 2), which is known to have an impact on vertical shoot growth in a broad range of plant species (Roychoudhry and Kepinski, 2015; González-Arcos et al., 2019). Similarity, homeologs also differed significantly among the *TAC1* copies. For example, the A11/D11 homeologs had high similarity at 96.8% at the amino acid level, while the A12/D12 and A08/D08 homeologs were 73% identical (Figure 3A; Supplementary Table S5). The regions of conserved sequence previously reported (Dardick et al., 2013) are highlighted as four domains and are conserved among all the homeologs, Figure 3A. Examination of the *GhTAC1* phylogeny within the allotetraploid shows A08/D08 and A11/D11 homeologous gene pairs showed separate clades forming from the A12/D12 gene pairs (Figure 3B).

Expression profiles of *GhTAC1* homeologs exhibit subgenome bias and vary by tissue type

To determine tissue-specific expression patterns of the six *GhTAC1* genes, we performed qPCR analysis in the Coker 312 cotton genotype at various stages of reproductive and vegetative growth utilizing gene and homoeolog-specific primers. Various *TAC1* homeologs displayed subgenome expression bias (A-subgenome dominance in all experiments) and differential expression in various cotton tissue sources (Figure 4A; Table 3). In vegetative tissues (leaf, stem, meristem, and roots) significant expression with subgenome bias was observed for the A11 *TAC1* homeolog (GhCoker.A11G109300) in stem tissue with over 10-fold expression compared to the other five homeologous copies (Figure 4B; Table 3). In leaf tissue GhCoker.A12G131200 was highly expressed compared to the other five homeologs with about a 2-fold expression compared to the second gene (GhCoker.D08G158700) that was highly expressed (Figure 4B and Table 3). Fairly low relative expression levels were observed in meristem and root tissue of all *GhTAC1* homeologous copies (Figure 4C; Table 3) which was expected, particularly in roots. Interestingly, in flower and fiber tissues, high expression levels were observed for GhCoker.A11G109300 (dominant copy) and GhCoker.D11G112200, respectively. The expression level of GhCokerA11.G109300 was over 7-fold higher in flower tissue and 20-fold higher in fiber compared to the other gene copies (Figure 4D; Table 3). The presence of multiple copies of the *GhTAC1* gene, along with its notably high expression in the flower tissue, may be the reason for the nearly horizontal growth pattern of the flower.

Targeted A11/D11 homeologous genome edits with dual guide RNAs

The dual guide RNAs used in this study to create a deletion were designed to target a 128-base pair (bp) region specific to the second exon of the GhCoker.D11G112200 and GhCoker.A11G109300 homeologous gene pair, Supplementary Table S6. Two types of mutations were observed at the target sites of SpCas9-edited lines that include a 94-base pair (bp) deletion (detected in most of the edited lines) and an 89bp inversion flanked by a 6bp and 10bp

10.3389/fpls.2023.1320638

TABLE 2 TAC1 gene distribution in cotton.

Genus_species (cultivar)	Gene name	Subgenome (A/D)	Chromosome	Paired
G. hirsutum (Coker312)	GhCoker.A12G131200.1	A	12	Yes
	GhCoker.D12G137700.1	D	12	
	GhCoker.A11G109300.1	A	11	Yes
	GhCoker.D11G112200.1	D	11	
	GhCoker.A08G143500.1	A	8	Yes
	GhCoker.D08G158700.1	D	8	
G. hirsutum (CSX8308)	GhCSX8308.A12G132100.1	A	12	Yes
	GhCSX8308.D12G133900.1	D	12	
	GhCSX8308.A11G111000.1	A	11	Yes
	GhCSX8308.D11G111700.1	D	11	
	GhCSX8308.A08G142400.1	A	8	Yes
	GhCSX8308.D08G156700.1	D	8	
G. hirsutum (UA48)	GhUA48.A12G135900.1	A	12	Yes
	GhUA48.D12G133200.1	D	12	
	GhUA48.A11G109300.1	A	11	Yes
	GhUA48.D11G112000.1	D	11	
	GhUA48.A08G145900.1	A	8	Yes
	GhUA48.D08G152400.1	D	8	
G. arboreum	Gar12G15480	A	12	N/A
	Gar11G10910	A	11	N/A
	Gar08G16700	A	8	N/A
G. raimodii	Gorai.008G127500.1	D	8	N/A
	Gorai.007G111200.1	D	7	N/A
	Gorai.004G150000.1	D	4	N/A
G. tometosum	Gotom.A12G133000.1	A	12	Yes
	Gotom.D12G139200.1	D	12	
	Gotom.A11G112900.1	A	11	Yes
	Gotom.A11G112900.1	D	11	
	Gotom.A08G147800.1	A	8	Yes
	Gotom.D08G162100.1			

N/A, Not applicable.

deletion, respectively, Figures 5A, B, Supplementary Table S7. The 94 bp deletion was found in both A and D subgenomes and included the PAM sequence, Figure 5A. In this event, the A and D subgenomes were clearly distinguished by numerous homeologous SNPs, arrow in Figure 5A. The 89 bp inversion flanked by 2 deletions that include approximately half of the gRNA1 and gRNA2 target sites was found only in the D-subgenome *TAC1* copy, (Figure 5B). The two edited lines with both 94bp deletions or a mosaic of the 94bp deletion and the 89bp inversion displayed the same branch/petiole angle inclination phenotype, Supplementary Figure S3.

Genetic mutations and segregation analysis in T_1 progeny

The T_1 seeds that were generated were sown to identify mutants and screen for individuals that may segregate with the genome editing reagents and the *TAC1* deletion. Among these mutants, *tac1*-73 and *tac1*-74 exhibited a favorable segregation ratio of 3:1, thus indicating a single copy (Figure 6). In our study, all 24 seeds of both *tac1*-73 and *tac1*-74 variants exhibited germination. Following screening for the presence of CAS9 in these mutants, 18 seedlings of *tac1*-73 and 17 of *tac1*-74 survived.

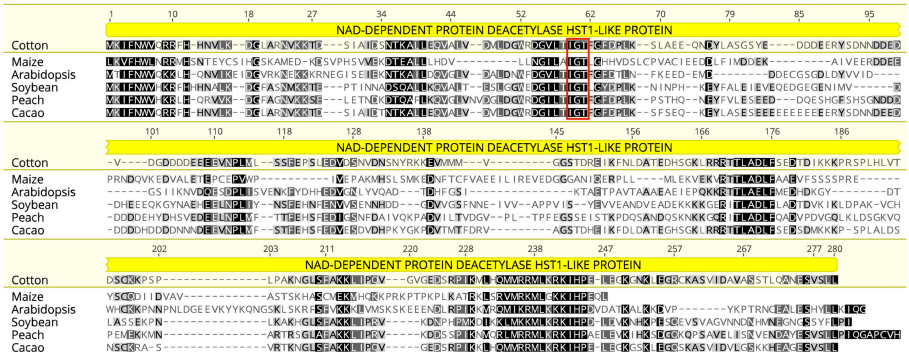


FIGURE 2
Multiple protein sequence alignment of predicted amino acid sequences of TAC1 orthologs cotton, peach, Arabidopsis, cacao, soybean, and maize. Highly conserved residues are highlighted in black with the IGT conserved domain indicated in red box at the 67-69bp position, which is typical of the IGT gene family.

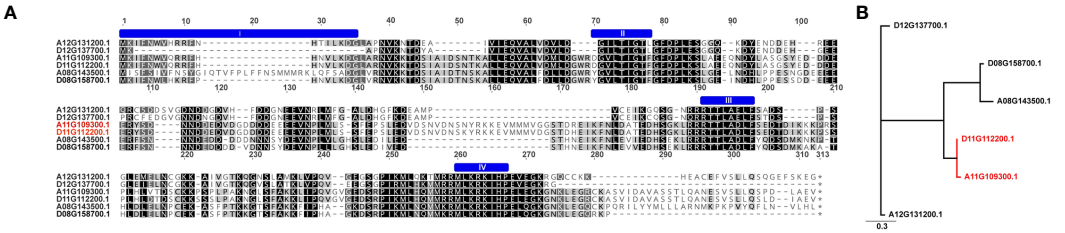


FIGURE 3
Multiple sequence alignment and dendrogram of TAC1 copies in (G) hirsutum (Coker312) and peach. **(A)** Six GhTAC1 protein sequences alignment. The previously reported 4 conserved domains (Dardick et al., 2013) are annotated by a blue "bar" above the sequence. **(B)** The homeologous copies on A12/D12 clade have the highest sequence identity to the single prunus copy. The six homologs within Coker312 are also quite diverged with only ~50% pairwise identity.

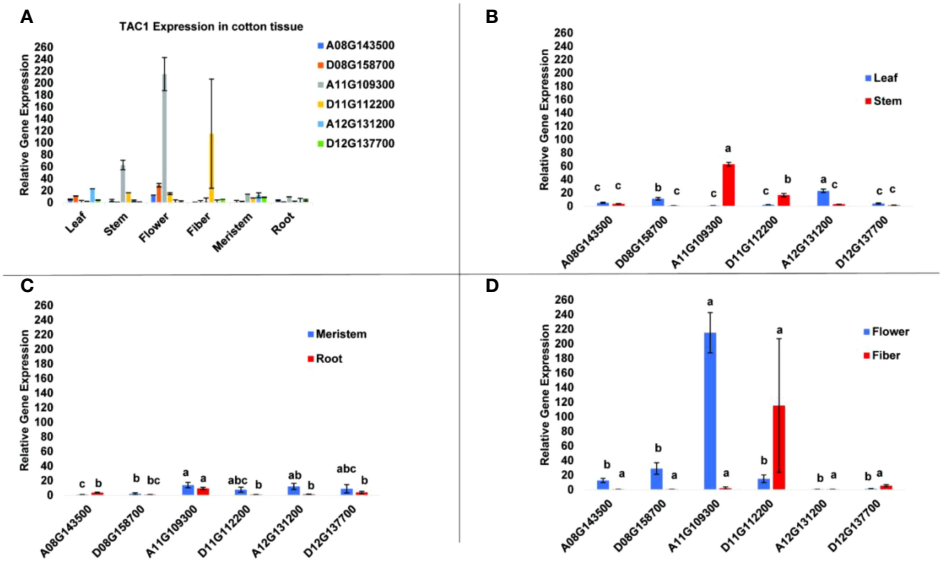


FIGURE 4
(A) Expression profiles of putative *TAC1* homologs in various cotton tissue. **(B)** leaf and stem, **(C)** meristem and root, **(D)** flower and fiber tissues. RT-qPCR was performed using template cDNA primed by Poly $T_{(25)}$ from total RNA that was isolated from 100 mg of tissue. The statistical difference between groups was determined by one-way ANOVA with a *post-hoc* Tukey HSD Test. Means not sharing the same letter are statistically significantly different ($P < 0.05$). All the data presented as the mean ± SD (n=3).

10.3389/fpls.2023.1320638

TABLE 3 Analysis of variance of TAC1 expression in multiple cotton tissues.

GENE	Leaf	Stem	Flower	Fiber	Meristem	Root
A08G143500	5.2a	3.5c	12.8b	1.0a	1.0a	3.8b
D08G158700	11.1b	1.0c	29.2b	1.0a	2.4a	1.0b
A11G109300	1.0c	63.0a	215.2a	2.4a	14.1a	9.5a
D11G112200	2.3c	16.6b	15.1b	115.4a	7.8a	1.0b
A12G131200	23.1a	3.0c	1.0b	1.0a	12.3a	1.4b
D12G137700	4.2a	1.3c	1.6b	5.6a	9.0a	4.1b
p-value (≤ 0.05)	<.0001	<.0001	<.0001	0.246	0.197	<.0001

Means not sharing the same letter are statistically significantly different (P < 0.05).

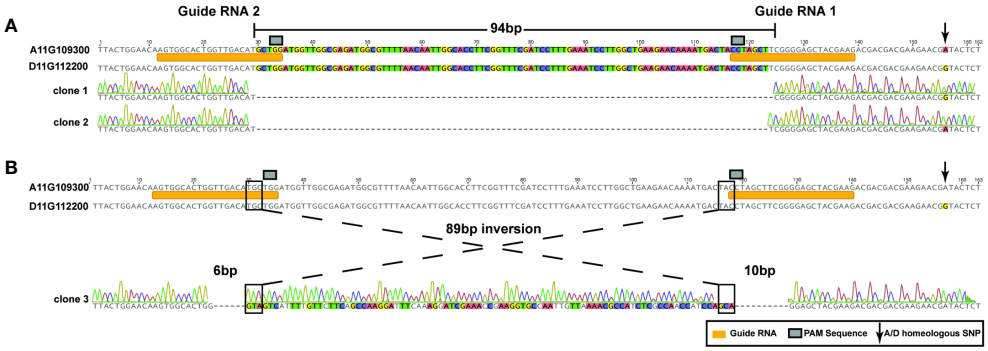


FIGURE 5
2 types of homeologous GhTAC1 edits by the CRISPR/Cas9 system in Coker312. **(A)** A 94-bp deletion that includes the PAM motif and a portion of the guide RNA. This deletion was found in the following events: tac1-72 (A and D subgenome) and in the A-subgenome of tac1-73 and tac1-74. **(B)** An 89bp inversion flanked by 6bp and 10bp deletions that include the PAM sequence and a portion of each guide RNA, respectively.

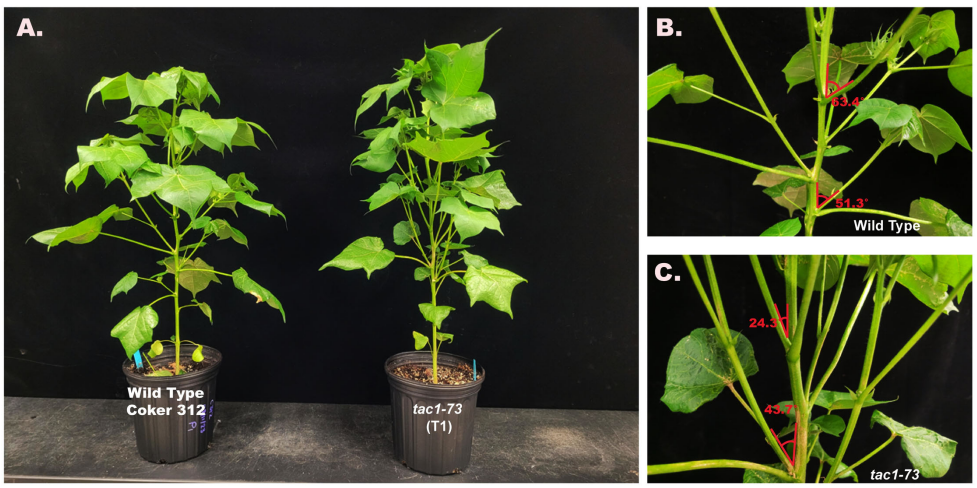


FIGURE 6
(A) Representative images depicting columnar phenotype inclination in wild type (Coker 312) *vs tac1-73* (T_1 generation). **(B)** Branch angle measurements of wild type (Coker 312) and **(C)** Branch angle measurements and *tac1-73* (T_1 generation).

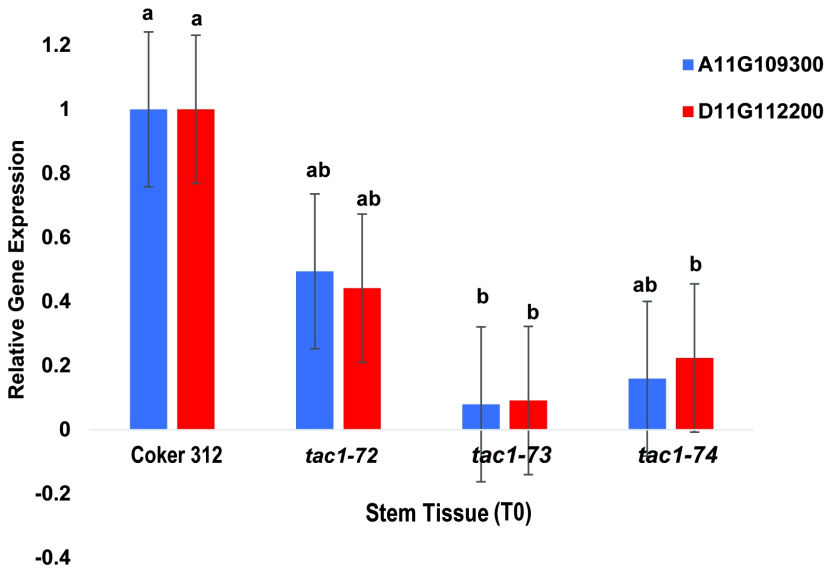


FIGURE 7
qPCR results of *GhTAC1* expression in stem tissue from wild type Coker 312 and *tac1-72*, *tac1-73*, and *tac1-74*. Error bars represent standard deviations of three biological replicates. Means not sharing the same letter are statistically significantly different ($P < 0.05$).

These results verify that the knockout strategy for the mutants developed was successful and could be further validated by the phenotype observed five weeks after emergence (Figure 6A). The T_1 plants of *tac1*-73 showed columnar branch formation as depicted in more inclined branch angles (Figure 6B) compared to wild type Coker 312 (Figure 6C).

Validation of the expression profiles of *GhTAC1* homeologs in mutants

To assess the expression profiles of *TAC1* genes and specifically investigate the knockout effects on the homoeologous gene pairs GhCoker.A11G109300 and GhCoker.D11G112200 in *TAC1* mutant lines (*tac1*-72, *tac1*-73, and *tac1*-74) in comparison to the wild-type Coker 312, a quantitative polymerase chain reaction (qPCR) analysis was conducted using stem tissue samples. Homoeolog-specific primers were designed to target the 128bp knockout region within the second exon of GhCoker.A11G109300 and GhCoker.D11G112200 homoeologous gene pair. Remarkably, all three mutant lines exhibited significantly reduced expression levels of both GhCoker.A11G109300 and GhCoker.D11G112200 homoeologous gene pairs (Figure 7). Notably, when compared to the wild-type Coker 312, the *tac1*-73 and *tac1*-74 mutants displayed very low expression, with reductions of approximately 10-fold and 5-fold in *tac1*-73 and *tac1*-74 mutants, respectively. However, the *tac1*-72 mutant demonstrated a more moderate 2-fold reduction in expression. Based on the gene-specific expression patterns, GhCoker.A11G109300 (the dominant copy) exhibited lower expression levels in all three mutant plant lines when compared to GhCoker.D11G112200, except in *tac1*-72 where GhCoker.D11G112200 showed lower expression in comparison to GhCoker.A11G109300 (Figure 7).

Alteration of A11/D11 *TAC1* expression leads to altered branch angle and impacts cotton morphology

To explore the role of *GhTAC1* in branch angle, the A11/D11 homoeologs were targeted because of their high levels of expression in stem tissue. Transgenic plants harboring targeted knockouts of the A11/D11 homoeologous gene pair were generated. Plants with the validated *GhTAC1* (A11/D11) knockouts (mutant) at the T_0 stage displayed significant differences in branch inclination angles when compared to wild type Coker 312 plants (Figure 8A). Branch inclination angles were measured at flower bud formation. Mean branch inclination angles widely differed among the transgenic A11/D11 knockouts (transgenic positive plant) and wild type Coker 312 for both monopodial and sympodial branches (Figures 8A–C). The phenotyping results revealed significant differences in branch inclination angles among the different plant variants. For sympodial branches the mutant exhibited an angle of 51.2° while the wild-type Coker 312 measured 74.6°. On the other hand, no significant differences were observed in monopodial branch angles, with the mutant measuring 47.5° and the wild type Coker 312 being 57.5°. (see Figure 8B; Table 4). Regarding petiole angles, the mutant showed narrower angles in sympodial branches compared to wild type Coker 312. However, no significant difference was found in monopodial branches (see Figure 8C; Table 5). Furthermore, it was noted that the mutant plants displayed increased branching and leaf biomass in the basal region while displaying fewer branches and leaves in the apical region though not significantly different from the wild-type Coker 312 plant (Supplementary Figures S4-S6). To assess the impact of *GhTAC1* on cotton morphology, measurements were taken for boll count, plant height, and branch length. The analysis revealed no significant differences in these variables between the wild-type Coker 312 and the mutants (Table 6).

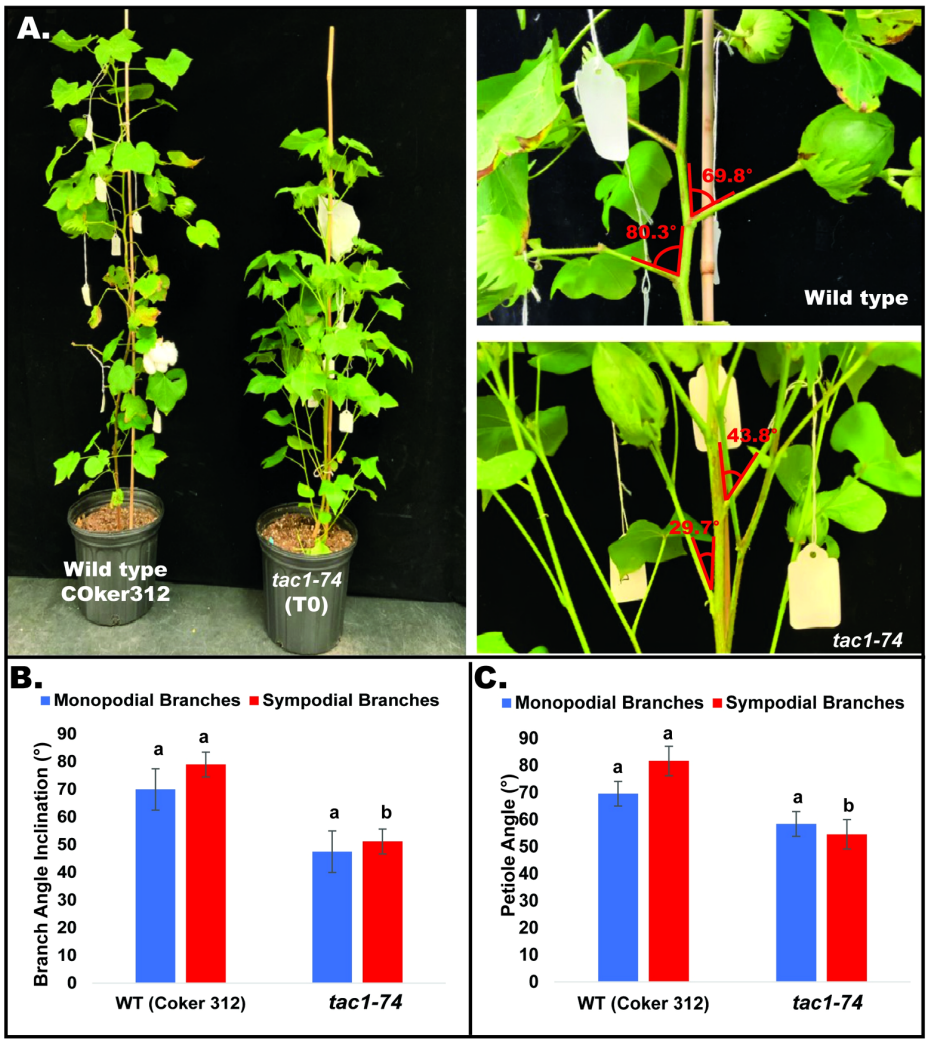


FIGURE 8
(A) Representative images depicting columnnar phenotype inclination in wild type (Coker 312) and mutant plant. **(B)** Mean branch angle measurements of wild type (Coker 312) and mutant plant (*tac1-74*). **(C)** Mean petiole angle measurements of wild type (Coker 312) and mutant plant (*tac1-74*). The statistical difference between groups was determined by one-way ANOVA with a *post-hoc* Tukey HSD Test. Means not sharing the same letter are statistically significantly different ($P < 0.05$). Errors bars represent SD.

Discussion

Upland cotton is a commonly cultivated fiber crop for the textile industry. Meeting the increasing food and fiber demands of the rising global population and addressing unpredictable climatic shifts requires enhancing crop productivity on the same or reduced land area, thus using the same or fewer resources. Combining higher-density planting with improved photosynthetic capacity is a promising approach to achieving this objective. This would maximize resource utilization while optimizing the plant's energy capture and conversion efficiency resulting in higher crop yields with resource efficiency, thereby benefiting growers and consumers. Understanding the impact of plant architecture on cotton growth is paramount to enable the breeding of cotton varieties with enhanced

TABLE 4 Analysis of variance for branch angle inclination and petiole Angle.

Genotype	Mean Branch Angle		Mean Petiole Angle	
	Monopodial Branches	Sympodial Branches	Monopodial Branches	Sympodial Branches
Wild Type (Coker 312)	57.5a	74.6a	69.6a	81.7a
tac1-74	47.5a	51.2b	58.4a	54.6b
p-value (≤ 0.05)	0.3333	0.0075	0.1328	0.0011

Means not sharing the same letter are statistically significantly different ($P < 0.05$).

TABLE 5 Analysis of variance for TAC1 expression in stem tissue of wild type (Coker 312) and mutants (tac1-72 tac1-73 and tac1-74).

Primer	A11G109300	D11G112200
Coker 312	1.0a	1.0a
tac1-72	0.5ab	0.4ab
tac1-73	0.1b	0.1b
tac1-74	0.2ab	0.2b
p-value (≤0.05)	0.04	0.02

Means not sharing the same letter are statistically significantly different (P < 0.05).

architectural traits (Kaggwa-Asiimwe et al., 2013; Hollender et al., 2018).

The genetic modification of plant architecture can offer several advantages such as increased carbon assimilation, improved light utilization in dense crop canopies, heightened mechanical process efficiency, decreased susceptibility to insects and diseases by reducing canopy humidity, and the potential for higher yields. The growth and yield of cotton plants are significantly impacted by light penetration, especially in dense cotton fields where the upper parts of the plant receive most of the light. This uneven distribution of light within the branch canopy critically impacts optimal growth and yield. Light distribution is influenced by various factors, such as genetics, environment, and management practices (Mao et al., 2014; Yao et al., 2017). Cotton varieties with columnar canopies tend to have an open structure that facilitates better radiance interception and light penetration throughout the canopy (Chapepa et al., 2020). The positioning of bolls is also crucial for yield, with the first position bolls proportionally contributing to plant yield due to their larger size and weight. To produce these bolls, older leaves must produce enough photosynthates, which require even and adequate light penetration throughout the plant (Jiang et al., 2012; Hollender et al., 2018; Huang et al., 2022). Therefore, developing cotton cultivars with modified plant architecture can significantly improve crop performance and have far-reaching benefits for the cotton industry. Here, we evaluate the genomic content, phylogenetic distribution, and functional implications of *TAC1* in cotton.

Phylogenetic analysis of *TAC1* revealed between 1 and 2 copies in representative diploid Angiosperms, except for the diploid progenitors of the *Gossypium* lineage (*G. arboreum* and *G. raimondii*, which contained 3 copies, Figure 1; Table 1, indicating that the *TAC1* gene duplication is specific to this lineage. In *Prunus*, a species with a single *TAC1* copy, altered expression of this gene resulted in pleiotropic shoot phenotypes. For example, silencing of *TAC1* resulted in plum trees with severely vertical branch orientations, while overexpression resulted in trees with more horizontal branch orientations. Collectively, alteration of *TAC1* in *Prunus* species leads to pleiotropic shoot phenotypes (Dardick et al., 2013; Hollender et al., 2020). Similarly in *Arabidopsis* plants with a single copy of the *TAC1* gene, mutant plants display a more inclined branch angle relative to the wild type (Hollender et al., 2020). A study on the functional characterization of *TAC1* in *Populus trichocarpa* (poplar), which contains two copies of the gene (Potri.014G102600 and Potri.002G175300), revealed that knocking out these genes resulted in narrower leaf angles and upright shoot growth (Fladung, 2021). In *Malus domestica*, two members of the *TAC1* gene (*MdTAC1a* and *MdTAC1b*) have been identified. Subcellular localization analysis of *MdTAC1a* showed that it is detected in the nucleus and cell membrane, while *MdTAC1b* is detected only in the cell membrane. Both genes play a role in regulating branch inclination in *Malus domestica*, and they are highly expressed in the shoot tips and vegetative buds of weeping cultivars. However, they exhibit weak expression in columnar cultivars (Li et al., 2022).

Our results revealed three copies in the diploid progenitor cotton species (*G. arboreum* and *G. raimondii*) indicating the duplication event supersedes the polyploidization event in allotetraploid cotton and underscores the potential importance retaining this gene duplication event in success of both the wild and domesticated allotetraploid cottons (Cronn et al., 1999; Meng et al., 2020). Visual comparisons of cotton plants with various Angiosperms reveals that most cotton genotypes have near horizontal branches which could be attributed to the presence of additional *TAC1* alleles. However, certain naturally occurring mutant cotton varieties exhibit more acute branch angles compared to their wild relatives, which could be a result of gene regulation ((Ji et al., 2021). Moreover, previous studies have identified two types of cotton mutants, known as cluster branching (*cl1*) in *G. hirsutum* (Upland cotton) and short branch (*cl2*) in *G. barbadense* (Pima cotton), which exhibit a distinct phenotype called "*nulliplex-branch*" (*nb*). These mutants display determinate sympodial growth (Kearney, 1930; Pathak and Singh, 1975). The *cl1* mutant is associated with a recessive allele located on chromosome D07, while the *cl2* mutant is associated with a recessive allele located on chromosome A07 (Stephens, 1955; Endrizzi and Ray, 1992). The *nulliplex-branch* mutants demonstrate unique characteristics, such as flowers developing directly from leaf axils on the main stem or occasionally from a

TABLE 6 Analysis of variance for some morphological traits of wild type (Coker 312) and mutants (tac1-72 tac1-73 and tac1-74).

Genotype	Monopodial branch length	Sympodial branch length	Boll Count	Plant Height(cm)
Coker 312 (WT)	32.8a	16.3a	14a	90.4a
tac1-72	37.3a	15.0a	14a	100.8a
tac1-73	38.1a	15.2a	15a	102.9a
tac1-74	38.2a	15.0a	14a	101.6a
p-value (≤0.05)	0.154	0.901	0.983	0.223

Means not sharing the same letter are statistically significantly different (P < 0.05).

short branch with a single node. These phenotypes typically lack fruiting branches and possess a compact plant structure. As a result, they are highly suitable for high-density planting and mechanical harvesting. Additionally, they mature early, and their growth does not require chemical regulation or manual pruning, which reduces labor inputs (Silow, 1946; McGarry et al., 2016; Sun et al., 2016b; Si et al., 2018; Chen et al., 2019; Huang et al., 2022; Sun et al., 2022).

Sequence similarity analysis of a representative *TAC1* gene in cotton showed expected regions of high identity when compared to other species (cacao, peach, soybean, *Arabidopsis*, and maize), Figure 2. Protein family analysis of *TAC1* identified an NAD-Dependent protein deacetylase HST1-like domain (Panther 38366). This functional domain (histone deacetylation) is involved in telomeric silencing and methylation maintenance. Other studies have shown that LAZY and DRO interact with a protein that has an RCC1 (regulator of chromatin condensation) domain. The strong sequence similarity between the *TAC1* gene in Arabidopsis and the wild diploid cotton species *Gossypium raimondii* confirms the functional role of *TAC1* in wild cotton. This finding supports the notion that domestication processes resulted in targeted artificial selection for agronomic traits, including yield, pest resistance, fiber length, compact branching, and reduced growth and maturity periods. The multiple alignment of *TAC1* amino acid sequences (Figure 2) also showed structural similarities in the *IGT* domain, suggesting that these ancient IGT genes play a consistent role in determining shoot growth angle orientation (Dardick et al., 2013; Dong et al., 2016; Montesinos et al., 2021).

An analysis of the six homeologous copies of *TAC1* in the Coker312 genotype revealed intriguing differences in the level of homeologous and homologous similarity among the six *TAC1* copies at the sequence identity level. Notably, the A11/D11 copies of *TAC1* exhibited a 96.8% identity, while their similarity with homologs on chromosome 8 or 12 was only up to 58% (Supplementary Table 1). We observed that this pair of *TAC1* homologs displayed dominant tissue-specific expression when compared to the other homologs. Additionally, their expression patterns displayed subgenome bias as shown in Figures 3A, B, 4A. The A08/D08 and A12/D12 copies exhibited a sequence identity of approximately 74% between homeologs, respectively (refer to Supplementary Table 1 for details). Significantly, the expression patterns of A11 and A12 *TAC1* homologs in various vegetative tissues, including leaf, stem, meristem, and root, revealed a distinct prevalence of expression in leaf and stem tissues (Figure 4). Previous studies have demonstrated that *TAC1* exhibits increased expression in actively growing vegetative buds, despite its low expression levels in meristem tissue (Xu et al., 2017). The low expression of *TAC1* in root tissue aligns with its regulatory role in upper branches of plants, with minimal or no expression in lower branches (Dardick et al., 2013). Our analysis showed a notable difference in the expression levels of A11 and D11 *TAC1* copies, with A11 exhibiting significantly high expression in flower tissue and D11 displaying the highest expression in fiber tissue. These results highlight the subgenome expression bias and suggest that the functional domains of the A11/D11 coding sequence (CDS) may be redundant, while their regulatory regions (5' and 3' UTRs and promoters) could be differentially activated in various tissues and developmental stages (Figure 4). Previous studies have indicated that *TAC1* exhibits high expression in reproductive structures such as flower buds, which may explain the horizontal growth of the flower and its subsequent development into fiber (Dardick et al., 2013) The A12/D12 copies displayed higher expression levels in vegetative tissues, with the A12 copy showing subgenome expression bias over D12 in leaf, stem, meristem, and root tissues, indicating the possibility of functional redundancy or sub-functionalization with a role in leaf angle.

Targeted knockout of the A11/D11 homeologous *TAC1* copies resulted in a cotton plant with an induced columnar phenotype with distinct branch and petiole inclination, Figure 6; Supplementary Figure S5 and S6. Overall, the mutant plants consistently exhibited narrower measurements in terms of branch inclination angle and petiole angle for sympodial branches. However, there were no notable differences observed in the measurements of monopodial branches and petiole angle for both the mutant and wild type Coker 312 (Figures 6B, C). No variations were observed in some morphological traits, including branch length (monopodial and sympodial branches), plant height, and boll count (Table 6). Sequence analysis of the mutants from three independent events with columnar branching (*tac1-72*, *tac1-73*, and *tac1-74*) revealed a high efficiency of the CRISPR-induced mutations (Figure 5A, B) resulting in hemizygous/homozygote mutations with large deletions and a fragment rearrangement (Supplementary Tables S6, S7). In *tac1-72* and *tac1-73* mutants (Figure 5A) a 94bp deletion was observed in both A11/D11 homeologs. Interestingly, the analysis of *tac1-74* mutant revealed an 89bp inversion flanked by two deletions at the target site in D11 and the 94bp deletion in A11 (Figure 6B). These mutations altered the *GhTAC1* gene function, rendering it non-functional, and subsequently led to the same observed columnar phenotype in all independent events. These findings align with previous studies in other Eudicots which have shown a correlation between mutations in *TAC*1 and the development of a nearly erect plant architecture phenotype (Dardick et al., 2013; Hollender et al., 2018; Hollender et al., 2020; Fladung, 2021; Dutt et al., 2022; Li et al., 2022). Intriguingly, we noted an enhanced canopy structure in the mutants accompanied by a significant increase in the number of leaves, auxiliary buds, and subtending leaves, particularly in the basal region of the plants compared to the apical region (see Supplementary Figures S5, S6). This pattern is characteristic of columnar plant phenotypes which have the potential to improve light distribution within the plant canopy (Chapepa et al., 2020). Validation of the expression of both A11/D11 homeologs in stem tissue revealed a decrease in expression levels for A11/D11 pairs. This decline in expression for both gene pairs contributes to the columnar phenotype observed in the mutants. These findings are consistent with earlier research which indicates that the loss/mutation of *TAC1* gene function results in decreased expression in mutant varieties (Dardick et al., 2013; Hollender et al., 2020).

The segregation analysis of T_1 progeny derived from the mutants, particularly *tac1-73* and *tac1-74*, revealed a notably favorable segregation ratio, emphasizing a single copy and a confirmation supported by the presence of resistance to the bar gene selectable marker (Supplementary Figures S5, S6). These findings provide strong evidence of transformation efficiency.

Several studies have indicated that the canopy microenvironment plays a critical role in regulating not only the distribution of light but also the temperature and relative humidity within the canopy. These factors collectively influence the radiation received by plants (Yang et al., 2014; Chapepa et al., 2020). The findings of this study are consistent with previous studies in peach and poplar, where mutant varieties with loss of function of the *TAC1* gene showed pronounced nearly upright growth of branches (Dardick et al., 2013; Fladung, 2021). The potential benefits of upright branch growth in peach, poplar, and other woody Eudicots include high density planting, increased yield, automated harvesting, and reduced pruning (Dardick et al., 2013; Fladung, 2021). Harnessing knowledge of plant architecture offers the potential for ultimately enhancing cotton productivity and increasing higher yields.

Conclusion

Plant architecture undergoes dynamic changes throughout their development. Our study revealed that the *TAC1* gene is duplicated and unique to the Gossypium lineage within the Angiosperms. Interestingly, we observed a correlation between the number of native *TAC1* copies and the horizontal branching pattern, with cotton exhibiting the most pronounced lateral growth among the Angiosperms. We also confirmed CRISPR-mediated knockout of the highly expressed copies of the *TAC*1 gene (A11G109300 and D11G112200) led to a significant reduction in both branch and petiole angles. The results of this study demonstrate the potential of gene editing technology to introduce novel traits into high-value crops, such as cotton, that would otherwise not exist naturally because of the presence of multiple subgenomes. These findings have significant implications for improving planting density, understanding optimal light interception and the correlation with yield potential, and enhancing our understanding of the evolution and domestication of the *Gossypium* lineage.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

FK: Investigation, Writing – original draft, Methodology. SK: Methodology, Writing – original draft, Supervision, Writing – review & editing. ZL: Supervision, Writing – review & editing, Data curation, Formal Analysis. AS: Data curation, Formal Analysis, Writing – review & editing. CD: Conceptualization, Formal Analysis, Writing – original draft, Writing – review & editing. DJ: Resources, Writing – review & editing. CS: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by Cotton Incorporated project number 20-655.

Acknowledgments

The authors extend appreciation to Stephen Parris, Lauren Carneal, Meredith Cobb, Quinn LaFave and Taylor Martin for assistance in branch angle data collection and maintenance of experimental plants in the greenhouse facility.

Conflict of interest

Author DJ was employed by the company Cotton Incorporated.

The remaining 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1320638/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Dual guide RNA design

SUPPLEMENTARY FIGURE 2
The total DNA amplified using primer pair: F-AGATGGGCT TGCACGAAATGTTAAG and R-CGTTTTTGGCAGGAAGAGRAGATG. The PCR products were cloned into pGEM-T-Easy vector and sequenced using standard Sanger sequencing techniques. The sequence alignments were conducted by Geneious Prime. For alignment, the sequences of A11G109300 and D11G112200 are from Coker312 whole genome sequence, the sequences of A1_Sp6 to A18_Sp6 are from sanger sequencing results. The 95-bp deletion showed in Clone A3-A4, the 95-bp deletion showed in Clone A1, A2, clone A6, A7, clone A10-A12, A14-A16 and A18; the one bp insertion (A) found in Clone A8 and A13; the 89-bp inversion detected in Clone A9 and A17. The highlighted line arrow indicated SNP between A11G109300 and D11G112200.

SUPPLEMENTARY FIGURE 3
A cartoon image describing the branch and petiole angle was measured in the study. In measuring branch angles, a line was drawn between the branch point and 2cm down the branch, and the angle with respect to the line and stem was recorded. For petiole angle measurements, a line was drawn

tangent to the branch and 1cm back from the last petiole on the branch, and the angle of this line with respect to the angle of the stem was recorded (Adapted from "Cotton Plant", Adapted from "Cotton Plant" by BioRender.com, (2023). Retrieved from https://app.biorender.com/biorender-templates.

SUPPLEMENTARY FIGURE 4
Branch and leaf count at the apical part (at 45cm height)

SUPPLEMENTARY FIGURE 5
Images of mutant plants (*tac1-72*, *tac1-73*, and *tac1-74*)

SUPPLEMENTARY FIGURE 6
Representative images illustrating the comparison between Wild Type Coker 312 and T_1 generation of mutant plants (*tac1-73* and *tac1-74*) at vegetative growth (before squaring).

References

Adams, C., Thapa, S., and Kimura, E. (2019). Determination of a plant population density threshold for optimizing cotton lint yield: A synthesis. *Field Crops Res.* 230, 11–16. doi: 10.1016/j.fcr.2018.10.005

Artico, S., Nardeli, S. M., Brilhante, O., Grossi-de-Sa, M. F., and Alves-Ferreira, M. (2010). Identification and evaluation of new reference genes in Gossypium hirsutumfor accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biol.* 10 (1), 49. doi: 10.1186/1471-2229-10-49

Basu, U., and Parida, S. K. (2021). Restructuring plant types for developing tailor-made crops. *Plant Biotechnol. J.* 21 (6), 1106-1122. doi: 10.1111/pbi.13666

BioRender.com. (2023). *Cotton plant*. Available at: https://app.biorender.com/biorender-templates (Accessed 2023).

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Cai, G., Yang, Q., Chen, H., Yang, Q., Zhang, C., Fan, C., et al. (2016). Genetic dissection of plant architecture and yield-related traits in Brassica napus. *Sci. Rep.* 6 (1), 21625. doi: 10.1038/srep21625

Chapepa, B., Mudada, N., and Mapuranga, R. (2020). The impact of plant density and spatial arrangement on light interception on cotton crop and seed cotton yield: an overview. *J. Cotton Res.* 3 (1), 18. doi: 10.1186/s42397-020-00059-z

Chen, W., Yao, J., Li, Y., Zhao, L., Liu, J., Guo, Y., et al. (2019). Nulliplex-branch, a TERMINAL FLOWER 1 ortholog, controls plant growth habit in cotton. *Theor. Appl. Genet.* 132 (1), 97–112. doi: 10.1007/s00122-018-3197-0

Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., et al. (2020). Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. *Nat. Genet.* 52 (5), 525–533. doi: 10.1038/s41588-020-0614-5

Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* 89 (4), 789–804. doi: 10.1111/tpj.13415

Cronn, R. C., Small, R. L., and Wendel, J. F. (1999). Duplicated genes evolve independently after polyploid formation in cotton. *Proc. Natl. Acad. Sci. U.S.A.* 96 (25), 14406–14411. doi: 10.1073/pnas.96.25.14406

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., et al. (2017). High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* 49 (7), 1099–1106. doi: 10.1038/ng.3886

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10 (2), 1-4. doi: 10.1093/gigascience/giab008

Dardick, C., Callahan, A., Horn, R., Ruiz, K. B., Zhebentyayeva, T., Hollender, C., et al. (2013). PpeTAC1 promotes the horizontal growth of branches in peach trees and is a member of a functionally conserved gene family found in diverse plants species. *Plant J.* 75 (4), 618–630. doi: 10.1111/tpj.12234

Dong, H., Zhao, H., Xie, W., Han, Z., Li, G., Yao, W., et al. (2016). A novel tiller angle gene, TAC3, together with TAC1 and D2 largely determine the natural variation of tiller angle in rice cultivars. *PloS Genet.* 12 (11), e1006412. doi: 10.1371/journal.pgen.1006412

Dutt, M., Mahmoud, L. M., Nehela, Y., Grosser, J. W., and Killiny, N. (2022). The Citrus sinensis TILLER ANGLE CONTROL 1 (CsTAC1) gene regulates tree architecture in sweet oranges by modulating the endogenous hormone content. *Plant Sci.* 323, 111401. doi: 10.1016/j.plantsci.2022.111401

Duursma, R. A., Falster, D. S., Valladares, F., Sterck, F. J., Pearcy, R. W., Lusk, C. H., et al. (2012). Light interception efficiency explained by two simple variables: a test using a diversity of small- to medium-sized woody plants. *New Phytol.* 193 (2), 397–408. doi: 10.1111/j.1469-8137.2011.03943.x

Duvick, D. N. (2005). "The Contribution of Breeding to Yield Advances in maize (Zea mays L.)," in *Advances in agronomy* (Ames, Iowa: Academic Press), 83–145.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf.* 5 (1), 113. doi: 10.1186/1471-2105-5-113

Endrizzi, J., and Ray, D. (1992). Mapping of the cl1, R1, yg1, and Dw loci in the long arm of chromosome 16 of cotton. *J. Heredity* 83 (1), 1–5. doi: 10.1093/oxfordjournals.jhered.a111148

Fladung, M. (2021). Targeted CRISPR/cas9-based knock-out of the rice orthologs TILLER ANGLE CONTROL 1 (TAC1) in poplar induces erect leaf habit and shoot growth. *Forests* 12 (12), 1615. doi: 10.3390/f12121615

Fok, M. A. C. (1998). "Cotton yield stagnation addressing a common effect of various causes," in *World cotton research conference* (U.S.A. Ccic: ICAC), 5–11.

González-Arcos, M., Fonseca, M. E., Zandonadi, D., Peres, L., Arruabarrena, A., Silva ferreira, D., et al. (2019). A loss-of-function allele of a TAC1-like gene (SlTAC1) located on tomato chromosome 10 is a candidate for the Erectoid leaf (Erl) mutation. *Euphytica* 215, 1–16. doi: 10.1007/s10681-019-2418-1

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40 (Database issue), D1178–D1186. doi: 10.1093/nar/gkr944

Gore, U. R. (1935). Morphogenetic studies on the inflorescence of cotton. *Botanical Gazette* 97 (1), 118–138. doi: 10.1086/334540

Guseman, J. M., Webb, K., Srinivasan, C., and Dardick, C. (2017). DRO1 influences root system architecture in Arabidopsis and Prunus species. *Plant J.* 89 (6), 1093–1105. doi: 10.1111/tpj.13470

Hahn, F., Mantegazza, O., Greiner, A., Hegemann, P., Eisenhut, M., and Weber, A. P. M. (2017). An efficient visual screen for CRISPR/cas9 activity in arabidopsis thaliana. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00039

Hollender, C. A., Hill, J. L.Jr., Waite, J., and Dardick, C. (2020). Opposing influences of TAC1 and LAZY1 on lateral shoot orientation in arabidopsis. *Sci. Rep.* 10 (6051), 1–13. doi: 10.1038/s41598-020-62962-4

Hollender, C. A., Waite, J. M., Tabb, A., Raines, D., Chinnithambi, S., and Dardick, C. (2018). Alteration of TAC1 expression in Prunus species leads to pleiotropic shoot phenotypes. *Hortic. Res.* 5, 26. doi: 10.1038/s41438-018-0034-1

Huang, X., Liu, H., and Ma, B. (2022). The current progresses in the genes and networks regulating cotton plant architecture. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.882583

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449 (7161), 463–467. doi: 10.1038/nature06148

Ji, G., Liang, C., Cai, Y., Pan, Z., Meng, Z., Li, Y., et al. (2021). A copy number variant at the HPDA-D12 locus confers compact plant architecture in cotton. *New Phytol.* 229 (4), 2091–2103. doi: 10.1111/nph.17059

Jiang, J., Tan, L., Zhu, Z., Fu, Y., Liu, F., Cai, H., et al. (2012). Molecular evolution of the TAC1 gene from rice (Oryza sativa L.). *J. Genet. Genomics* 39 (10), 551–560. doi: 10.1016/j.jgg.2012.07.011

Jin, S.-x., Liu, G.-z., Zhu, H.-g., Yang, X.-y., and Zhang, X.-l. (2012). Transformation of upland cotton (Gossypium hirsutum L.) with gfp gene as a visual marker. *Journal of Integrative Agriculture* 11 (6), 910–919. doi: 10.1016/s2095-3119(12) 60081-1

Kaggwa-Asiimwe, R., Andrade-Sanchez, P., and Wang, G. (2013). Plant architecture influences growth and yield response of upland cotton to population density. *Field Crops Res.* 145, 52–59. doi: 10.1016/j.fcr.2013.02.005

Kant, S., Seneweera, S., Rodin, J., Materne, M., Burch, D., Rothstein, S., et al. (2012). Improving yield potential in crops under elevated CO2: Integrating the photosynthetic and nitrogen utilization efficiencies. *Front. Plant Sci.* 3. doi: 10.3389/fpls.2012.00162

Kearney, T. H. (1930). A study of hybrids in Egyptian cotton. *J. Agric. Res.* 52 (622/623), 16.

Khan, N., Han, Y., Xing, F., Feng, L., Wang, Z., Wang, G., et al. (2020). Plant density influences reproductive growth, lint yield and boll spatial distribution of cotton. *Agronomy* 10 (1), 14. doi: 10.3390/agronomy10010014

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22 (3), 568–576. doi: 10.1101/gr.129684.111

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35 (21), 4453–4455. doi: 10.1093/bioinformatics/btz305

Ku, L., Wei, X., Zhang, S., Zhang, J., Guo, S., and Chen, Y. (2011). Cloning and characterization of a putative TAC1 ortholog associated with leaf angle in maize (Zea mays L.). *PloS One* 6 (6), e20621. doi: 10.1371/journal.pone.0020621

Kumar, S., Ruggles, A., Logan, S., Mazarakis, A., Tyson, T., Bates, M., et al. (2021). Comparative transcriptomics of non-embryogenic and embryogenic callus in semi-recalcitrant and non-recalcitrant upland cotton lines. *Plants (Basel)* 10 (9), 1-16. doi: 10.3390/plants10091775

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*, 1-13.

Li, F., Fan, G., Lu, C., Xiao, G., Zou, C., Kohel, R. J., et al. (2015). Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nat. Biotechnol.* 33 (5), 524–530. doi: 10.1038/nbt.3208

Li, Y., Tan, X., Guo, J., Hu, E., Pan, Q., Zhao, Y., et al. (2022). Functional Characterization of MdTAC1a Gene Related to Branch Angle in Apple (Malus x domestica Borkh.). *Int. J. Mol. Sci.* 23 (3), 1-17. doi: 10.3390/ijms23031870

Li, H., Zhang, L., Hu, J., Zhang, F., Chen, B., Xu, K., et al. (2017). Genome-wide association mapping reveals the genetic control underlying branch angle in rapeseed (Brassica napus L.). *Front. Plant Sci.* 8, 1054. doi: 10.3389/fpls.2017.01054

Li, Y.-h., Zhao, S.-C., Ma, J.-X., Li, D., Yan, L., Li, J., et al. (2013). Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* 14 (1), 579. doi: 10.1186/1471-2164-14-579

Lonardi, S., Muñoz-Amatriain, M., Liang, Q., Shu, S., Wanamaker, S. I., Lo, S., et al. (2019). The genome of cowpea (Vigna unguiculata [L.] Walp.). *Plant J.* 98 (5), 767–782. doi: 10.1111/tpj.14349

Mao, L., Zhang, L., Zhao, X., Liu, S., van der Werf, W., Zhang, S., et al. (2014). Crop growth, light utilization and yield of relay intercropped cotton as affected by plant density and a plant growth regulator. *Field Crops Res.* 155, 67–76. doi: 10.1016/j.fcr.2013.09.021

McGarry, R. C., Prewitt, S. F., Culpepper, S., Eshed, Y., Lifschitz, E., and Ayre, B. G. (2016). Monopodial and sympodial branching architecture in cotton is differentially regulated by the Gossypium hirsutum Single Flower Truss and Self-Pruning Orthologs. *New Phytol.* 212 (1), 244–258. doi: 10.1111/nph.14037

Meng, F., Pan, Y., Wang, J., Yu, J., Liu, C., Zhang, Z., et al. (2020). Cotton duplicated genes produced by polyploidy show significantly elevated and unbalanced evolutionary rates, overwhelmingly perturbing gene tree topology. *Front. Genet.* 11. doi: 10.3389/fgene.2020.00239

Mi, H., and Thomas, P. (2009). PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol. Biol.* 563, 123–140. doi: 10.1007/978-1-60761-175-2_7

Montesinos, A., Dardick, C., Rubio-Cabetas, M. J., and Grimplet, J. (2021). Polymorphisms and gene expression in the almond IGT family are not correlated to variability in growth habit in major commercial almond cultivars. *PloS One* 16 (10), e0252001. doi: 10.1371/journal.pone.0252001

Motamayor, J. C., Mockaitis, K., Schmutz, J., Haiminen, N., Iii, D. L., Cornejo, O., et al. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol.* 14 (6), r53. doi: 10.1186/gb-2013-14-6-r53

Murchie, E. H., and Burgess, A. J. (2022). Casting light on the architecture of crop yield. *Crop Environ.* 1 (1), 74–85. doi: 10.1016/j.crope.2022.03.009

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., et al. (2014). The genome of Eucalyptus grandis. *Nature* 510 (7505), 356–362. doi: 10.1038/nature13308

Norton, E. R., Silvertooth, J. C., and Stedman, S. W. (1995). *"Plant population evaluation for upland cotton"* (University of Arizona (Tucson, AZ: College of Agriculture).

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., et al. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature* 492 (7429), 423–427. doi: 10.1038/nature11798

Pathak, R. S., and Singh, R. B. (1975). Genetic analysis of the duplicate loci, cluster and short branch in. *Theor. Appl. Genet.* 46 (6), 281–287. doi: 10.1007/BF00281150

Peng, J., Richards, D. E., Hartley, N. M., Murphy, G. P., Devos, K. M., Flintham, J. E., et al. (1999). 'Green revolution' genes encode mutant gibberellin response modulators. *Nature* 400 (6741), 256–261. doi: 10.1038/22307

Phytozome (2022a) *Gossypium hirsutum Coker v1.1* (Department of Energy's Joint Genome Institute). Available at: http://phytozome.jgi.doe.gov/info/GhirsutumCoker_v1_1 (Accessed September10, 2022).

Phytozome (2022b) *Gossypium hirsutum CSX8308 v1.1* (Department of Energy's Joint Genome Institute). Available at: http://phytozome.jgi.doe.gov/info/GhirsutumCSX8308_v1_1 (Accessed September 10, 2022).

Phytozome (2022c) *Gossypium hirsutum UA48 v1* (Department of Energy's Joint Genome Institute). Available at: http://phytozome.jgi.doe.gov/info/GhirsutumUA48_v1_1 (Accessed September 10, 2022).

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6), 841–842. doi: 10.1093/bioinformatics/btq033

Rathore, K. S., Chowdhury, V. K., and Hodges, T. K. (1993). Use of bar as a selectable marker gene and for the production of herbicide-resistant rice plants from protoplasts. *Plant Mol. Biol.* 21 (5), 871–884. doi: 10.1007/bf00027118

Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., and Korbel, J. O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28 (18), i333–i339. doi: 10.1093/bioinformatics/bts378

Reinhardt, D., and Kuhlemeier, C. (2002). Plant architecture. *EMBO Rep.* 3 (9), 846–851. doi: 10.1093/embo-reports/kvf177

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29 (1), 24–26. doi: 10.1038/nbt.1754

Roychoudhry, S., and Kepinski, S. (2015). Shoot and root branch growth angle control-the wonderfulness of lateralness. *Curr. Opin. Plant Biol.* 23, 124–131. doi: 10.1016/j.pbi.2014.12.004

Schmidt, J. E., Bowles, T. M., and Gaudin, A. C. M. (2016). Using ancient traits to convert soil health into crop yield: impact of selection on maize root and rhizosphere function. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00373

Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3 (6), 1101–1108. doi: 10.1038/nprot.2008.73

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463 (7278), 178–183. doi: 10.1038/nature08670

Shahbandeh, M. (2021) *Cotton yield Per Harvested acre in the U.S. 2020* (Statista). Available at: https://www.statista.com/statistics/191494/cotton-yield-per-harvested-acre-in-the-us-since-2000/ (Accessed April, 20 2023).

Shao, P., Peng, Y., Wu, Y., Wang, J., Pan, Z., Yang, Y., et al. (2022). Genome-wide association study and transcriptome analysis reveal key genes controlling fruit branch angle in cotton. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.988647

Si, Z., Liu, H., Zhu, J., Chen, J., Wang, Q., Fang, L., et al. (2018). Mutation of SELF-PRUNING homologs in cotton promotes short-branching plant architecture. *J. Exp. Bot.* 69 (10), 2543–2553. doi: 10.1093/jxb/ery093

Silow, R. A. (1946). Evidence on chromosome homology and gene homology in the amphidiploid new world cottons. *J. Genet.* 47 (2), 213–221. doi: 10.1007/BF02986788

Song, X., and Zhang, T. (2009). Quantitative trait loci controlling plant architectural traits in cotton. *Plant Sci.* 177 (4), 317–323. doi: 10.1016/j.plantsci.2009.05.015

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30 (9), 1312–1313. doi: 10.1093/bioinformatics/btu033

Stephens, S. G. (1955). LINKAGE IN UPLAND COTTON. *Genetics* 40 (6), 903–917. doi: 10.1093/genetics/40.6.903

Sun, Q., Du, X., Cai, C., Long, L., Zhang, S., Qiao, P., et al. (2016b). To be a flower or fruiting branch: insights revealed by mRNA and small RNA transcriptomes from different cotton developmental stages. *Sci. Rep.* 6 (1), 23212. doi: 10.1038/srep23212

Sun, C., Wang, B., Wang, X., Hu, K., Li, K., Li, Z., et al. (2016a). Genome-wide association study dissecting the genetic architecture underlying the branch angle trait in rapeseed (Brassica napus L.). *Sci. Rep.* 6 (1), 33673. doi: 10.1038/srep33673

Sun, Q., Xie, Y., Li, H., Liu, J., Geng, R., Wang, P., et al. (2022). Cotton GhBRC1 regulates branching, flowering, and growth by integrating multiple hormone pathways. *Crop J.* 10 (1), 75–87. doi: 10.1016/j.cj.2021.01.007

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., et al. (2014). An improved genome release (version Mt4.0) for the model legume Medicago truncatula. *BMC Genomics* 15 (1), 312. doi: 10.1186/1471-2164-15-312

Thompson, C. J., Movva, N. R., Tizard, R., Crameri, R., Davies, J. E., Lauwereys, M., et al. (1987). Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus. *EMBO J.* 6 (9), 2519–2523. doi: 10.1002/j.1460-2075.1987.tb02538.x

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, populus trichocarpa (Torr. & Gray). *Science* 313 (5793), 1596–1604. doi: 10.1126/science.1128691

Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., et al. (2013). The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat. Genet.* 45 (5), 487–494. doi: 10.1038/ng.2586

Waite, J. M., and Dardick, C. (2018). TILLER ANGLE CONTROL 1 modulates plant architecture in response to photosynthetic signals. *J. Exp. Bot.* 69 (20), 4935–4944. doi: 10.1093/jxb/ery253

Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., et al. (2012). The draft genome of a diploid cotton Gossypium raimondii. *Nat. Genet.* 44 (10), 1098–1103. doi: 10.1038/ng.2371

Whipple, C. J., Kebrom, T. H., Weber, A. L., Yang, F., Hall, D., Meeley, R., et al. (2011). grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *Proc. Natl. Acad. Sci. U.S.A.* 108 (33), E506–E512. doi: 10.1073/pnas.1102819108

Xu, D., Qi, X., Li, J., Han, X., Wang, J., Jiang, Y., et al. (2017). PzTAC and PzLAZY from a narrow-crown poplar contribute to regulation of branch angles. *Plant Physiol Biochem* 118, 571–578. doi: 10.1016/j.plaphy.2017.07.011

Yang, X. C., and Hwa, C. M. (2008). Genetic modification of plant architecture and variety improvement in rice. *Heredity* 101 (5), 396–404. doi: 10.1038/hdy.2008.90

Yang, G.-z., Luo, X.-j., Nie, Y.-c., and Zhang, X.-L. (2014). Effects of plant density on yield and canopy micro environment in hybrid cotton. *J. Integr. Agric.* 13 (10), 2154–2163. doi: 10.1016/S2095-3119(13)60727-3

Yao, H., Zhang, Y., Yi, X., Zuo, W., Lei, Z., Sui, L., et al. (2017). Characters in light-response curves of canopy photosynthetic use efficiency of light and N in responses to

plant density in field-grown cotton. *Field Crops Res.* 203, 192–200. doi: 10.1016/j.fcr.2016.12.018

Yu, B., Lin, Z., Li, H., Li, X., Li, J., Wang, Y., et al. (2007). TAC1, a major quantitative trait locus controlling tiller angle in rice. *Plant J.* 52 (5), 891–898. doi: 10.1111/j.1365-313X.2007.03284.x

Yuan, D., Tang, Z., Wang, M., Gao, W., Tu, L., Jin, X., et al. (2015). The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres. *Sci. Rep.* 5 (1), 17662. doi: 10.1038/srep17662

Zhang, L., Chia, J.-M., Kumari, S., Stein, J. C., Liu, Z., Narechania, A., et al. (2009). A genome-wide characterization of microRNA genes in maize. *PloS Genet.* 5 (11), e1000716. doi: 10.1371/journal.pgen.1000716

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., et al. (2015). Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* 33 (5), 531–537. doi: 10.1038/nbt.3207

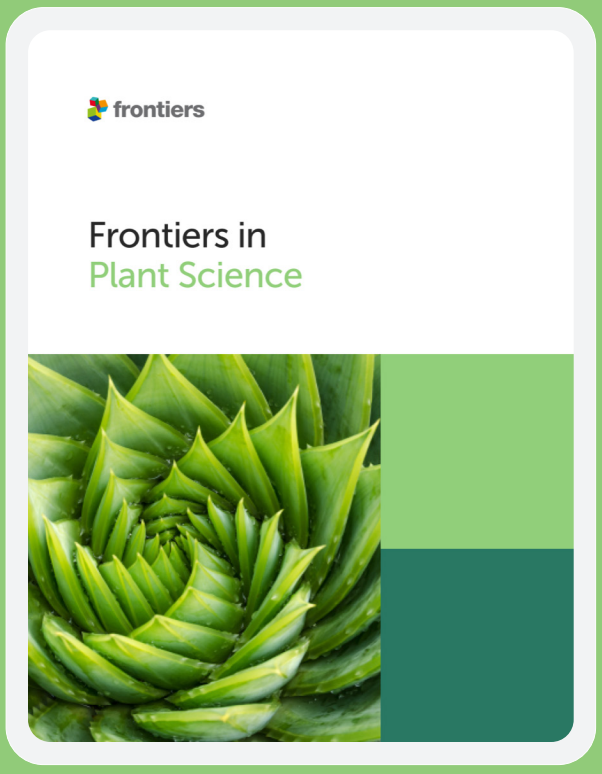
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact